



## Die USA als Argument:<sup>1</sup> Anmerkungen anlässlich der Debatte um Bildungsstandards

*S. Karin Amos*

*Johann Wolfgang Goethe Universität Frankfurt a.M.*

### *Abstract*

The article takes a closer look at the mode of reference to international controversies on educational standards in the German academic debate in favour of their introduction. It argues that the references to the US-American state of the debate are highly selective and do not yet sufficiently address the central issue behind the present American standards-based educational reform: it is embedded in the educational initiative of the current political administration known as the No Child Left Behind (NCLB) Act. The Act is directly linked to a massive national mobilisation and claims to solve the problems associated with equity and quality by holding the educational system and not the individual responsible for test-measured learning outcomes. Hence, performance standards have been linked to high stakes testing rewarding or punishing schools. Even before massive disillusionment with NCLB has set in, critics have warned against the negative effects of high stakes testing especially with regards to students at risk. The example shows that the attempt to force schools to ensure that students meet minimum standards has adverse effects.

### 1. Qualität und Egalität

Nachfolgend nehme ich einen Aspekt in den Blick, der im Kontext der Durchsetzung neuer Steuerungsmodelle im Bildungssystem eine zentrale, aber nicht ausdrücklich thematisierte Rolle spielt: die Vereinbarkeit von Qualität und Egalität. Breiter wissenschaftlicher, öffentlicher und politischer Konsens für die Umorientierung von Input auf Output kommt nur zustande, wenn die Gültigkeit des folgenden Grundsatzes erhalten bleibt: Im modernen gesellschaftlichen Selbstverständnis entscheiden nicht mehr – wie in feudalen, vormodernen Gesellschaften – Geburt oder askriptive Merkmale, sondern „durch eigenes Handeln erworbene Eigenschaften“ über die gesellschaftliche Positionierung von Individuen (vgl. Berger, 2005, S. 8). Hierbei hat der Bildungserfolg als erreichte formale Stufe im Bildungssystem mit der Zertifizierung eine wesentliche Rolle. So wurde in den Auseinandersetzungen um die Befunde der *Large Scale As-*

essment Studien – wie jüngst bei PISA – nochmals dezidiert auf den (inakzeptabel) engen Zusammenhang zwischen sozialer Herkunft und Bildungserfolg im deutschen Bildungssystem hingewiesen und damit nicht nur grundsätzlich die Notwendigkeit von output-orientierten Bildungsreformen begründet, sondern auch auf eine breite politische, professionelle und wissenschaftliche Akzeptanz dieser Umsteuerungen abgezielt. Die geringe Übereinstimmung zwischen gemessener Kompetenzstufe, Noten und sogar den besuchten Schulformen der Schülerinnen und Schüler widerspricht dem meritokratischen Selbstverständnis von Schule (vgl. auch Radtke, 2004).

Kernstück der aktuellen, auch als ‚ergebnisorientiert‘ oder ‚Results Based‘<sup>2</sup> bezeichneten, Bildungsreformen sind Debatten und Maßnahmen zur Konzeptualisierung und Umsetzung von Bildungsstandards. Diesen liegen wiederum didaktische und psychologische Kompetenzmodelle zugrunde. Die ergebnisorientierten Bildungsstandards sollen zum Systemmonitoring, zur Schulevaluation und – eingeschränkt – auch für die individuelle Rückmeldung an Schülerinnen und Schüler genutzt werden. Auf den Punkt gebracht heißt es in der vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Expertise zu nationalen Bildungsstandards – häufig als ‚Klieme-Gutachten‘ zitiert:

Nationale Bildungsstandards formulieren verbindliche *Anforderungen an das Lehren und Lernen in der Schule*. Sie stellen damit innerhalb der Gesamtheit der Anstrengungen zur *Sicherung und Steigerung der Qualität schulischer Arbeit* ein zentrales Gelenkstück dar. Bildungsstandards benennen präzise, verständlich und fokussiert die wesentlichen Ziele der pädagogischen Arbeit, ausgedrückt als erwünschte Lernergebnisse der Schülerinnen und Schüler. ... Bildungsstandards ... *legen fest, welche Kompetenzen die Kinder oder Jugendlichen bis zu einer bestimmten Jahrgangsstufe mindestens erworben haben sollen*. Die Kompetenzen werden so konkret beschrieben, dass sie in Aufgabenstellungen umgesetzt und prinzipiell mit Hilfe von Testverfahren erfasst werden können. ...

Die *erste Funktion* der Bildungsstandards besteht in der *Orientierung* der Schulen auf verbindliche Ziele. Lehrkräfte, aber auch Lernende und deren Eltern können sich darauf bei der pädagogischen Weiterentwicklung von Schule und Unterricht beziehen. ...

Die *zweite Funktion* der Bildungsstandards besteht darin, dass auf ihrer Grundlage *Lernergebnisse erfasst und bewertet werden*. So lässt sich feststellen, inwieweit das Bildungssystem seinen Auftrag erfüllt hat (*Bildungsmonitoring*), und die Schulen erhalten eine Rückmeldung über die Ergebnisse ihrer Arbeit (*Schulevaluation*). Die Standards können auch Auskunft geben über individuelle Diagnostik und Förderung.“ Von letzterer Verwendung wird jedoch aus methodischen Gründen abgeraten.

Festgelegt wird ferner, dass die ... verbindliche Festsetzung der Standards und die standardbezogene Evaluation von Schulen *Angelegenheit der einzelnen Länder* sind (Klieme et al., 2003, S. 9 f.).

Für die Entwicklung von Standards und Tests sowie das (nationale) Bildungsmonitoring – das an PISA und andere internationale Untersuchungen gekoppelt werden könnte – schlägt die Expertise eine nationale Agentur vor.

Die Einführung und Durchsetzung von Bildungsstandards hat aber neben direkten Konsequenzen der Umsteuerung für die Schulen und die Entwicklung des Bildungs-

systems weitere Folgen: Sie ist mit einer Rekonfiguration des Bildungsbereichs in seinen wissenschaftlichen, politischen und handlungspraktischen (professionsbezogenen) Dimensionen verbunden (vgl. auch Van Zanten, 2005). Dieser Prozess kann nicht unabhängig von bestimmten Interessenlagen betrachtet werden, um deren Durchsetzung zurzeit gerungen wird. Der Wissenschaft kommt bei der Legitimierung und Dissemination von Argumentationsstrategien eine zentrale Rolle zu, und zwar auch aufgrund ihrer Referenzen auf internationale Erfahrungen und Entwicklungen. Diese Referenzen beziehen sich auf die Fachdebatte zu Bildungsstandards sowie auf das Thema Qualität und Egalität.

Nachfolgend geht es um die Betrachtung dieser Expertise unter der Perspektive, wie die angesprochenen Aspekte der US-amerikanischen Debatte aufgenommen werden. Der Fall USA ist interessant, weil die Einführung ergebnisorientierter Bildungsstandards in Verbindung mit der Diagnose einer dem Selbstverständnis moderner Gesellschaften widersprechenden Bildungsungleichheit in einem äußerst kontroversen Gesetz zusammen geführt wurden: dem 2002 verabschiedeten *No Child Left Behind Act*. Daher wird im zweiten Teil das Gesetz und werden die dabei auftauchenden Fragen und Probleme thematisiert. Vor diesem Hintergrund und mit Bezug auf die deutsche Bildungsstandard-Diskussion werden in einem Resümee vor allem Fragen generiert (und einige Antworten gegeben).

## 2. Referenzen auf internationale Entwicklungen in der Expertise zu nationalen Bildungsstandards

Das ‚Klieme-Gutachten‘ bezieht sich in mehreren Passagen auf die internationale Diskussion. Im Rahmen der Diskussion von Merkmalen guter Bildungsstandards heißt es:

Beispiele für unrealistische Erwartungshaltungen finden sich in dem No child left behind Gesetz, das im Frühjahr 2002 in den USA verabschiedet wurde. ... Das US-Gesetz sieht vor, dass alle US-Bundesstaaten Standards und zugehörige Testverfahren entwickeln und – jeweils nach eigenen Kriterien – Mindesterwartungen in Grundkompetenzen, vor allem in der Lesefähigkeit, festsetzen. Das Gesetz erwartet sodann von den einzelnen Schulen, von Jahr zu Jahr den Anteil der Schülerinnen und Schüler, die das Mindestniveau verfehlen, kontinuierlich zu senken. Wissenschaftler haben darauf aufmerksam gemacht, dass derartige geradlinige Entwicklungen auf Schulebene über 10 Jahre hinweg höchst unwahrscheinlich sind (Klieme et al., 2003, S. 30).

In der nächsten Referenz auf die US-amerikanische Debatte geht es um die Unterscheidung verschiedener Formen von Bildungsstandards. So werden vor allem drei Punkte behandelt: Sind Bildungsstandards auf Input- oder auf Output-Messung ausgerichtet? Auf welcher Ebene werden die Standards formuliert? Auf welches Zielniveau sind sie bezogen: auf den Mindeststandard, den Regelstandard oder den Maximalstandard. In der Erörterung der angloamerikanischen Debatte wird vor allem zwischen *process* und *performance standards* unterschieden, wobei *performance standards*

(Leistungsstandards) als stark verallgemeinerte Form, ja als Synonym für Standards schlechthin, gelten und *process standards* vor allem im Rahmen mathematikdidaktischer Erwägungen zur Sprache kommen. An dieser und an weiteren Stellen bezieht sich die Expertise positiv auf Publikationen des *National Council of Teachers of Mathematics* (NCTM). Die Ausführungen zu Formen von Standards enden mit einer kontrastierenden Betrachtung der die Verwendung bestimmter Konzeptionen von Standards beeinflussenden unterschiedlichen kulturellen und politischen Traditionen und Steuerungsmodellen anhand der Beispiele der USA und Skandinaviens (Klieme et al., 2003, S. 34 f.).

Auffällig ist dabei die quantitative Verteilung: Auf mehr als einer Seite geht es um die USA, deren dezentral gesteuertes Bildungssystem und den daraus erwachsenden Bedarf an verbindlichen Vorgaben und Tests wie die NCTM-Standards. Diese Vorgaben und Tests sollen Staaten und Schulen vergleichbar machen. Besonders bei den NCTM-Maßgaben geht es um den Anspruch, durch interessante und anspruchsvolle Inhalte den Leistungsstand aller Schülerinnen und Schüler zu verbessern. Andererseits ließe sich auch die in einigen Bundesstaaten beobachtbare Tendenz ablesen, dass als Folge der Einführung ‚reformbasierter Standards‘ (wobei dieser Ausdruck ohne Erläuterung bleibt) eine disproportionale Orientierung an Testperformanz bei gleichzeitiger Beschränkung des Curriculums auf stark reduzierte Inhalte erfolge. Die Expertengruppe resümiert, dass solche Praktiken den Intentionen der anspruchsvollen Standards des NCTM widersprächen.

Anschließend wird auf einer halben Seite über das nationale Curriculum in Schweden berichtet, das sich an allgemeinen Zielen orientiere, etwa der Autonomie der einzelnen Schule, der Individualität der Schülerinnen und Schüler, dem kulturellen Erbe, der Internationalisierung der Gesellschaft und des Bildungssystems sowie dem technologischen Wandel. Daneben hätten auch andere nordeuropäische Länder, wie Finnland, auf Flexibilität und individuelle Wahlmöglichkeiten als Eckpfeiler der Reform gesetzt – allerdings seien in jüngster Zeit Kerncurricula, Basiskompetenzen und entsprechende Kompetenzniveaus verbindlich definiert und landesweit eingeführt worden.

Die Schlussfolgerung: Die Beispiele zeigten, dass Standards auf völlig unterschiedlichen theoretischen Annahmen, Wertvorstellungen und gesellschaftlichen Zielen basieren können. Diese Annahmen, Werte und Ziele bestimmten aber ganz wesentlich die Akzeptanz von Standards, ihre Umsetzung in der Schule und die Möglichkeiten der täglichen Arbeit mit diesen Standards.

Im vierten Kapitel geht es um Standards für die Weiterentwicklung von Schule und Unterricht. Dabei wird auf die wünschenswerte Kooperation zwischen Schule und Eltern, respektive zwischen Schule und älteren Schülerinnen und Schülern verwiesen. So lautet das Desiderat: Bildungsstandards sollten so formuliert sein, dass Eltern und ältere Schülerinnen und Schüler sie nachvollziehen könnten. Standards vermittelten mit

größerer Klarheit als zuvor, auf welche Kompetenzen es in der Schule ankomme und wie diese über die Schullaufbahn hinweg aufzubauen seien. Wünschenswert sei die veränderte Beteiligung und Mitwirkung von Schülerinnen und Schülern und Eltern, etwa im Sinne einer stärkeren Übernahme der Planung oder auch einer Beteiligung und Überwachung von Lernprozessen (Klieme et al., 2003, S. 47 f.). Komplementär ist daran gedacht, dass sich die Schule den Eltern oder anderen ‚Vertragspartnern‘ (die Bezeichnung stammt von der Verf.) gegenüber verpflichtet, auf bestimmte Ziele hinzuarbeiten und dafür Fördermaßnahmen bereitzustellen. In diesem Sinne spricht die Expertise davon, dass Standards einen wichtigen Bezugspunkt für Gespräche zwischen Eltern, Schülern und Lehrern bilden könnten. Klare und verbindliche Erwartungen ermöglichten es, den individuellen Lernweg zu planen, Lernhindernisse zu erkennen und bestmögliche Fördermöglichkeiten anzusprechen.

Ein weiterer zentraler Aspekt der Expertise sind Konsequenzen der Festlegung von Standards und Leistungsmessungen: Wenn Mindeststandards festgelegt würden, folgten Fragen der Bewertung von Lernergebnissen. Es müsse klar unterschieden werden zwischen der Diskussion über die Qualität der Lernergebnisse und der über die Festlegung von Noten, die Vergabe von Zertifikaten und die Entscheidung über die Schullaufbahn von Schülerinnen und Schülern. Standards sollten ein Kriterium für die Lernergebnisse jeder Schule, jeder Klasse und konsequenterweise jeder einzelnen Schülerin und jedes einzelnen Schülers sein. Sie dürften aber nicht mit Benotungs- und Prüfungsregeln oder gar mit Regelungen zum Übergang auf weiterführende Schulen verwechselt werden. Die Standards ließen die professionelle Kompetenz der Lehrenden unangetastet.

Wichtig wird diese Abgrenzung bei auf Standards bezogenen Tests: Erfahrungen mit dem *high stakes assessment* (high stakes bedeutet, dass mit den Testergebnissen Sanktionen bzw. Gratifikationen verbunden sind) in den USA verdeutlichten die negativen Konsequenzen des Missbrauchs in der Schule, denn der Test beginne den Unterricht zu dominieren – was bei intelligenten Tests, die Raum für komplexes Denken, Problemlösen und Kreativität ließen, durchaus fruchtbar sein könne, bei schlechten Tests hingegen fatale Auswirkungen zeitige. Zudem seien Tests, die für Bildungsmonitoring und Schulevaluation entwickelt wurden, in der Regel ungeeignet für die individuelle Diagnostik. In jedem Falle gehe es aber um ergebnis-, nicht um prozessorientierte (*opportunity to learn*) Standards.

Interessant ist zudem die Aussage, dass langfristig die Koexistenz nationaler Bildungsstandards mit detaillierten curricularen Vorgaben auf Länderebene unwahrscheinlich erscheine. Kerncurricula lösten vermutlich die jetzt bestehenden Lehrpläne und Rahmenrichtlinien ab (Klieme et al., 2003, S. 53).

Die problematischen Aspekte der amerikanischen Erfahrung werden nicht nur mit den positiven skandinavischen kontrastiert, sondern ebenso mit dem *Large Scale Assessment* Typus PISA (ebd., S. 104).

Wie überzeugend die Argumentation der die Einführung von ergebnisorientierten Bildungsstandards empfehlenden Expertengruppe wirkt, illustriert folgendes Beispiel. Eine den Titel ‚Nationale Bildungsstandards – Wundermittel oder Teufelszeug?‘ tragende Publikation der Gewerkschaft für Erziehung und Wissenschaft (GEW, 2003), bezieht sich ausgesprochen positiv auf die Bildungsstandard-Expertise, wobei folgende Punkte hervorgehoben werden:

- Das ‚Klieme-Gutachten‘ berücksichtigt internationale Entwicklungen und Erfahrungen, um aus guten Beispielen wie aus Fehlern zu lernen.
- Die Empfehlungen beruhen auf einem optimistischen Menschenbild und einer an individueller Förderung, nicht an Selektion orientierten Pädagogik: „Jedes Kind ist wichtig, keines bleibt zurück“.
- Leistungsfähigkeit und Chancengleichheit sind Leitkategorien.
- Der Lehrerprofessionalität wird für die Entwicklung von Standards und Beispielaufgaben ein hoher Stellenwert zugemessen.
- Vorbild ist die skandinavische Schul- und Qualitätsphilosophie, nicht das angelsächsische Konkurrenzmodell. Jede Schule soll sich ohne jährliches Schulranking und ohne die Androhung öffentlicher Bloßstellung verbessern und entwickeln können.
- Pädagogische Freiräume sollen gesichert werden. Tests und Vergleichsarbeiten sollen sparsam eingesetzt werden und die Bewertung durch die Lehrerin bzw. den Lehrer nicht ersetzen (ebd., S. 10).

In dieser Lesart wird einem als Konkurrenzmodell perzipierten angloamerikanischen ein im Umkehrschluss als demokratisch und inklusiv wahrgenommener skandinavischer Konterpart gegenübergestellt. Daraus werden besonders für die Konzeption der Lehrerprofessionalität gravierende Konsequenzen abgeleitet: kein öffentliches *namings and shaming*, kein *teaching to the test*, keine Einmischung in die Leistungsbeurteilungs- und Übergangsempfehlungskompetenz der Lehrerinnen und Lehrer.

### 3. Die Mission von *No Child Left Behind*

Im Gegensatz zum ausführlichen und wiederholt aufgenommenen Rekurs auf die Mathematikstandards der NCTM, spielt das *No Child Left Behind* Gesetz trotz seines grundsätzlichen Charakters – der bislang gravierendste Eingriff des amerikanischen Bundes in die Bildungshoheit der Staaten – in der deutschen Bildungsstandard-Diskussion eine eher randständige, tendenziell negative Rolle. Dies spiegelt sich auch in der zitierten GEW-Publikation wider, die zwar die ‚Klieme-Expertise‘ mit einem positiven Menschenbild ‚Kein Kind soll zurück gelassen werden‘ assoziiert, dabei aber nicht thematisiert, dass diese Formulierung eine Übertragung des programmatischen Titels *No Child Left Behind* ist. Damit stellt sich die Frage, mit welchen Begründungen das jeweils reklamierte positive und inklusive Menschenbild in unterschiedlichen Län-

derkontexten in weit divergierende bildungspolitische Maßnahmen übersetzt wird. Was ist in diesem Zusammenhang der Hintergrund für das angloamerikanische Konkurrenzmodell?

Die Wahl des programmatischen Titels *No Child Left Behind* ist ein bildungspolitisches Glanzstück, allein schon aufgrund der geschickten Konstruktion des sprachlichen Ausdrucks, der nicht zwischen Desiderat *no child (shall be) left behind* und Resultat *no child (is) left behind* unterscheidet. Signifikanter ist die konnotative Ebene: Hier trifft *No Child Left Behind* den Kern des nationalen Selbstverständnisses in einem doppelten Sinne. In erster Linie ist *No Child Left Behind* im spezifisch amerikanischen Gesellschaftskontext verortet und trägt die deutliche Signatur der patriotischen Stimmung in der Folge des Elften September, zum anderen, dies illustriert die GEW-Publikation, trifft dieser sprachliche Ausdruck im allgemeinen Sinne einen zentralen Nerv moderner Gesellschaften: Wenn das Wohl aller Kinder dezidiert angesprochen ist, wird der enge Zusammenhang zwischen Nationalstaat und öffentlicher Bildung und Erziehung augenfällig. Schule als evolutionäre Universalie (Adick, 1992) soll die vertikale Differenzierung der Gesellschaft überformen; sie beansprucht, aus allen Kindern eigenständige und verantwortliche zukünftige (Staats)Bürger zu machen. Auf die Einlösung dieses Anspruchs zielt *No Child Left Behind* (NCLB) ab:

The purpose of this title is to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments (Änderung des Titels I, ESEA: Improving the Academic Achievement of the Disadvantaged, Sec. 1001).

Da dieses hehre Ziel alle Partikularinteressen überformt – jedenfalls auf semantischer Ebene und solange keine negativen Erfahrungen vorliegen –, kann NCLB als Fallbeispiel für den Versuch gelten, im Dienste eines übergreifenden Interesses föderale Regeln soweit außer Kraft zu setzen, dass der Bund bis in die einzelne Schule ‚hineinregiert‘. Legitimiert wird dies mit einer dem deutschen Publikum bekannten Diagnose (wenn auch als Resultat eines sich vom deutschen stark unterscheidenden Bildungsgefüges): ein für ein demokratisches Land inakzeptables Maß an Bildungsungleichheit und die entsprechende Kritik des Bildungssystems. Dies bestätigt auch die PISA-Studie. Insbesondere lautet der Vorwurf, dass in den USA zwischen den Leistungsstärksten und den Leistungsschwächsten die international größten Disparitäten lägen, wobei zwischen sozialer Herkunft und Bildungserfolg bzw. Misserfolg ein starker Zusammenhang bestehe – auf diesen Vorwurf antwortet NCLB.

Die Säulen des Gesetzes: Rechenschaftspflicht der Schulen, freie Schulwahl, größere Entscheidungsfreiheit auf lokaler Ebene und wissenschaftliche Fundierung der Bildungsmaßnahmen, sind nicht neu, sie rekombinieren Elemente vorangegangener Reformperioden. Im Detail sieht NCLB folgendes Procedere vor: Der jeweilige Bundesstaat beantragt Mittel (*grants*) bei der Bundeszentrale, muss dabei aber durch seine *State Educational Agency* (SEA) Bildungsstandards definieren und begründen. Die

SEA verpflichtet sich, Mittel (*subgrants*) für die *Local Educational Authorities* bereitzustellen, und überwacht mittels Assessments, welche Schulen den staatlichen Minimalanforderungen gerecht werden. Um den Staaten eine Orientierung für ihre Bildungsstandard-Konzeptionen zu geben, hat die Bundeszentrale einen Leitfaden erstellt (vgl. US Department of Education, 2003). Die Steuerung verläuft also über die Argumentation, dass sich kein Staat dem Ziel, kein Kind zurückzulassen, widersetzen kann – erst recht nicht in einer Zeit nationaler Mobilisierung – sowie über das Medium Geld, denn in allen Staaten der USA gibt es Schuldistrikte, die zur Deckung ihres Finanzbedarfs auf Bundesmittel zurückgreifen. In weniger als zehn Jahren, so das Ziel, sollen alle Kinder aller Leistungsgruppen in zentralen Fächern (*reading, language arts, mathematics*) die benchmark *proficiency*, die dem Regelstandard entsprechenden Kompetenzen, erreichen. Dafür werden alle Kinder von der dritten bis zur achten Klasse<sup>3</sup> jährlich hinsichtlich der Lernfortschritte getestet (*Adequate Yearly Progress – AYP*), und die Ergebnisse dieser Tests werden in Form von Ranglisten der Schulen veröffentlicht. Damit sollen sich Eltern ein Bild machen können über die Schulleistung, die Fortschritte der Bildungsleistungsbilanz ihres Staates, die Qualifikationen der Lehrer, die Qualität der Schule und den Lernzuwachs ihres Kindes in den getesteten Schlüsselfächern.

Schneiden Schulen sehr gut ab, werden Schulleiter mit Geldleistungen belohnt, schneiden sie schlecht ab und gelten als versagende Schulen, kann der Bund diese Schule ‚übernehmen‘, sie umgestalten oder auch schließen. Gilt eine Schule als *failing*, als ‚versagend‘, ist den Eltern freigestellt, ihr Kind auf einer anderen öffentlichen Schule, einer *Charter School*<sup>4</sup> oder einer privaten Schule anzumelden.

Bei NCLB geht es nicht in erster Linie um Ursachen für die Defizite der Kinder, der Familien oder des Umfelds, sondern um die Verbesserung des Lernerfolgs; vor allem Resultate interessieren – wie sie zustande kommen wird zunächst nicht gefragt.

#### 4. Der historische Kontext von NCLB

Das 2002 von Präsident George W. Bush unterzeichnete Gesetz hatte zuvor mit großer Mehrheit beider Parteien den Senat passiert. NCLB bezeichnet die aktuelle Änderung des *Elementary and Secondary Education Act* (ESEA)<sup>5</sup>, einem Gesetz, das 1965 unter der Johnson-Administration verabschiedet wurde und – vor allem unter dem Rechtstitel I – der Bekämpfung der Armut (*War on Poverty*) dienen sollte. Zentrales Ziel der Intervention war, den engen Zusammenhang zwischen sozialer Herkunft und Bildungserfolg aufzubrechen und perspektivisch zu beseitigen. Hintergrund für die prominente gesellschaftliche Platzierung des Problems bildeten ökonomische und politische Überlegungen, die sich in Konzeptionen zur größeren Ausschöpfung der Humankapitalreserven trafen. So lassen sich mehrere Bezugspunkte identifizieren: Einer waren die seitens des amerikanischen Wohlfahrtsstaats gefürchteten Folgekosten mangelnder Bildung und Qualifikation; ein weiterer ist im Kalten Krieg zu sehen. Im tech-

nologischen Wettbewerb zwischen den Supermächten USA und UdSSR ging es den USA um die Ausschöpfung aller Begabungsreserven. Stand wenige Jahre zuvor die Elitebildung im Zentrum bildungspolitischer Anstrengungen, ging es spätestens seit der Mitte der 1960er Jahre um die umfassende Förderung aller Kinder. In sozialpolitischer Hinsicht diente ESEA im Kern der Wahrung des inneren Friedens in äußerst konfliktreichen Zeiten (erinnert sei an die Schwarze Bürgerrechtsbewegung, die zunehmende Militarisierung vor allem der urbanen schwarzen Jugend und die Ghettonruhen Mitte der 1960er Jahre). Bildungsungleichheit wurde vor allem als ökonomisches Problem gesehen – dies betraf besonders ethnische Minderheiten, vor allem Amerikaner afrikanischer Herkunft, aber auch weiße Bevölkerungsteile in abgelegenen ländlichen Gebieten.

ESEA mit dem Flaggschiff kompensatorische Erziehung setzte auf mehr Input, vor allem durch umfassende Förderungsangebote. Diese wurden sukzessive auf das ganze Bildungssystem ausgeweitet. Die Bildungsökonomien unter Milton Friedman setzten auf Wettbewerb – daher die Ursprünge des ‚Konkurrenzmodells‘ in der öffentlichen Bildung –, weil sie davon ausgingen, dass die durch Bildungsgutscheine ermöglichte Wahl zwischen öffentlichen und privaten Schulen für mehr Anreize zur schulischen Qualitätsentwicklung sorgen würde. Bei allen Differenzen in den Antworten – Inputsteigerung ist sicher eine andere Form der Intervention als etwa ein Bildungsgutscheinmodell – herrschte in der Diagnose weitgehend Einhelligkeit: Das amerikanische Bildungssystem werde seiner Kernaufgabe, *alle* Kinder auf ein eigenständiges Leben vorzubereiten, nur unzureichend gerecht. Dies werde deutlich am gruppenspezifisch unterschiedlich verteilten Bildungserfolg, der in deutlichem Widerspruch stehe zu einer der Grundüberzeugungen moderner Gesellschaften: der gruppenspezifischen Verteilung von Begabung. Andererseits stellen Wissenschaftler (vgl. Berliner, 2004) bei der Betrachtung der Funktion des Bildungssystems mit Blick auf die vertikale Differenzierung der Gesellschaft fest, dass das amerikanische Schulsystem hoch effizient sei. Es erbringe hervorragende Resultate vor allem, wenn auch nicht ausschließlich, für die Euro-Amerikaner der privilegierten Schichten. Dieser Befund stimmt im Kern mit der deutschen Bildungsungleichheits-Diskussion überein.

## 5. Die Rolle standardisierter Tests in der amerikanischen Bildungspolitik

Unabhängig von den Mustern der Bildungsreform der 1960er Jahre mit Input und der aktuellen mit Output ist zu beachten, dass Tests und Standards in der amerikanischen Bildungsdebatte schon lange eine zentrale Rolle spielen und in der tiefen Überzeugung von der Effektivität technologischer Intervention im Bildungssystem wurzeln (Koretz, 1992; Resnick, 1982).<sup>6</sup> Zwar waren Tests zum Systemmonitoring und zur Evaluation der Lehrer von Anfang an intendiert, einige Forscher sehen den Ursprung bereits im ESEA-Gesetz von 1965 (Airasian, 1987), allerdings wurden Tests seit den 1970er Jahren in größerem Umfang vor allem für die individuelle Diagnostik der Lernenden ein-

gesetzt. Damit verschob sich die amerikanische Bildungsreform von einer Input und Chancengleichheitsorientierung auf Output und Leistung. Auslöser waren u.a. die Veröffentlichung schlechter Ergebnisse im zentralen College-Eingangstest, dem *Scholastic Aptitude Test* (SAT), die Unzufriedenheit von Betrieben mit den schlechten Voraussetzungen der *High School* Absolventen und eine allgemeine Sorge um mangelnde Kompetenzen und Fertigkeiten der nachwachsenden Generation. Zu dieser Zeit begann das Begriffspaar Qualität – bezogen auf die Schülerleistung – und Egalität – bezogen auf die gesellschaftlichen Teilhabemöglichkeiten – eine zentrale Rolle in den nationalen Debatten zu spielen. Tests avancierten zum Hauptreformerinstrument der Intervention. Je mehr sie sich durchsetzten und als Mindeststandards zu einem Kriterium für den erfolgreichen Abschluss der *High School* oder auch nur für die Versetzungsent-scheidung in die nächst höhere Klasse genutzt wurden, desto offensichtlicher wurde, dass vor allem Schülerpopulationen aus Minderheiten, entgegen der Intention der Maßnahmen, negativ von den Tests betroffen waren. Eine Reihe von Rechtsklagen war die Folge (Baratz-Snowden, 1993, S. 314).

In Folge der Veröffentlichung von *A Nation at Risk* (National Commission of Excellence in Education, 1983) kam es zur nächsten bildungspolitischen Verschiebung zwischen *excellence* und *equity* als besonderen Varianten der Qualitäts- und Egalitätsdebatte. Die Sorge galt der Bildungselite und deren umfassender Förderung sowie der fortdauernden Bildungskluft. Der *National Council on Education Standards and Testing* (1992) sah das Heil der Bildungsreform in neuen Assessment-Strategien und prozessorientierten *opportunity to learn*-Standards. Mit dieser Intervention sollte der Motor des Bildungssystems einen ‚Schockstart‘ erleben. Zentral war der Einsatz von Tests in Verbindung mit Verantwortlichkeit und Rechenschaft. Die Tests wurden komplexer und anspruchsvoller; mit schlechten Testergebnissen verbundene Sanktionen wurden auf die Schulen ausgeweitet. Zwar erlahmte der Enthusiasmus für Tests Ende der 1980er Jahre, da die Praxis des *teaching to the test* die Testergebnisse verfälschte (was war nun tatsächlicher Lernzuwachs und was nicht), und aufgrund der einseitigen Konzentration auf kognitive Leistungen auch der Unterrichtsqualität schadete (Koretz, Linn, Dunbar & Shephard, 1991). Allerdings wurde an der test-basierten Rechenschaft festgehalten. Um das *Performance Assessment*, die Leistungsmessung, differenzierter vornehmen zu können, wurden neue Formen eingeführt, die beispielsweise Portfolios einbezogen, das Durchführen von naturwissenschaftlichen Experimenten vorsahen oder auch die Bearbeitungsform variierten und zwischen Einzel- und Gruppenbearbeitung wechselten.

Haney (2002) schließt aus der Analyse der Testdaten für Massachusetts, dass sich Anreize oder Strafen als Folge guter oder schlechter Testergebnisse vor allem für kleine Schulen negativ auswirkten, während sie bei größeren kaum Wirkungen zeitigten. Weiter würden häufig Schulen mit heterogener Schülerschaft bestraft. Schließlich seien

Tests ungeeignet, um *Best practice*-Schulen zu identifizieren; es gebe effizientere Methoden, um Informationen über nachahmenswerte Schulen zu erhalten.

Abgesehen von lokalen und zeitlich begrenzten Erfolgen ist es in den vergangenen vierzig Jahren seit Inkrafttreten des ESEA und trotz zahlreicher Bildungsreformen nicht gelungen, den engen Zusammenhang zwischen Bildungserfolg und sozialer Herkunft zu lockern. Nach wie vor herrscht in den USA eine tiefe Kluft zwischen privilegierten und benachteiligten Kindern hinsichtlich ihrer Bildungsbeteiligung; nationale und internationale Studien (vgl. nur Blossfeld & Shavit, 1993) haben dies aus unterschiedlichen Perspektiven erforscht und bestätigt. Neben den oben angesprochenen Reformbemühungen ließen sich weitere Stichworte nennen wie *ethnic monitoring*, *vouchers* (Bildungsgutscheine), *Charter Schools*, oder *Community Control*. Wie diese Reformen auch kategorisiert und zugeordnet werden: Mit keiner wurde der Durchbruch erzielt, keine erwies sich als *Magic Bullet*, als Wundermittel.

## 6. ‚Ergebnisorientierte Intervention‘ im Kontext von NCLB

Die erste Zustimmung zu NCLB überformte zunächst die unterschiedlichen Interessenlagen. Dies ist u.a. auf den Dissens geradezu verbotenen genialisch einfachen Titel zurückzuführen, der zudem deutlich von der patriotischen Grundstimmung direkt nach *Nine Eleven* gekennzeichnet ist. Ebenso spielte die Logik der ergebnisorientierten Steuerung eine Rolle, mittels der die Intervention erfolgen soll. Zwar werden Bildungsstandards in den USA seit langem konzeptioniert und implementiert, neu ist an den Debatten, dass Bildungsstandards zum Dreh- und Angelpunkt der aktuellen Bildungsreformwelle auf Bundesebene und nicht nur auf der Ebene einzelner Staaten avancierten. Diese so genannte standard-basierte Reform geht zurück auf unterschiedliche, aber miteinander verbundene Entwicklungen: das Interesse der Gouverneure der Bundesstaaten an der wirtschaftlichen Wettbewerbsfähigkeit ihrer Staaten (National Governors Association, 1991) und der vom Kongress der Vereinigten Staaten eingerichtete nationale Ausschuss zu Standards und Assessment (National Council on Education Standards and Testing, 1992) sowie die nationalen Ausschüsse zu Bildungszielen und zum Berichtswesen (1991–1993). Diese Initiativen gingen in die Erneuerungen von ESEA unter den Vorgängeradministrationen ein, so in Präsident Bush Seniors *America 2000* (1991) und in Präsident Clintons *Goals 2000* (1993). Unter dem am Modell der *Results Based Reforms* (RBR), der ergebnisorientierten Reformen, ausgerichteten NCLB werden bildungsstandard-basierte Schülerleistungen jedoch zur einzigen Grundlage des Bewertungssystems. Schulen und Bildungsadministration übernehmen die Verantwortung für die Erreichung der Ziele und werden für das Erreichen bzw. Nichterreichen der Ziele belohnt oder bestraft. Die Logik der *Results Based Reform* funktioniert wie folgt: Zunächst ist festzulegen, was gelernt werden soll, daran hat sich der Vermittlungsprozess zu orientieren, die Lernenden werden bei der Erreichung der Ziele unterstützt, der Lernfortschritt wird gemessen, die gemessenen Ergeb-

nisse an das System zurückgemeldet, der Vermittlungsprozess wird gegebenenfalls modifiziert und der Zyklus neu gestartet, so lange, bis das gewünschte Ergebnis erreicht ist (vgl. Baker, 2004). Diese Logik findet sich auch in anderen Bereichen, wie dem *Instructional Systems Design* (ISD) oder liegt – beispielsweise wenn Pilotstudien durchgeführt werden – dem Design empirischer Studien zugrunde. Gemeinsam mit einem weiteren Instrument, der Rechenschaftslegung, gilt RBR als präferierte Interventionsmaßnahme der aktuellen Bildungspolitik.

Vor diesem Hintergrund ist die Kopplung an Sanktionen und Gratifikationen der folgerichtige Schritt, um zu verhindern, dass Rückmeldungen folgenlos verpuffen. Ob dies in Form öffentlicher Anprangerung oder Belobigung einzelner Schulen geschieht oder ob drastischere Maßnahmen ergriffen werden – die Schließung einer Schule bei anhaltendem Verfehlen der Ziele – ist dann Sache der politischen Aushandlung.

Im Fall von NCLB sind folgende Komponenten im Kontext der *Results Based Reform* bedeutsam:

- (a) Die mit den Zwecken verbundenen Nutzen oder Sanktionen,
- (b) Ziele oder Standards,
- (c) die in den Blick genommenen Begünstigten,
- (d) die *benchmarks*, in diesem Falle die intendierte Kompetenzstufe *proficient*,
- (e) die Fortschrittsrate bei der Erreichung der *benchmark*, häufig ausgedrückt in Zielgruppenprozenten, die die Kompetenzstufe erreichen – in der Sprache von NCLB der angemessene jährliche Fortschritt (*Adequate Yearly Progress – AYP*),
- (f) die Inputs,
- (g) die Ressourcen,
- (h) die Verfahren,
- (i) die Maßnahmen,
- (j) die Ergebnisse, die nach bestimmten Maßgaben zu veröffentlichen sind.

Die Ergebnisse setzen eine weitere Ereignisfolge in Gang:

- (a) Die Daten führen zu system-externen Schlussfolgerungen (gut; auf dem richtigen Wege; in bestimmten Bereichen schwach; desaströs),
- (b) die Beurteilung führt zu an die Rechenschaftspflicht gekoppelten Konsequenzen,
- (c) daraus folgt die Adjustierung der Ziele für alle Lernenden oder für bestimmte Lerngruppen,
- (d) daraufhin wird die Reform-Implementation neu gestaltet,
- (e) Ressourcen werden umverteilt,
- (f) der Ertrag des neuen Anlaufs wird in die nächste Schleife eingespeist (nach Baker, 2004, S. 4).

## 7. Kritik an NCLB

Die auf unterschiedlichen Ebenen geäußerte Kritik lässt sich wie folgt zusammenfassen: generelle Unterfinanzierung (vgl. Mathis, 2003); Bestrafung der Schulen mit heterogener Schülerschaft oder allgemeiner: Bestrafung der Schulen, die ohnehin die schlechteren Bedingungen und Ausstattungen aufweisen; stupides *teaching to the test*, also Ausrichtung des Unterrichts auf die Vorbereitung der Tests; Senkung der Standards; (grundsätzlich) eine verfehlte Fokussierung auf die Testergebnisse zu Lasten der professionellen Kompetenzen der Lehrer und der pädagogischen Dimension des Unterrichts. Kritisiert wird aber vor allem, dass letztlich die schwachen und die ‚Risikoschüler‘ bestraft werden (vgl. etwa Meier & Wood, 2004; Haney et al., 2004 zum Anstieg der Schulabbrecherquote durch *high stakes testing*; Mathis, 2003 sowie die breiten Diskussionen in erziehungswissenschaftlichen Fachzeitschriften wie *Education Week* und *Phi Delta Kappan*). Es ist zu vermuten, dass sich diese Problemlage (in Rede stehen hier nicht nur die ‚dropouts‘, sondern auch die ‚pushouts‘ und die Klassenwiederholer) nach der Einführung von NCLB nicht wesentlich verändern wird (Darling-Hammond, 2004, S. 18–24; Wood, 2004, S. 36–38), obwohl das Gesetz im Rahmen der Rechenschaftspflicht der Schulen explizit die Angabe der High-School-Abschlussquoten einfordert (zur Problematik der Kalkulation von Abschlussquoten vgl. Miao & Haney, 2004). Es ist nicht unplausibel anzunehmen, dass das Spannungsverhältnis zwischen eingeforderter Schülerleistung und Erhöhung der Abschlussquoten dazu führen wird, dass sich bereits beobachtbare Verdrängungsprozesse eher intensivieren als abnehmen werden. Schließlich haben die Schulen ein Interesse daran, leistungsschwache Schülerinnen und Schüler nicht aufzunehmen oder sie vor Beginn der jährlichen Testperiode loszuwerden, damit sie das Leistungsniveau nicht nach unten drücken.

Diese Befunde provozieren weitere Fragen, wie die nach der Notwendigkeit interner Differenzierung: Sollten ‚Problemschulen‘ den gleichen Regeln unterworfen werden wie privilegierte? Sollten die Testergebnisse je nach Besonderheit der Schule differenziert bewertet werden?

Der Widerstand gegen NCLB ist erheblich und zahlreiche Staaten wehren sich gegen bestimmte Teile oder gegen das Programm insgesamt, z.B. Arizona, Hawaii, New Mexiko und Vermont (Meier & Wood, 2004).

Insgesamt geht ein großer Teil der Kritik zurück auf die Folgen und Nebenwirkungen des Eingriffs von Außen, des Belohnungs- bzw. Bestrafungskatalogs, der mit dem Leistungserfolg der Lernenden verbunden ist und der auf die Schulen bzw. die Schülerinnen und Schüler zurückgerechnet wird. Allerdings besitzt die Kopplung der Testergebnisse an Anreizsysteme zunächst eine hohe Plausibilität, um Systemmodifikationen hervorzurufen, denen unterstellt wird, dass sie besonders benachteiligten Schülergruppen zugute kämen.

Die Kritik weist zudem eine immanente Dimension auf, die sich auf die Konstruktionslogik der Intervention bezieht. Die angeführten Komponenten von NCLB setzen ein Passungsverhältnis, präziser ein Ausrichten oder ‚Ausfluchten‘ (*alignment*) der Bestandteile voraus: Inhalte, Unterrichtsmaterialien, Lehrerbildung, Didaktik und Kompetenzen müssen aufeinander abgestimmt sein. Wird dies nicht vorausgesetzt, so hängt nichts mit nichts zusammen: Lernziele werden in der Praxis umgesetzt oder nicht, Kinder lernen oder lernen nicht was von ihnen erwartet wird; die Testergebnisse spiegeln Standards wider oder auch nicht (Baker, 2004, S. 5). Ohne ‚Ausfluchten‘ ist der Eingriff Glück oder Magie.

Ein weiterer Einwand zielt auf die Validität von Tests. Diese hängt von zwei Hypothesen ab: dass mittels Tests eruiert werden kann, ob Schüler über hinreichende Kompetenz in einer bestimmten Domäne verfügen, und dass Schüler ihre Kompetenzen steigern können, was sich wiederum durch Tests messen lässt.

Ein Problem der ersten Hypothese ist, dass potentielle Verfälschungen der Testergebnisse bereits in der Testkonstruktion angelegt sein können (vgl. Koretz, Bertenthal & Green, 1999). So stellt etwa die Frage, ob die gewählten Items eines Tests die Domäne hinreichend abbilden, hohe Anforderungen an die Testkonstruktion. Ein weiteres Problem liegt in der Ergebnisvariation alternativer, aber ähnlicher Tests. Die Versagerquote kann auch bei stark korrelierenden Tests signifikant variieren. Das heißt, dass auch bei sehr ähnlich konstruierten Tests für die gleiche Domäne bereits geringe Abweichungen zu sehr unterschiedlichen Testergebnissen führen können. Im Ernstfall kann dies bedeuten, dass bis zu einem Viertel der Schülerinnen und Schüler, die Test A bestanden haben, in Test B durchfallen. Hinzu kommen Faktoren wie die Wahl des Tests, der Zeitpunkt der Durchführung, der Umgang mit besonderen Schülergruppen, beispielsweise mit solchen mit besonderem Förderbedarf (nehmen diese teil oder nicht) oder solche, die die Unterrichtssprache nur unzureichend beherrschen – all dies hat Einfluss auf die Testergebnisse (Koretz, 2002) und macht skeptisch gegenüber einem einfachen Rückschluss von Testergebnis auf Schülerleistung. Weiter wurde in dem Zusammenhang ‚Testergebnis = Schülerleistung‘ das Phänomen inflationärer Testzugewinne untersucht (vgl. Shephard, 1990). In den 1980er Jahren beobachtete man einen merkwürdigen statistischen Effekt: Nach den Rückmeldungen der Mehrheit der Schulbezirke in allen amerikanischen Bundesstaaten lag jeder Bundesstaat in den national genormten standardisierten Tests über dem nationalen Durchschnitt – dies ist logisch aber unmöglich. Das Phänomen wird auch unter dem Stichwort ‚Lake Wobegon Phänomen‘ diskutiert, nach dem von Garrison Keillor erfundenen Ort im Herzen Amerikas. In Lake Wobegon sind alle Kinder überdurchschnittlich, alle Frauen stark und alle Männer gutaussehend (vgl. Cannell, 1987; zur Volatilität der Testergebnisse auf der Ebene der einzelnen Bundesstaaten auch Bolon, 2001; Kane & Staiger, 2001). Das Problem besteht ungeachtet von Fortschritten der Testentwicklung. So entwickelte z.B. das *Kentucky Instructional Results Information System* (KIRIS) elaborierte, ‚intel-

ligente‘ Tests. Auch KIRIS sah ambitionierte Ergebnisgewinne über einen Zeitraum von 20 Jahren vor, aber bereits im ersten Jahr kam es zu enormen Verbesserungen. Die statistischen Effekte provozierten Untersuchungen, mit ernüchternden Befunden: Die Verbesserungen gehen nicht auf tatsächliche Lernfortschritte, sondern auf Effekte der Testeinführung zurück (Koretz, 2002, S. 761 f.). Ähnlich Kritisches gilt für das so genannte *Texas Miracle*: Die Einführung des *Texas Assessment of Academic Skills* (TAAS) in den frühen 1990er Jahren schien ein veritables Bildungswunder zu induzieren: Die Schülerleistung stieg signifikant an, bei gleichzeitiger drastischer Senkung der Schulabbrecherquote. Eine ausführliche Analyse (Haney, 2000) relativiert die Erfolge stark. So nahm eine signifikant hohe Rate von Schülern nicht an den Tests teil, die Schüler wurden unter der Rubrik *special education* geführt, also aus dem Regelschulsystem ‚ausgelagert‘. NCLB orientierte sich, dies sei beiläufig erwähnt, am Modell Texas, d.h., die unhinterfragten Erfolgsmeldungen aus Texas und die dort ‚erprobte‘ Interventionsstrategie dienten als Modell für das Bundesgesetz.

Eine weitere Hypothese betrifft den Zusammenhang von Tests und Anreizen. Die Kritik richtet sich gegen die Annahme, dass Tests den notwendigen Ansporn bereitstellen, um Lehrende, Lernende oder die Bildungsadministration zu einer Leistungsverbesserung zu veranlassen. Diese Annahme ist mit einer Reihe weiterer Probleme und Fragen behaftet. Sie betreffen die Folgen der Nutzung von Testergebnissen. In der Optik ergebnisorientierter Bildungsstandards erscheint der Output als eine Art Platzhalter von Bildungsqualität. Wird die Leistungsmessung der Schüler den Lehrern zugerechnet, so wird unterstellt, diese hätten einen wesentlichen Anteil am Erfolg oder Misserfolg der Lernenden. Die Prämisse des *high stakes testing* wie der ergebnisorientierten Bildungsstandard-Diskussion besteht somit darin, die Leistung der Schulen oder des Lehrers indirekt zu ermitteln, indem von der Schülerleistung auf den Lehrer, die Organisation (Schule) oder das Bildungssystem geschlossen wird. Diese Zurechnungen lassen unberücksichtigt, dass ein Großteil der Varianz in den Testergebnissen auf Faktoren außerhalb des Klassenzimmers zurückzuführen ist: auf die ethnische Herkunft der Schülerschaft, auf den Bildungsstand und das Einkommen der Eltern, das familiäre Umfeld usw. Unberücksichtigt bleibt auch, dass die meisten Lehrer die Schülerinnen und Schüler nicht kontinuierlich, sondern nur punktuell in ihrer Schullaufbahn begleiten. Zudem ist das konkrete Unterrichtsgeschehen für die Forschung noch weitgehend eine Art *Black Box*.

Insgesamt ist aus diesen Kontroversen auch für die hiesige Auseinandersetzung diskussionswürdig: 1. Das Problem, die Testresultate als direktes und unzweideutiges Maß der Schülerleistung zu interpretieren und 2. eine stark verkürzte und vereindeutigende Interpretation von Lernzuwachs.

## 8. Resümee

Ausgangspunkt der Beschäftigung mit dem ‚Fall‘ USA war die Rezeption der amerikanischen Bildungsstandard-Debatte in einem zentralen Dokument der aktuellen Auseinandersetzung um die Einführung ergebnisorientierter Standards. Die Rezeption ausländischer Entwicklungen und Erfahrungen in einem generellen Modus des ‚Auslands‘ als Argument entspricht einem typischen Muster. Dies liegt nahe, weil bestimmte Begründungen, die für die Einführung ergebnisorientierter Bildungsstandards sprechen, durch Rekurs auf die internationale Erfahrung gestützt und bekräftigt werden sollen.

Generell lässt sich, bezogen auf die intentionale Ebene, auf die allgemeine Ausrichtung und Zielsetzungen der Bildungsstandard-Diskussion, zumal in Verbindung mit Rechenschaftspflicht und Transparenz, eigentlich nur rhetorisch fragen: *What could be bad?* Im Sinne von: Was könnte daran schlecht sein? Sowohl die allgemeine Diskussion um Standards und Rechenschaftspflicht der Schulen als auch die besondere Variante von NCLB versprechen eine angemessene Antwort auf die Forderung nach bestmöglicher Vorbereitung und Förderung der nachwachsenden Generation als hohes gesellschaftliches Gut zu geben. Dazu einen Beitrag zu leisten und sicher zu stellen, dass die nächste Generation auf zukünftige gesellschaftliche Anforderungen möglichst gut vorbereitet ist, markiert folglich einen zentralen gemeinsamen Bezugspunkt der internationalen Bildungsstandard-Diskussion. Vor diesem grundsätzlichen Hintergrund ist dann auch die große Zustimmung zu verstehen, die NCLB zunächst von unterschiedlichsten Interessenvertretern zuteil wurde. Die Konzeption und Intention von Bildungsstandards sind also nicht an sich problematisch. Die mit den ergebnisorientierten Bildungsstandards verbundenen Schwierigkeiten liegen, auch ohne *high stakes assessments*, überwiegend in deren Nutzung. Diese Nutzung kann in Deutschland und in den USA vor dem Hintergrund bestehender Bildungsungleichheit, angesichts von Egalität und Qualität betrachtet werden. Dies macht beide Länder vergleichbar, bei aller Unterschiedlichkeit der Ursachen dieses Effekts, und dies unterscheidet sie von den skandinavischen Staaten. Diese Gemeinsamkeit lässt es sinnvoll erscheinen, die amerikanische Erfahrung eingehender zu betrachten und dies als Folie zu nutzen, um abschließend Fragen zu generieren und einige Bedenken zu formulieren.

Mit dem erklärten Ziel, die Bildungsbeteiligung benachteiligter Schülergruppen zu erhöhen, wurden in den USA Standards an Tests und Testergebnisse an ein Anreizsystem gebunden. In Deutschland ist bisher nicht an ein solches Anreizsystem gedacht. Verhindert dies die negativen Effekte? Auch ohne *naming and shaming*, ohne Liga-Tabellen und ohne direkte materielle Konsequenzen in Form von Ressourcenzuwendung bzw. Schließung von Schulen ist zu erwarten, dass auch in Deutschland Schulen die Testergebnisse der Schülerleistungen nutzen werden. Gutes Abschneiden ermöglicht Schulen, Schülerströme zu beeinflussen (vgl. hierzu Radtke, 2004) und zur Leistungshomogenität beizutragen. Dies kann die Polarisierung des Bildungssystems verstärken. Andererseits stellt sich in Abwesenheit eines steuernden Anreizsystems die

Frage, wie die mit der Einführung ergebnisorientierter Bildungsstandards beabsichtigten Wirkungen gesichert werden können. Wem gegenüber sind die Schulen zur Rechenschaft verpflichtet? Welche Konsequenzen hat es, wenn die gemessenen Leistungen unbefriedigend sind? Die nächste Frage betrifft die Aussagekraft ergebnisorientierter Bildungsstandards in einem *unaligned*, einem ‚unausgerichteten‘ System, in dem Rahmenrichtlinien und Lehrpläne bis auf weiteres den Unterricht bestimmen und die standardbezogenen Tests (additiv) hinzukommen. Dem soll abgeholfen werden, indem die Bildungsstandards an ein Kerncurriculum gekoppelt werden (vgl. Tenorth, 2004). Dies wäre in der Tat ein wichtiger Schritt auf dem Weg der Ausrichtung des Bildungssystems. Bedeutet dies mittel- und langfristig, dass den Testergebnissen ein größeres Gewicht zukommt, so ist man bei der Frage des Umgangs mit leistungsschwachen und anderen ‚Problemschülern‘ angelangt. In den USA zeigt sich, dass Schulen die Verantwortungszurechnung unterlaufen, indem sie leistungsschwache Schüler möglichst nicht aufnehmen und andernfalls vor Beginn der Testperioden loswerden. Solange nämlich bei der Einführung von Standards und Tests nicht gleichzeitig darüber gewacht wird, dass alle schulpflichtigen Schülerinnen und Schüler berücksichtigt werden, besteht die Gefahr, die ein amerikanischer Autor in Bezug auf NCLB so auf den Punkt brachte: Es ist, als führe ein Krankenhaus Buch über die erfolgreich behandelten Patienten, aber nicht über diejenigen, welche die Behandlung nicht überlebt haben. Dies kann auch hierzulande eine Erhöhung der Schulabbrecherquote und der Ausschulungsrate bewirken. Die Stichworte Schulversagen und Schulabsentismus berühren ein Feld, in dem ohnehin schwierige statistische Verhältnisse herrschen.

Die amerikanische Erfahrung zeigt, dass die Wirkung von Bildungsstandards entscheidend davon abhängt, in welches Gefüge von Instrumenten und Maßnahmen sie eingebettet sind. So wäre zu fragen, ob alleine ergebnisorientierte Bildungsstandards angewendet werden sollen oder nicht auch auf die Lehrerprofessionalität bezogene Standards sowie prozessorientierte Standards hinzutreten sollten. Letztere zielen auf die *Black Box* Unterricht ab (Black & Dylan, 1998; Stiggins, 2002; Stigler, Gonzales, Kawanaka, Knoll & Serrano, 1999), indem sie nicht nur Lernergebnisse messen, sondern auch den Weg in den Blick nehmen, der zu den Ergebnissen führt, also neben *assessments of learning standards* auch *assessments for learning standards* vornehmen. Weiter ließe sich fragen, wie Standards und ihre Messung im Verhältnis zu anderen Entscheidungsinstrumenten und professionellen Kompetenzen gewichtet werden sollen und wie über diese Gewichtung entschieden wird.

## Anmerkungen

1. Mit diesem Titel nehme ich sprachliche Anleihe an Bernd Zymeks Schrift: *Das Ausland als Argument in der pädagogischen Reformdiskussion* (1975).
2. Ich verwende die Bezeichnung *Results-Based Reform*, um Missverständlichkeiten des Sprachgebrauchs von Output zu vermeiden. Herrmann (2003) stellt hierzu klar, dass Input-Output Rechnungen auf an Effizienz orientierten quantitativen Kosten-Nutzen-Kalkülen bezogen sind. Zu unterscheiden sind Effektivität und Effizienz des Bildungssystems und parallel dazu: output und outcome, also das „*quantitativ zu bewertende* Resultat bzw. Ergebnis eines *Prozesses*“ (ebd., S. 627).
3. Nach PISA entstand eine Diskussion darüber, ob auch in der High School häufiger getestet werden solle. Ab 2007 ist vorgesehen, auch naturwissenschaftliche Fächer in die Tests einzubeziehen.
4. Charter Schools sind frei wählbare öffentliche Schulen, die sich durch größere Freiheiten von anderen öffentlichen Schulen unterscheiden. Der Charter, der Freibrief, begründet die Etablierung. Sie ist ein Leistungsvertrag, in dem Mission, Programm, Ziele und Zielgruppe, Methoden und Assessment-Strategien festgelegt werden. Die hat eine begrenzte Laufzeit von meistens drei bis fünf Jahre und muss dann erneuert werden. Voraussetzungen sind eine Verbesserung der Schülerleistungen und die Erfüllung der Vertragspflichten. Rechenschaftspflichtig sind Charter Schools gegenüber ihren Vertragspartnern, dem Bundesstaat oder einer lokalen Schulbehörde. Leitkonzepte sind Schulautonomie und Rechenschaftspflicht.
5. Das Dokument ist im Internet verfügbar unter <http://www.ed.gov/print/policy/elsec/leg/ede02/beginning.html>.
6. Vgl. die Rolle des *National Center for Education Statistics* (NCES), das in den USA federführend war für die Durchführung von PISA, des *National Assessment on Educational Progress* (NAEP) auch bekannt als *the Nation's report card*, sowie der *National Assessment Governing Board* (NAGB).

## Literatur

- Adick, C. (1992). *Die Universalisierung der modernen Schule*. Münster: Schöningh.
- Airasian, P. (1987). State mandated testing and educational reform: Context and consequences. *American Journal of Education*, 95 (3), 393–412.
- Baker, E. L. (2004). Aligning curriculum, standards, and assessments: Fulfilling the promise of school reform. Verfügbar unter: <http://www.cresst96.cse.ucla.edu/reports/r645.pdf>. [Forthcoming as: Baker, E. L. (in press). Aligning curriculum, standards, and assessments: Fulfilling the promise of school reform. In C. Dwyer (Ed.), *Measurement and research in the accountability era*. Mahwah, NJ: Earlbaum].
- Baratz-Snowden, J. (1993). Opportunity to learn: Implications for professional development. *Journal of Negro Education*, 62, 311–323.
- Berger, P. (2005). Deutsche Ungleichheiten – eine Skizze. *Aus Politik und Zeitgeschichte*, 37, 7–16.
- Berliner, D. (2004). If the underlying premise for No Child Left Behind is false, how can that Act solve our problems? The Iowa Academy of Education. Occasional Research Paper No. 6. [Auch In K. Goodman, P. Shannon & Y. Goodman (Eds.), *Saving Our Schools: The Case for Public Education in America*. Oakland, CA: RDR Books in press.]
- Black, P. & Dylan, W. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80 (2), 139–148. Verfügbar unter: <http://www.pdkintl.org/kappan/kbla9810.htm>.

- Blossfeld, H.-P. & Shavit, Y. (1993). Dauerhafte Ungleichheiten. Zur Veränderung des Einflusses der sozialen Herkunft auf die Bildungschancen in dreizehn industrialisierten Ländern. *Zeitschrift für Pädagogik*, 39 (1), 25–52.
- Bolon, C. (2001). Significance of test-based ratings for metropolitan Boston schools. *Education Policy Analysis Archives*, 9 (42). Verfügbar unter: <http://epaa.asu.edu/epaa/v9n42>.
- Cannell, J. (1987). *Nationally normed elementary achievement testing in America's public schools: How all fifty states are above the national average*. Daniels, W.VA: Friends of Education.
- Darling-Hammond, L. (2004). From “Separate but Equal” to “No Child Left Behind”: The collision of new standards and old inequalities. In D. Meier & G. Wood (Eds.), *Many children left behind. How the No Child Left Behind Act is damaging our children and our schools* (pp. 3–32). Boston: Beacon Press.
- The Elementary and Secondary Education Act (*No Child Left Behind*) 2001. (2002). Verfügbar unter: <http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- GEW (Gewerkschaft Erziehung und Wissenschaft). (2003). *Nationale Bildungsstandards. Wundermittel oder Teufelszeug? Funktionen, Hintergründe und Positionen der GEW*. Frankfurt a.M.: Spitzdruck Darmstadt.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8 (41). Verfügbar unter: <http://epaa.asu.edu/epaa/v8n41>.
- Haney, W. (2002). Lake Wobeguaranteed: Misuse of test scores in Massachusetts, Part I. *Education Policy Analysis Archive*, 10 (24), 27 pp. Verfügbar unter: <http://epaa.asu.edu/epaa/v10n24>.
- Haney, W., Madaus, G., Abrams, L., Wheelock, A., Miao, J. & Gruia, I. (2004). *The education pipeline in the United States, 1970–2000*. Education Pipeline Project. National Board on Educational Testing and Public Policy. Center for the Study of Testing, Evaluation, and Educational Policy. Boston: Boston College.
- Hermann, U. (2003). „Bildungsstandards“ – Erwartungen und Bedingungen, Grenzen und Chancen. *Zeitschrift für Pädagogik*, 49 (5), 625–639.
- Kane, T. & Staiger, D. (2001). *Volatility in School Test Scores: Implications for Test-based Accountability Systems*. Working paper of the National Bureau of Economic Research. Verfügbar unter: <http://www.nber.org/papers/w8156>.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H.-J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Berlin: Bundesministerium für Bildung und Forschung.
- Koretz, D. (1992). State and national assessment. In M. Alkin (Ed.), *Encyclopedia of educational research* (pp. 1262–1267). Washington, D.C.: American Educational Research Association.
- Koretz, D. M. (2002). Limitations in the use of achievement tests as measures of educator's productivity. *The Journal of Human Resources*, 37, 752–777.
- Koretz, D., Linn, R., Dunbar, S. & Shephard, L. (1991). *The effects of high-stakes testing on achievement: Preliminary findings about generalizations across tests*. Beitrag anlässlich des Symposiums der American Educational Research Association. Chicago, Ill.
- Koretz, D., Bertenthal, M. & Green, B. (Eds.). (1999). *Embedding questions: The pursuit of a common measure in uncommon tests*. Committee on Embedding Common Test Items in State and District Assessments. National Research Council. Washington, D.C.: National Academies Press.
- Mathis, W. (2003). No Child Left Behind: Costs and benefits. *Phi Delta Kappan*, 84 (9), 679–686.
- Meier, D. & Wood, G. (Eds.). (2004). *Many children left behind. How the No Child Left Behind Act is damaging our children and our schools*. Boston: Beacon Press.
- Miao, J. & Haney, W. (2004). High school graduation rates: Alternative methods and implications. *Education Policy Analysis*, 12 (55), 70. Verfügbar unter: <http://epaa.asu.edu/epaa/v12n55>.

- National Council on Education Standards and Testing (1992). *Raising standards for American education*. Washington, D.C.: National Council on Education Standards and Testing.
- National Commission on Excellence in Education (1983). *A nation at risk*. Washington, D.C.: U.S. Government Printing Office.
- National Governors Association (1991). *Results in education, 1989: The governors' 1991 report on education* (ERIC Document Reproduction Service No. ED 313338). Washington, D.C.: National Governors Association.
- Radtke, F.-O. (2004). Die Illusion der meritokratischen Schule. Lokale Konstellationen der Produktion von Ungleichheit im Erziehungssystem. In K. J. Bade & M. Bommers (Hrsg.), *Migration – Integration – Bildung. Grundfragen und Problembereiche* (IMIS-Beiträge, Bd. 23) (S. 143–178). Osnabrück: Institut für Migrationsforschung und Interkulturelle Studien.
- Resnick, D. (1982). History of educational testing. In A. Wigdor & W. Garner (Eds.), *Ability testing: Uses, consequences, and controversies, part II* (pp. 173–194). Washington, D.C.: National Academy Press.
- Shephard, L. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement. Issues and Practice*, 9 (3), 15–22.
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83 (10), 758–765.
- Stigler, J., Gonzales, P., Kawanaka, T., Knoll, S. & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States*. Washington, D.C.: National Center for Education Statistics.
- Tenorth, H.-E. (2004). Stichwort: „Grundbildung“ und „Basiskompetenzen“. Herkunft, Bedeutung und Probleme im Kontext allgemeiner Bildung. *Zeitschrift für Erziehungswissenschaft*, 7 (2), 169–182.
- U.S. Department of Education (2003). *Standards and assessments: Non-regulatory guidance*. Washington, D.C.: U.S. Department of Education.
- Van Zanten, A. (2005). New Modes of Reproducing Social Inequality in Education: the changing role of parents, teachers, schools, and education policies. *European Educational Research Journal*, 4 (3), 155–169.
- Wood, G. (2004). A view from the field: NCLB's effects on classroom and schools. In D. Meier & G. Wood (Eds.), *Many children left behind. How the No Child Left Behind Act is damaging our children and our schools* (pp. 33–50). Boston: Beacon Press.
- Zymek, B. (1975). *Das Ausland als Argument in der pädagogischen Reformdiskussion. Politische Selbstrechtfertigung, Auslandspropaganda, internationale Verständigung und Ansätze einer Vergleichenden Erziehungswissenschaft in der internationalen Berichterstattung deutscher pädagogischer Zeitschriften 1871–1952*. Ratingen: Henn.