

Klieme, Eckhard

## Begründung, Implementation und Wirkung von Bildungsstandards: Aktuelle Diskussionslinien und empirische Befunde. Einführung in den Thementeil

*Zeitschrift für Pädagogik 50 (2004) 5, S. 625-634*



Quellenangabe/ Reference:

Klieme, Eckhard: Begründung, Implementation und Wirkung von Bildungsstandards: Aktuelle Diskussionslinien und empirische Befunde. Einführung in den Thementeil - In: Zeitschrift für Pädagogik 50 (2004) 5, S. 625-634 - URN: urn:nbn:de:0111-opus-48317 - DOI: 10.25656/01:4831

<https://nbn-resolving.org/urn:nbn:de:0111-opus-48317>

<https://doi.org/10.25656/01:4831>

in Kooperation mit / in cooperation with:

# BELTZ

<http://www.beltz.de>

### Nutzungsbedingungen

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

### Terms of use

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

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

### Kontakt / Contact:

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

Mitglied der

  
Leibniz-Gemeinschaft

## **Inhaltsverzeichnis**

### *Thementeil: Bildungsstandards*

*Eckhard Klieme*

Begründung, Implementation und Wirkungen von Bildungsstandards: Aktuelle Diskussionslinien und empirische Befunde Einführung in den Thementeil .....	625
--	-----

*Kristina Reiss*

Bildungsstandards und die Rolle der Fachdidaktik am Beispiel der Mathematik .....	635
--	-----

*Heinz-Elmar Tenorth*

Bildungsstandards und Kerncurriculum – Systematischer Kontext, bildungstheoretische Probleme .....	650
---	-----

*Jürgen Rost*

Psychometrische Modelle zur Überprüfung von Bildungsstandards anhand von Kompetenzmodellen .....	662
---	-----

*Olaf Köller/Jürgen Baumert/Kai S. Cortina/Ulrich Trautwein/Rainer Watermann*

Öffnung von Bildungswegen in der Sekundarstufe II und die Wahrung von Standards. Analysen am Beispiel der Englischleistungen von Oberstufenschülern an integrierten Gesamtschulen, beruflichen und allgemein bildenden Gymnasien .....	679
---	-----

Linktipps zum Thema Bildungsstandards .....	701
---	-----

### *Allgemeiner Teil*

*Alfred Schäfer*

Alterität: Überlegungen zu Grenzen des Pädagogischen Selbstverständnisses .....	706
---	-----

<i>Maria Fölling-Albers/Andreas Hartinger/Dženana Mörtl-Hafizović</i> Situieretes Lernen in der Lehrerbildung .....	727
<i>Peter Jörg Alexander/Matthias Pilz</i> Die Frage der Gleichwertigkeit von allgemeiner und beruflicher Bildung in Japan und Deutschland im Vergleich .....	748
 <i>Besprechungen</i>	
<i>Daniel Gredig/Elena Wilhelm</i> Erika Steinert/Gisela Thiele: Sozialarbeitsforschung für Studium und Praxis. Einführung in die qualitativen und quantitativen Methoden Hanne Schaffer: Empirische Sozialforschung für die Soziale Arbeit. Eine Einführung Hans-Uwe Otto/Gertrud Oelerich/Heinz G. Micheel (Hrsg.): Empirische Forschung und Soziale Arbeit. Ein Lehr- und Arbeitsbuch Cornelia Schweppe (Hrsg.): Qualitative Forschung in der Sozialpädagogik .....	770
<i>Cristina Allemann-Ghionda</i> Martina Weber: Heterogenität im Schulalltag. Konstruktion ethnischer und geschlecht- licher Unterschiede .....	779
<i>Andreas Krapp</i> Monique Boekaerts/Paul R. Pintrich/Moshe Zeidner (Eds.): Handbook of Self-Regulation .....	781
<i>Peter Martin Roeder</i> Kurt A. Heller (Hrsg.): Begabtenförderung im Gymnasium. Ergebnisse einer zehnjährigen Längsschnittstudie .....	783
 <i>Dokumentation</i>	
Pädagogische Neuerscheinungen .....	788

Eckhard Klieme

# **Begründung, Implementation und Wirkung von Bildungsstandards: Aktuelle Diskussionslinien und empirische Befunde**

*Einführung in den Thementeil*

## **1. Standard-Begriffe und deren Wurzeln im bildungspolitischen und pädagogischen Diskurs**

Die aktuelle Diskussion um Standards für die pädagogische Arbeit in Schulen knüpft an sehr unterschiedliche Diskurse an: (a) den traditionellen Diskurs um Maßstäbe für verantwortliches professionelles Handeln, (b) die seit Ende der 1980er-Jahre international zu beobachtende Tendenz zu stärkerer Lenkung und Kontrolle der Qualitätsentwicklung in Schulsystemen, (c) die durch empirische Studien wie TIMSS, BIJU, PISA und IGLU neuerdings wieder aufgeflamnte Debatte über die große Variation von Anforderungsniveaus und Bewertungsmaßstäben zwischen Klassen, Schulen, Schulformen und Ländern (Köller/Baumert/Schnabel 1999; Baumert/Trautwein/Artelt 2003). In die Vorschläge zur Entwicklung und „Implementation“ von Bildungsstandards, wie sie gegenwärtig in Deutschland, aber beispielsweise auch in der Schweiz vorgelegt werden, gehen darüber hinaus Konzepte verschiedener Forschungsrichtungen der Erziehungswissenschaft und ihrer Nachbardisziplinen ein: (d) Strategien der Lehrplanarbeit und insbesondere der Curriculumreformen der 1960er und 1970er-Jahre, deren Grundfragen jetzt teilweise in neuer Einkleidung verhandelt werden, (e) Ansätze der fachdidaktischen, lernpsychologischen und diagnostischen Forschung zur Konzeptualisierung, Messung und Förderung von Kompetenzen, sowie (f) Konzepte der Schulentwicklungsforschung.

Dieser einleitende Beitrag soll in seinem ersten Teil den Diskussionsrahmen zumindest ansatzweise aufspannen und die Beiträge des vorliegenden Thementils einordnen. Die vier nachfolgenden Texte beleuchten Grundlagen der Einführung, Legitimation, Operationalisierung und Nutzung von Bildungsstandards aus Sicht verschiedener Teildisziplinen der Erziehungswissenschaft: der fachdidaktischen Lehr-Lern-Forschung, der Bildungstheorie, der pädagogischen Diagnostik und der empirischen Schulforschung. Die Beiträge werden ergänzt durch eine Zusammenstellung einschlägiger Internet-Links, die Zugang zu aktuellen Reformdokumenten und Forschungsergebnissen eröffnen. Ein zweiter Abschnitt dieser Einleitung soll – basierend auf einer Sichtung neuerer Literatur aus den USA – erste empirisch fundierte Antworten auf die Frage geben, welche Auswirkungen – auch im Sinne nicht intendierter Folgen – die Einführung von Bildungsstandards in Schulen haben kann.

Seit jeher verbindet sich der Begriff des „Standards“ mit der Frage nach den Normen professionellen Handelns. Professionen definieren sich nicht zuletzt durch die teils expliziten, teils impliziten Erwartungen an das verantwortliche Handeln ihrer Angehörigen, die durch wissenschaftliche Ausbildung, durch die berufliche Sozialisation und den kollegialen Diskurs vermittelt und immer wieder eingefordert werden. Gerade weil professionelles Handeln – bei Lehrkräften wie in anderen Professionen – durch einen hohen Grad an Komplexität, Nicht-Planbarkeit und Eigenverantwortung gekennzeichnet ist, kommt Standards im Sinne tradierter Erwartungen eine große Bedeutung zu. Professionelles Handeln kann und soll durch solche Normen nicht „standardisiert“, technologisch ausgerichtet werden, aber die Qualität des Handelns und seine Konsequenzen werden rational überprüfbar. Besonders herausgefordert sehen sich Professionelle in pädagogischen Berufen nun durch neuere Versuche, Handlungserwartungen explizit und systematisch auszuarbeiten, als Forderungen an die Handlungskompetenz von Lehrkräften zu formulieren und normativ zur Grundlage der Lehrerbildung zu erklären (Oser/Oelkers 2001; Terhart 2002). Diese Tendenz ist vor allem für die deutschsprachigen Länder neu, in denen die Autonomie professionellen Handelns traditionell stärker gewichtet wird als die Sicherung von Qualitätsnormen, während beispielsweise in den angelsächsischen Ländern und in den Niederlanden die standard-bezogene Zertifizierung von Lehrpersonen oder die Formulierung expliziter Standards für „guten Unterricht“ durch Schulinspektionen schon seit längerem Realität sind (vgl. Döbert/Klieme/Sroka 2004). Aus der Perspektive einer Prozessanalyse von Bildungsinstitutionen erfassen solche Standards Aspekte des „Inputs“ der Schulen (hier: der Qualifikationen des pädagogischen Personals) sowie der Lehr-Lern-Prozesse.

Dass überhaupt nach der Qualität pädagogischer Arbeit gefragt und hierfür – über professionsinterne Normen hinaus – Kontrollsysteme etabliert werden, ist ein Ergebnis von Veränderungen in der politischen Strategie zur „Steuerung“ von Bildungssystemen. Ökonomische Argumente (u.a. die Theorie des Humankapitals), neue Konzepte des Verwaltungshandelns („new public management“, Einführung von Kosten-Leistungskalkülen und marktanalogen Strukturen), wiederholte Hinweise auf die Schwächen der Schulsysteme (z.B. in internationalen Vergleichsstudien der IEA oder in der weit über die USA hinaus rezipierten Expertise „A Nation at Risk“ von 1983) sowie Befunde der Schuleffektivitätsforschung verstärkten zunächst in den USA, aber auch in anderen westlichen Industrienationen die Bemühungen, den „Output“ von Bildungseinrichtungen stärker zu kontrollieren und durch Veränderung der Inputs und Prozesse zielorientiert zu beeinflussen. Die OECD fasste diese Tendenzen in ihrem Bericht „Schulen und Qualität“ (deutsch 1991, Original 1989) zusammen, in dem das Konzept der Standards eine zentrale Bedeutung hatte. Standards waren hier auf Lernergebnisse bezogen – sei es empirisch verstanden als Niveau der tatsächlich erreichten Lernergebnisse (so bis heute in Großbritannien), sei es normativ als Erwartung an zu erreichende Ergebnisse (so in den USA und neuerdings in Deutschland). Jedenfalls wurde der Erhebung und Rückmeldung von Lernergebnissen eine zentrale Funktion bei der Steuerung (eigentlich, da über *feedback* vermittelt: Regelung) des Schulsystems beigemessen, die wiederum Grundlage für die Weiterentwicklung der Bildungsqualität sei. In den USA und teilweise

in Kanada entwickelten sich Steuerungssysteme, die staatliche *content standards* bzw. *curriculum standards* (Lernziele) mit Testverfahren koppelten, wobei *performance standards* als Grenzwerte auf einer Testskala festgelegt wurden (Cizek 2001). Diese Standards sind Bezugsgrößen für die *accountability* der Schulen und Schulbezirke, d.h. für Verantwortungsübernahme und Rechenschaftslegung gegenüber Aufsichtsgremien und Öffentlichkeit (Fuhrman/Elmore 2004). In Großbritannien entwickelte sich ein breites System von Schulinspektionen, das auch – aber nicht nur – Daten aus zentralen Prüfungen und regelmäßigen Tests aller Altersstufen verwendet (Matthews 2004). Angesichts einer traditionell hohen Autonomie der Einzelschulen und Schuldistrikte in den angelsächsischen Staaten dienten diese Instrumentarien einer regelmäßigen, einheitlichen und objektiven Bewertung des erreichten Stands der Qualitätsentwicklung. Im Vordergrund stand und steht dabei der „output“ der Schulen, aber auch Input und Prozesse, z.B. die Lerngelegenheiten der Schüler werden erfasst und anhand von *opportunity to learn-standards* bewertet.

Für diesen Bewertungsprozess hat sich im Deutschen mittlerweile der Begriff „Qualitätssicherung“ eingebürgert. Es bedurfte jedoch erst der großen repräsentativen Schulleistungsvergleiche der vergangenen Jahre, um einen solchen Prozess – länderintern und länderübergreifend – in Gang zu setzen. Eines der Ergebnisse der PISA-Studie lautete, dass deutsche Schülerinnen und Schüler vor allem im unteren Leistungsbereich deutlich schlechter abschneiden als Schülerinnen und Schüler aus anderen OECD-Staaten. Dies wurde von Bildungsforschern dahingehend interpretiert, dass in Deutschland „Mindeststandards“ nicht ausreichend eingehalten würden. Die Ergänzungsstudie PISA-E, die Schulen der 16 Bundesländer systematisch verglich, wies zudem nach, dass Leistungsniveaus und Benotungsmaßstäbe zwischen den Ländern, aber auch innerhalb der Länder zwischen Schulformen und Einzelschulen, deutlich schwanken.<sup>1</sup> Aus dem Vergleich mit anderen Staaten, die mit output-orientierten Qualitätssicherungssystemen arbeiten, wurde vielfach geschlossen, dass es in Deutschland an verbindlichen Setzungen normativer Standards fehle. So entstand in der bildungspolitischen Diskussion der Ruf nach national einheitlichen Standards, die Art und Niveau von schulischen Anforderungen regeln und prüfbar machen sollten. Eine der wichtigsten Leitlinien der deutschen Bildungspolitik „nach PISA“ könnte man daher folgendermaßen formulieren: *Leistungsstandards im empirischen Sinne* (= gemessene Lernergebnisse) sollen länder-übergreifend auf ein ähnliches Niveau gebracht und zugleich – soweit möglich – angehoben werden, und *Leistungsstandards im diagnostischen Sinne* (= Benotungsmaßstäbe) sollen angeglichen werden, indem *Standards im normativen Sinne* (= Erwartungen (a) an die Lernergebnisse von Schülerinnen und Schülern – „performance standards“ –, (b) an die

1 Klassenspezifisch variierende Benotungsmaßstäbe hatte schon Ingenkamp (1989) festgestellt. Baumert und Mitarbeiter (siehe Köller u.a. 1999) hatten im Zuge ihrer BIJU-Studie unterschiedliche Leistungsnormen in der Oberstufe von Gymnasien und Gesamtschulen in Nordrhein-Westfalen ermittelt; vgl. auch Köller u.a. in diesem Heft.

Unterrichtsinhalte – „content standards“ – oder (c) an den Unterricht selbst – „opportunity-to-learn-standards“ –) verbindlich vereinbart und überprüft werden<sup>2</sup>.

Die Kultusminister der Länder griffen diese Forderung schließlich auf und beschlossen im Juni 2002, binnen kurzem für den mittleren Schulabschluss, den Hauptschulabschluss und den Abschlussjahrgang der Primarschule (4. Klasse) länderübergreifende Standards in zentralen Fächern (Mathematik, Deutsch, erste Fremdsprache, Naturwissenschaften) zu verabschieden, deren Einhaltung durch Vergleichstests überprüft werden sollte. Inzwischen liegen die ersten Dokumente dieser Art vor, die von Lehrplanexperten unter Beratung durch Fachdidaktiker erstellt wurden. Diese Dokumente beschreiben zentrale Kompetenzen, die Schülerinnen und Schüler bis zur jeweiligen Jahrgangsstufe erreicht haben sollen, und bieten erste Ansätze zur Konkretisierung bzw. Operationalisierung dieser Erwartungen durch Beispielaufgaben, wobei nach Möglichkeit mehrere Kompetenzniveaus unterschieden werden. Die KMK-Standards haben prinzipiell den Anspruch, Erwartungen an fachliches Lernen im Kontext allgemeiner Bildungsziele zu spezifizieren; sie werden daher häufig – so auch im vorliegenden Themenheft – als Bildungsstandards bezeichnet. Es handelt sich bei diesen Standards im Wesentlichen um Curricula, die auf Kernbereiche fokussiert und exemplarisch operationalisiert sind. Es kann daher nicht verwundern, dass in der bildungspolitischen und erziehungswissenschaftlichen Debatte viele Themen der früheren Curriculumforschung (Hameyer/Frey/Haft 1983) wieder auflebten, beispielsweise die Frage nach deren bildungstheoretischen Grundlagen, die auch die Legitimation des Kanons der Schulfächer und Lernziele berührt (siehe hierzu Tenorth, in diesem Heft).

Die Ausarbeitung von Bildungsstandards in der KMK, orientiert an internationalen Vorbildern, wird am Beispiel der Mathematik im vorliegenden Heft durch Reiss diskutiert. Ihr Beitrag macht deutlich, dass Standards der fachdidaktischen und pädagogisch-psychologischen Fundierung bedürfen. Dies betrifft insbesondere die Formulierung von so genannten Kompetenzmodellen, die darstellen sollen, in welchen Teildimensionen und Graduierungen („Kompetenzstufen“) Erwartungen an das fachliche Lernen beschrieben und überprüft werden. Es wird deutlich, dass selbst im relativ gut erforschten Bereich mathematischer Kompetenzen viele Fragen der Grundlagenforschung offen sind – insbesondere wenn Kompetenzmodelle nicht nur Leistungsniveaus im Querschnitt, sondern auch Lernentwicklungen abbilden sollen. Die Kompetenzstufen, wie sie

2 Es wäre einer historischen Rekonstruktion wert, nachzuvollziehen, wie sich die hier angedeutete Bedeutungsverschiebung im bildungspolitischen Diskurs durchgesetzt hat. Köller, Baumert und Schnabel sprechen (1999) in ihrem einflussreichen Aufsatz von „Mindeststandards“ und meinen damit die in Gesetzen und Verordnungen festgelegten Kriterien für als „ausreichend“ benotete Leistungen, also Leistungsstandards im diagnostischen Sinne. Als „pragmatischen Indikator“ verwenden sie die mittlere Leistung der mit „4“ benoteten Schüler in einem Fachtest. Aus dem Befund, dass diese Bewertungsmaßstäbe nicht einheitlich sind, folgern sie die Notwendigkeit von „konstruktiven Schulentwicklungsprozessen“, unterstützt durch „administrative, auch standardsichernde Vorgaben“. Aus der Idee solcher „Vorgaben“ ist offenbar das KMK-Konzept normativer Standards hervorgegangen, das in der Expertise von Klieme et al. (2003) aufgegriffen und – nun wieder aus der Perspektive von Diagnostik und Schulentwicklung – diskutiert wurde.

etwa bei PISA vorgelegt wurden, beruhen auf nachträglichen detaillierten Analysen der Anforderungsmerkmale aller verwendeter Testitems. Der Beitrag von Rost verweist auf die Grenzen des PISA-Ansatzes und diskutiert am Beispiel der Naturwissenschaften die Möglichkeiten psychometrisch anspruchsvollerer Modellierungsmethoden; auch hier ergeben sich offene Forschungsfragen.

Die KMK-Standards für den mittleren Bildungsabschluss werden seit Sommer 2004 in allen Bundesländern eingeführt („implementiert“), häufig begleitet durch Fortbildungsaktivitäten, die Landesinstitute, Schulleitungen und Fachkollegien einbeziehen. Zugleich arbeiten Experten an der Entwicklung von Testaufgaben, die Standards konkretisieren und messbar machen. Erstmals sollen im Jahr 2006 – angekoppelt an die PISA-Studie, deren dritter Zyklus dann durchgeführt wird – Aufgaben für das Fach Mathematik bundesweit normiert werden. Weitere Arbeitsschritte wird das *Institut für Qualitätsentwicklung im Bildungswesen* übernehmen, das die Kultusministerkonferenz an der Berliner Humboldt-Universität eingerichtet hat. Die dort entwickelten Aufgabenpools und Tests werden als Messinstrumente für länderinterne und ländervergleichende Untersuchungen wie auch für die Evaluation von Einzelschulen zur Verfügung stehen, um die Erfüllung der Standards zu überprüfen. Der Beitrag von Köller und Mitarbeitern im vorliegenden Heft zeigt exemplarisch, welche Art von Erkenntnissen aus solchen standard-bezogenen Untersuchungen resultieren könnte. Sie verwenden (mangels nationaler Standards) ein Testverfahren, das in den USA bei der Studienzulassung ausländischer Bewerber eingesetzt wird, als Standard zur Beurteilung der Englischleistungen von Oberstufenschülern und zeigen – unter sorgfältiger Kontrolle von Hintergrundfaktoren wie etwa dem sozio-ökonomischen Status der Herkunftsfamilie –, welche Differenzen zwischen den Lernergebnissen in unterschiedlichen Schulformen bestehen. Für eine Strategie der Qualitätsentwicklung ist besonders wichtig, dass sie Wege zur Korrektur von Leistungsrückständen aufzeigen wie z.B. Auslandsaufenthalte.

In der bildungspolitischen Diskussion ist in Deutschland noch weitgehend ungeklärt, wie die Standards an den Schulen implementiert werden, wie oft und nach welchen Regeln Schulen zukünftig evaluiert werden usw. An dieser Stelle bedarf die Debatte über Bildungsstandards dringend der Anbindung an Diskurse der Schulentwicklungsforschung, die von Forschern und Praktikern dieses Bereichs nachdrücklich eingefordert wird. Erforderlich wäre es zudem, mittels empirischer Designs die intendierten und nicht intendierten Wirkungen der Einführung der Standards zu prüfen. Hierzu lohnt es sich, einschlägige nordamerikanische Studien zu konsultieren.

## **2. Zur Wirkung von Bildungsstandards: Konzepte und empirische Befunde aus den USA**

Kein anderes Bildungssystem ist so stark durch das Konzept der (normativen) Standards geprägt wie das US-amerikanische. Mitte der 1980er-Jahre verbanden sich die Forderungen nach anspruchsvolleren Lerninhalten, höheren Lernergebnissen und mehr Chancengleichheit im Bildungssystem mit den Traditionen des lernzielorientierten Un-

terrichtens und der ausgeprägten Testpraxis (vor allem beim Übergang von der Schule in berufliche und akademische Ausbildungen) zu der Idee der Standards (vgl. die sehr instruktiven Darstellungen bei Horn 2004 und Ravitch 1995). Selbst Kritiker des „standards-movement“ wie die Schulentwicklungsforscherin Darling-Hammond (1997, S. 212) konstatieren eine im Vergleich zu Europa (sic!) geringe Professionalität nord-amerikanischer Lehrkräfte und das Fehlen klarer, fachlich-systematisch strukturierter Curricula. Auch unter diesen Kritikern ist daher die Forderung nach professionellen Standards als Leitlinien der Unterrichtspraxis kaum umstritten. Diskutiert wird aber mit großer Schärfe, welche Art von Standards gebraucht werden (z.B. *content, performance* oder *opportunity to learn-standards*), wie sie entwickelt, implementiert und überprüft werden sollen. Hinzu kommen politische Streitfragen (beispielsweise die Rollenverteilung zwischen den Staaten und der Bundesregierung), Verteilungskämpfe (z.B. um die Chance der Geographie, in den engen Kreis der Standard-basierten Fächer aufgenommen zu werden) und ideologische Auseinandersetzungen, wie sie bei der Vorbereitung von Standards für Englisch und Geschichte auftraten. Diese Konfliktlinien haben Anfang der 1990er-Jahre die Entwicklung nationaler Bildungsstandards gestoppt (Ravitch 1995). Forciert wurde aber – gerade unter der Administration von Bush jun. – der Ausbau von Systemen der Rechenschaftslegung durch Testverfahren.

Zum Gesamtkomplex der Standard-Diskussion in den USA gehören die folgenden Stränge der Bildungsreform:

1. *State standards*: Praktisch alle US-Bundesstaaten haben inzwischen Dokumente veröffentlicht, die verbindliche Erwartungen an Lernergebnisse für Schülerinnen und Schüler ausdrücken. Diese Standards sind durchweg fach- und inhaltsbezogen.
2. *Accountability systems*: Das von der Bush-Administration verabschiedete „no child left behind (NCLB)“-Gesetz verlangt von allen Staaten, Systeme der Rechenschaftslegung zu etablieren. Überall wurden engmaschige Testprogramme etabliert, die flächendeckend jeden Schüler und jede Schülerin in den Jahrgangsstufen 3 bis 8 erfassen. Die eingesetzten Tests sollen sich an den jeweiligen Standards orientieren (*standard-based-testing*). Im Zentrum stehen vorerst Lesefähigkeit und mathematische Kompetenzen. Rückmeldungen aus dem Testverfahren gehen an Schulen und Distrikte, zum Teil auch an einzelne Lehrer, Schüler und Eltern. Schulen müssen nachweisen, dass sie die Lernergebnisse – auch bei Schülern aus ethnischen Minderheiten – von Jahr zu Jahr steigern, bis 95 Prozent aller Schülerinnen und Schüler einen definierten Mindeststandard erreichen. Für „*failing schools*“ sind besondere Interventionen vorgesehen.
3. *Standard-based school reform*: Eine ganze Reihe von Schulentwicklungsinitiativen auf Einzelschul- und Distriktebene tritt mit dem Anspruch auf, den Unterricht, die Arbeit der Schulleitungen und der Kollegien stärker auf das Erreichen der Standards auszurichten und dadurch Lernergebnisse zu verbessern. Hier wird die Implementierung von Standards zum Ausgangspunkt und Maßstab der Schulentwicklung genommen. Unterschiedliche Projekte dieser Art werden durch die US-Bundesregierung oder Stiftungen gefördert.

4. *Professional standards*: Unabhängige Organisationen wie die Mathematiklehrervereinigung NCTM haben seit den 1980er-Jahren Standards entwickelt, die Erwartungen an Lernergebnisse, aber auch an die Unterrichtspraxis, die Qualifikation der Lehrerinnen und Lehrer und das System der Leistungsbewertung formulieren. Sie sind nicht notwendig mit Testsystemen verknüpft. Die didaktische Konzeption etwa der NCTM-Standards führt, wie systematische Vergleiche gezeigt haben, über die eher oberflächlichen *state standards* weit hinaus.
5. *Standard-based curriculum reform*: An den Leitideen der professionellen Standards orientieren sich Reformprojekte vor allem im Bereich der Mathematik (Senk/Thompson 2003) und der Naturwissenschaften, welche die Philosophie des verständnis-orientierten Lehrens und Lernens in Lehrmaterialien, Lehrpläne, Weiterbildungsprogramme, Beratungs- und Coaching-Aktivitäten umsetzen. Zumeist sind solche Reforminitiativen auch mit der Entwicklung von spezifischen Tests verbunden, die – enger als staatliche *large scale assessments* – an das Curriculum und seine Lernziele angepasst sind. Das prominenteste Beispiel ist das von Lauren Resnick vom *Learning Research and Development Center* in Pittsburgh inspirierte *New Standards Project* (Tucker/Codding 1998).

In der Praxis überlagern sich die verschiedenen Reformstränge mit ihren unterschiedlichen Standard-Begriffen. Insbesondere bilden Standards und darauf bezogene Tests bzw. Evaluationsverfahren eine Einheit. Viele Autoren meinen aber eine klare Grenze ziehen zu können zwischen „schlechten“ technischen Standards (Horn 2004) bzw. bürokratischen Rechenschaftsmodellen (O'Day 2004) einerseits, „guten“ professionellen Standards und Rechenschaftssystemen, *standards of complexity* (Horn 2004) bzw. *standards without standardization* (Darling-Hammond 1997) andererseits. Ausgehend von Theorien des Organisationslernens und Befunden der Schulentwicklungsforschung sind sich Experten darin einig, dass Standards nur dann Schule und Unterricht verändern und Lernergebnisse verbessern können, (a) wenn Lehrer über die notwendigen Kompetenzen und Einstellungen verfügen bzw. diese erwerben, (b) wenn Schulen – gestützt durch externe Beratung – beginnen, ihre eigene Praxis zu evaluieren und weiter zu entwickeln, (c) wenn die eingesetzten Testverfahren die anspruchsvollen Lernziele tatsächlich abbilden und (d) wenn Schulen bzw. Lehrkräften und Schülern, deren Ergebnisse unbefriedigend sind, gezielte Unterstützung gegeben wird (vgl. etwa den wichtigen Sammelband von Fuhrman/Elmore 2004.)

Diese Empfehlungen sind allerdings nur zum geringen Teil Ergebnis empirischer Forschungs- oder Evaluationsstudien zu den Bedingungen der Wirksamkeit von Standards. Angesichts der sich überlagernden Reformstränge ist es nahezu unmöglich, die Einführung von Standards im Sinne eines „reforms as experiments“-Ansatzes (Campbell 1969) zu evaluieren. Eine Arbeitsgruppe des *National Research Council* hat unlängst (2002) ein Rahmenkonzept für Forschung über die Wirkungen von Standards vorgelegt, das bezeichnenderweise als Desiderat formuliert ist, nicht als Zusammenfassung vorhandener Studien. Die Autoren gehen davon aus, dass Standards unter dem Einfluss von Kontextfaktoren (öffentliche Meinung, politische Strategien, Einfluss professioneller

Organisationen) über drei „Kanäle“ die Lehrer und ihre Unterrichtspraxis beeinflussen und sich schließlich auf Lernprozesse und Lernergebnisse auswirken können. Diese Kanäle sind: Curriculum (Verteilung der Lernzeit auf Fächer und Inhalte, Setzung von Schwerpunkten, Unterrichtsmaterialien usw.), Lehrerbildung (insbesondere berufs begleitendes, kollegiales professionelles Lernen) sowie schließlich *assessment* und *accountability*, also das System der Leistungsüberprüfung und Evaluation.

Einige empirische Befunde liegen vor, die unterschiedliche Stränge der Standard-Diskussion betreffen. Sehr gut dokumentiert und – teils unter Nutzung von Kontrollgruppen – systematisch evaluiert sind verschiedene Curricula, mit denen NCTM-Standards implementiert werden sollen. Der Sammelband von Senk/Thompson (2003) informiert über Studien im Bereich von Primarschulen bis zu High Schools. Die Ergebnisse werden dahingehend interpretiert, dass sich in den Curriculumprojekten tatsächlich die Unterrichtsinhalte und -methoden verändern. Schülerinnen und Schüler erbringen in den spezifischen Zieldimensionen der Reformprojekte (konzeptuelles Verständnis, Problemlösefähigkeiten) tatsächlich höhere Leistungen, unterscheiden sich aber in den Standard-Leistungsdimensionen, wie sie durch staatliche Tests erfasst werden, nicht von Vergleichsgruppen. Eine neuere Beobachtungs- und Teststudie von Thompson (2004) stellt fest, dass drei aus den NCTM-Standards abgeleitete Unterrichtspraktiken (Projektarbeit, Nutzung von Anschauungsmitteln und Selbsteinschätzung der Lernenden) selbst auf standardisierte Tests positive Effekte haben. Dies wird als Effekt von *standards-based instruction* interpretiert.

Zahlreiche Untersuchungen widmen sich der Frage, welche Effekte Tests auf den Unterricht haben. Neben den Verdacht, „teaching to the test“ enge das Curriculum und die Unterrichtsmethoden ein, treten auch positive Erwartungen, wonach von anspruchsvollen Tests ein Reformimpuls ausgehen könnte. Je nach Art des Lernziels und nach der Passung (*alignment*) zwischen Test und Ziel können die so genannten *washback*-Effekte positiv oder negativ interpretiert werden. Neuere Forschungsübersichten von Herman (2004) sowie Cheng/Watanabe/Curtis (2004) kommen zu dem Ergebnis, dass Lehrer in den USA ihre Unterrichtspraxis tatsächlich in bedeutsamem Ausmaß an den Testsystemen ausrichten – eher noch als an den Standards selbst. Herman (2004) folgert daraus, dass standard-orientiertes Testen Schulen und Lehrer ermutigen könne, den Unterricht auf zentrale Lerninhalte zu fokussieren. Sie betont aber, dass Systeme der Leistungsmessung sehr viel differenzierter ausgelegt sein müssten als gängige staatliche Testverfahren, bei denen in den vergangenen Jahren sogar ein Rückschritt auf enge, oberflächliche Aufgabentypen zu beobachten sei. Stattdessen gelte es, multiple Kriterien anzuwenden, komplexe Aufgabenformen und Portfolios zu nutzen, die Leistungsprüfung in den Unterricht zu integrieren. Als Problem wird übereinstimmend festgehalten, dass nicht getestete Fächer bzw. Themengebiete vernachlässigt werden.

Eine dritte Linie von Forschungen versucht herauszufinden, welche Auswirkungen *accountability*-Systeme als Ganze haben. Die Bildungsökonominnen Wyckoff/Naples (2000) kommen anhand von Daten aus dem Staat New York zu dem Ergebnis, dass sich durch anspruchsvollere, verpflichtende Abschlussprüfungen das Leistungsniveau insgesamt heben lässt, allerdings ohne die Chancenungleichheit zwischen ethnischen Grup-

pen zu beseitigen. Carnoy/Loeb (2004) zeigen, dass Mathematikleistungen der 8. Jahrgangsstufe in den Staaten, die ein strengeres System der Rechenschaftslegung implementiert haben, zwischen 1996 und 2000 stärker angestiegen seien als in anderen Bundesstaaten. Entsprechende Unterschiede konnten allerdings nicht für die Primarschule berichtet werden. O'Day (2004, S. 25) weist darauf hin, dass die Interpretation solcher Leistungsanstiege umstritten ist. Grosse Veränderungen, die unter einzelstaatlichen Testprogrammen wie z.B. in Texas sichtbar sind, lassen sich in national repräsentativen Studien nur zum Teil bestätigen (Herman 2004, S. 152). Überdies gibt es Hinweise darauf, dass ausgerechnet die schlechter abschneidenden Schulen weniger gut aus Rückmeldungen lernen (O'Day 2004, S. 26). Die Befundlage zur Auswirkung von *accountability*-Systemen ist also insgesamt unbefriedigend. Insbesondere die Frage der Nachhaltigkeit der Effekte bleibt offen. Im Kontext des NCLB-Gesetzes erwarten Experten aufgrund von Simulationsstudien und Erfahrungen der ersten Jahre große Turbulenzen, weil die Erfolgskriterien zu hoch angesetzt sind, sehr viele Schulen zu *failing schools* stempeln und zu Fehlklassifikationen führen werden (Baker/Linn 2004, S. 60).

Angesichts der heftig geführten Debatten und der vielen offenen Forschungsfragen, die in den USA nach fast 20-jähriger Erfahrung mit Bildungsstandards vorzufinden sind, lässt sich ermesen, welche politischen Herausforderungen auf das deutsche Bildungssystem zukommen und wie groß der Bedarf an systematischer, solider Forschung und Evaluation ist. Die zentralen Probleme dürften in der für Deutschland noch weitgehend ungeklärten Verknüpfung zwischen Standards, Schulentwicklung, Schulevaluation und Rechenschaftslegung liegen.

## Literatur

- Baker, E.L./Linn, R.L. (2004): Validity Issues for Accountability Systems. In: Fuhrman, S.H. & Elmore, R.F. (Eds.): *Redesigning Accountability Systems for Education*. New York/London: Teachers College Press, S. 47-72.
- Baumert, J./Trautwein, U./Artelt, C. (2003): Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In: Deutsches PISA-Konsortium (Hrsg.): *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich, S. 261-331
- Campbell, D.T. (1969): Reforms as Experiments. In: *American Psychologist* 24, S. 409-429.
- Carnoy, M./Loeb, S. (2004): Does External Accountability Affect Student Outcomes? A Cross-State Analysis. In: Fuhrman, S.H./Elmore, R.F. (Eds.): *Redesigning Accountability Systems for Education*. New York/London: Teachers College Press, S. 189-219.
- Cheng, L./Watanabe, Y./Curtis, A. (2004) (Eds.): *Washback in Language Testing. Research Contexts and Methods*. Mahwah/London: Lawrence Erlbaum.
- Cizek, G.J. (2001) (Eds.): *Setting Performance Standards. Concepts, Methods, and Perspectives*. Mahwah/London: Lawrence Erlbaum.
- Darling-Hammond, L. (1997). *The Right to Learn. A Blueprint for Creating Schools that Work*. San Francisco: Jossey-Bass.
- Döbert, H./Klieme, E./Sroka, W. (2004) (Eds.): *Conditions of School Performance in Seven Countries. A Quest for Understanding the International Variation of PISA Results*. Münster: Waxmann.
- Fitzner, T. (2004) (Hrsg.): *Bildungsstandards. Internationale Erfahrungen – Schulentwicklung – Bildungsreform*. Bad Boll: Evangelische Akademie.

- Fuhrman, S.H./Elmore, R.F. (2004) (Eds.): *Redesigning Accountability Systems for Education*. New York/London: Teachers College Press.
- Hameyer, U./Frey, K./Haft, H. (Hrsg.) (1983): *Handbuch der Curriculumforschung*. Weinheim: Beltz.
- Herman, J.L. (2004): The Effects of Testing on Instruction. In: Fuhrman, S.H./Elmore, R.F. (Eds.): *Redesigning Accountability Systems for Education*. New York/London: Teachers College Press, S. 141-166.
- Horn, R.A. Jr. (2004): *Standards*. New York: Peter Lang.
- Ingenkamp, K. (1989): *Diagnostik in der Schule*. Weinheim und Basel: Beltz.
- 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.
- Köller, O./Baumert, J./Schnabel, K. (1999): Wege zur Hochschulreife: Offenheit des Systems und Sicherung vergleichbarer Standards. Analysen am Beispiel der Mathematikleistungen von Oberstufenschülern an integrierten Gesamtschulen und Gymnasien in Nordrhein-Westfalen. *Zeitschrift für Erziehungswissenschaft* 2, S. 385-422.
- Matthews, P. (2004): The Work and Impact of OFSTED. In: Fitzner, T. (Hrsg.): *Bildungsstandards. Internationale Erfahrungen – Schulentwicklung – Bildungsreform*. Bad Boll: Evangelische Akademie, S. 220-233.
- National Research Council (2002): *Investigating the Influence of Standards: A framework for research in mathematics, science, and technology education*. Washington, D.C.: National Academy Press.
- O'Day, J.A. (2004): Complexity, Accountability, and School Improvement. In: Fuhrman, S.H./Elmore, R.F. (Eds.): *Redesigning Accountability Systems for Education*. New York/London: Teachers College Press, S. 15-43.
- OECD (1991): *Schulen und Qualität*. Frankfurt a.M.: Lang.
- Oser, F./Oelkers, J. (Hrsg.) (2001): *Die Wirksamkeit der Lehrerbildungssysteme: Von der Allrounderbildung zur Ausbildung professioneller Standards*. Chur, Zürich: Rüegger.
- Ravitch, D. (1995): *National Standards in American Education. A Citizen's Guide*. Washington, D.C.: Brookings Institution Press.
- Senk, S.L./Thompson, D.R. (2003) (Eds.): *Standards-Based School Mathematics Curricula. What Are They? What Do Students Learn?* Mahwah/London: Lawrence Erlbaum.
- Terhart, E. (2002): *Standards für die Lehrerbildung. Gutachten für die Kultusministerkonferenz*. ZKL-Text Nr. 23. Universität Münster.
- Tucker, M.S./Coddling, J.B. (1998): *Standards for our schools: How to set them, measure them, and reach them*. San Francisco: Jossey-Bass.
- Wyckoff, J.H./Naples, M. (2000): Educational finance to support high learning standards: a synthesis. In: *Economics of Education Review*, 19, S. 305-318.

*Anschrift des Autors:*

Prof. Dr. Eckhard Klieme, Deutsches Institut für Internationale Pädagogische Forschung, Schloss-Str. 29, 60486 Frankfurt a.M., E-Mail: klieme@dipf.de.