

Köller, Olaf

## **Bildungsstandards in einem Gesamtsystem der Qualitätssicherung im allgemeinbildenden Schulsystem Deutschlands**

*Klieme, Eckhard [Hrsg.]; Tippelt, Rudolf [Hrsg.]: Qualitätssicherung im Bildungswesen. Weinheim u.a. : Beltz 2008, S. 59-75. - (Zeitschrift für Pädagogik, Beiheft; 53)*

urn:nbn:de:0111-opus-72692



in Kooperation mit / in cooperation with:

# **BELTZ JUVENTA**

<http://www.juventa.de>

### **Nutzungsbedingungen / conditions of use**

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.

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.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.  
By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

**peDOCS**  
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)  
Mitglied der Leibniz-Gemeinschaft  
Informationszentrum (IZ) Bildung  
Schloßstr. 29, D-60486 Frankfurt am Main  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Zeitschrift für Pädagogik · 53. Beiheft

# **Qualitätssicherung im Bildungswesen**

Eine aktuelle Zwischenbilanz

Herausgegeben von  
Eckhard Klieme und Rudolf Tippelt

Beltz Verlag · Weinheim und Basel

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder ein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder auf ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen oder sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopie hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder genutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG Wort, Abteilung Wissenschaft, Goethestr. 49, 80336 München, bei der die einzelnen Zahlungsmodalitäten zu erfragen sind.

©2008 Beltz Verlag · Weinheim und Basel  
Herstellung: Klaus Kaltenberg  
Gesamtherstellung: Druckhaus »Thomas Müntzer«, Bad Langensalza  
Printed in Germany  
ISSN 0514-2717  
Bestell-Nr. 41154

# Inhaltsverzeichnis

<i>Eckhard Klieme/Rudolf Tippelt</i>	
Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz .....	7
<b>Bereiche</b>	
<i>Wolfgang Tietze</i>	
Qualitätssicherung im Elementarbereich .....	16
<i>Hartmut Ditton</i>	
Qualitätssicherung in Schulen .....	36
<i>Olaf Köller</i>	
Bildungsstandards in einem Gesamtsystem der Qualitätssicherung im allgemeinbildenden Schulsystem Deutschlands .....	59
<i>Thomas Olk/Karsten Speck</i>	
Qualität und Qualitätsentwicklung in der Sozialpädagogik .....	76
<i>Philipp Gonon</i>	
Qualitätssicherung in der beruflichen Bildung. Eine Bilanz .....	96
<i>Klaus Meisel</i>	
Qualitätsmanagement und Qualitätsentwicklung in der Weiterbildung .....	108
<i>Lutz von Rosenstiel</i>	
Qualitätssicherung in der betrieblichen Weiterbildung .....	122
<i>Andrä Wolter/Christian Kerst</i>	
Akkreditierung als Verfahren der Qualitätssicherung von Studiengängen in Deutschland. Eine Policy-orientierte Analyse .....	135
<i>Bernhard Schmidt</i>	
Qualität der Lehre an Hochschulen .....	156
<i>Rudolf Tippelt</i>	
Qualitätsstandards und Forschungsevaluationen in der Erziehungswissenschaft .....	171

**Implementierung von Qualitätssicherung**

*Helmut Fend*

Dimensionen von Qualität im Bildungswesen. Von Produktindikatoren zu  
Prozessindikatoren am Beispiel der Schule ..... 190

*Esther Berner/Jürgen Oelkers/Kurt Reusser*

Implementation von Bildungsstandards: Bedingungen des Gelingens  
(und Scheiterns) aus internationaler Sicht ..... 210

Olaf Köller

# **Bildungsstandards in einem Gesamtsystem der Qualitätssicherung im allgemein bildenden Schulsystem Deutschlands**

## **1. Überblick**

Als Folge der Publikationen zu den Ergebnissen der *Programme of International Student Assessment-Studie* (PISA) 2000 (vgl. u.a. Deutsches PISA-Konsortium 2001, 2002) haben sich in den deutschsprachigen Ländern Österreich, Schweiz, Luxemburg und Deutschland erhebliche Veränderungen in der Steuerung des allgemein bildenden Schulsystems ergeben, die gern als Paradigmenwechsel von der Input- zur Outputsteuerung beschrieben werden (vgl. Klieme u.a. 2003). Erreichte fachliche Kompetenzniveaus auf Seiten der Schülerinnen und Schüler sollen Rückschlüsse auf den Erfolg des Bildungssystems zulassen. Ausdruck findet diese Outputsteuerung unter anderem in der langfristigen Beteiligung an internationalen *Large-Scale-Assessments* wie PISA (vgl. z.B. Deutsches PISA-Konsortium 2007) oder der Progress in Reading Literacy Study (PIRLS) / Internationalen Grundschul-Lese-Untersuchung (IGLU) (vgl. z.B. Bos u.a. 2007), aber auch in der Verabschiedung von Bildungsstandards, in denen für ausgewählte Fächer definiert ist, über welche Kompetenzen Schülerinnen und Schüler zu festen Zeitpunkten ihrer schulischen Karriere verfügen sollten.

Mit ihren Beschlüssen von 2003 und 2004 hat die Ständige Konferenz der Kultusminister (KMK) für alle 16 Länder der Bundesrepublik Deutschland solche verbindlichen Bildungsstandards für die Grundschule und das Ende der Sekundarstufe I verabschiedet. Trotz Föderalismus sind die darin vereinbarten Ziele schulischer Lehr-Lernprozesse für alle 16 Länder bindend, dementsprechend haben sich auch alle Länder verpflichtet, Sorge zu tragen, dass die Standards und die mit ihnen verbundenen Ziele im Unterricht umgesetzt werden. Für die Sekundarstufe I wurden bewusst keine schulformspezifischen, sondern abschlussbezogene Standards für den Hauptschul- und Mittleren Schulabschluss definiert. Damit wurde dem Umstand Rechnung getragen, dass in allen Ländern Deutschlands derselbe Abschluss an unterschiedlichen Schulformen vergeben wird (vgl. z.B. Köller u.a. 2004).

Mit den KMK-Beschlüssen vom Oktober 2007 ist der Auftrag dahingehend erweitert worden, auch Standards für das Ende der gymnasialen Oberstufe zu erarbeiten, und zwar erneut für die Fächer Deutsch, erste Fremdsprache (Französisch, Englisch), Mathematik und die drei Naturwissenschaften (Biologie, Chemie, Physik). Damit werden in absehbarer Zeit in den Kernfächern abschlussbezogene Standards für das gesamte allgemein bildende Schulsystem in Deutschland vorliegen. Dies fasst noch einmal die Tabelle 1 im Überblick zusammen.

Tab. 1: **Fächer, für die laut KMK-Beschluss Bildungsstandards vorliegen bzw. erarbeitet werden, nach Abschluss**

	Ende der 4. Jahrgangsstufe	Ende der Sekundarstufe I		Ende der gymnasialen Oberstufe
		Hauptschulabschluss	Mittlerer Schulabschluss	Allgemeine Hochschulreife
Deutsch	■	■	■	■
Mathematik	■	■	■	■
Englisch		■	■	■
Französisch		■	■	■
Biologie			■	■
Chemie			■	■
Physik			■	■

An die Verabschiedung der Bildungsstandards hat die KMK in ihren Beschlüssen vom Juni 2006 (vgl. KMK 2006) ein breites System der Qualitätssicherung gekoppelt, das neben stichprobenbasierten internationalen und nationalen Vergleichen auch regelmäßige flächendeckende Vergleichsarbeiten vorsieht und dazu dienen soll zurückzumelden, ob Schülerinnen und Schüler die in den Standards formulierten Leistungserwartungen einlösen können. Flankiert wird dieses System durch ein indikatorengestütztes Monitoring auf gesamtstaatlicher Ebene im Rahmen der nationalen Bildungsberichterstattung (vgl. Klieme u.a. 2007) und durch die systematische Evaluierung von Einzelschulen, wofür nahezu alle Länder inzwischen Schulinspektionsteams eingerichtet haben.

Das Ziel dieses Beitrages besteht darin, das Gesamtsystem der Qualitätssicherung im deutschen Schulsystem, wie es sich aus den Bildungsstandards ergibt, vorzustellen und mit seinen Implikationen und Nebenwirkungen zu diskutieren. Darüber hinaus soll im Einklang mit den Ausführungen von Klieme u.a. (2003) präzisiert werden, welches Potenzial die in Deutschland verabschiedeten Bildungsstandards für die Unterrichtsentwicklung haben und dass sie ihre Wirkung auch nur dann entfalten werden, wenn sie die Grundlagen sowohl von qualitätssichernden als auch qualitätsentwickelnden Maßnahmen sind.

## 2. Bildungsstandards im allgemeinbildenden Schulsystem Deutschlands

Die KMK hat als Reaktion auf PISA 2000 im Mai 2002 beschlossen, länderübergreifende Bildungsstandards in den Fächern Deutsch und Mathematik sowie für die erste Fremdsprache (Englisch, Französisch) für den Mittleren Schulabschluss und für den Hauptschulabschluss zu erarbeiten. Die anschließende konzeptionelle Rahmung um die Ein-

führung der Bildungsstandards wurde maßgeblich durch die Expertise „Zur Entwicklung nationaler Bildungsstandards“ (Klieme u.a. 2003) vorangebracht. Sie beschreibt die Funktionen, konzeptuellen Grundlagen und die Entwicklung von Bildungsstandards. In ihrer Expertise haben Klieme u.a. (2003) ausgeführt, wodurch sich Standards auszeichnen sollen. Danach benennen sie fachspezifisch die wesentlichen Ziele pädagogischer Arbeit, ausgedrückt als erwünschte Lernergebnisse bzw. Kompetenzen der Schülerinnen und Schüler zu bestimmten Zeitpunkten ihrer Bildungsbiographien. In den entsprechenden KMK-Dokumenten wurde dies dahin gehend umgesetzt, dass Standards als sogenannte *Can do Statements* formuliert sind. Beispielsweise heißt es im Bereich der ersten Fremdsprache für den Mittleren Schulabschluss für den Kompetenzbereich Lesen: „Die Schülerinnen und Schüler können Korrespondenz lesen, die sich auf das eigene Interessengebiet bezieht und die wesentliche Aussage erfassen (B2); klar formulierte Anweisungen, unkomplizierte Anleitungen, Hinweise und Vorschriften verstehen (B1/ B2) (...)“ (KMK 2004, S. 14). Kompetenzen sind hier als gezeigtes Verhalten definiert. Folgt man einem stärker psychologisch geprägten Kompetenzkonzept (vgl. z.B. Weinert 2001), so wird man nicht das Verhalten selbst, sondern die bei den Schülerinnen und Schülern verfügbaren oder von ihnen erlernbaren kognitiven Fähigkeiten und Fertigkeiten, die notwendig sind, um bestimmte Probleme bzw. Aufgaben lösen zu können, als Kompetenzen verstehen. In diesem Fall gewinnen Kompetenzen den Charakter psychologischer Konstrukte, welche mit Hilfe von Messinstrumenten operationalisiert werden können.

### *2.1 Von Bildungsstandards zu nationalen Skalen – Aufgabenentwicklung und Normierung*

Bildungsstandards sind in den entsprechenden Dokumenten der KMK zunächst verbal formuliert (vgl. z.B. für die Fremdsprachen KMK 2004). Sie benennen mehr oder weniger konkret die fachspezifischen Kompetenzen, die Schülerinnen und Schülern in der Regel erreichen sollten (sogenannte Regelstandards). Was genau „in der Regel“ heißt, ist dabei unklar und nicht ganz zu Unrecht verschiedentlich kritisiert worden (vgl. z.B. Bremerich-Vos 2003; Spinner 2005). Man wird dieser Kritik allerdings erst begegnen können, wenn auf der Basis der Standards normierte Aufgabenpools und damit verbunden nationale Skalen entwickelt wurden, die eine Verortung von Schülerinnen/ Schülern und Aufgaben (Items) erlauben. Dann wird es möglich sein, Leistungsbereiche zu definieren, die Schülerinnen und Schüler in der Regel (Regelstandards) oder mindestens (Mindeststandards; vgl. zu dieser Debatte auch Klieme u.a. 2003) am Ende der 4. Jahrgangsstufe oder am Ende der Sekundarstufe I erreicht haben sollten.

Mit der Gründung des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) hat die KMK im Jahr 2004 die Voraussetzung geschaffen, entsprechende Aufgabenpools und nationale Skalen zu entwickeln (vgl. zu den Aufgaben des IQB Köller 2005). Der Prozess – von den Standards zu nationalen Skalen – soll im Folgenden allgemein beschrieben werden. Für das Fach Mathematik in der Sekundarstufe I findet sich eine aus-



fürliche Beschreibung des Prozesses bei Ehmke u.a. (2006). Rupp u.a. (2008) beschreiben den Aufgabenentwicklungsprozess für das Fach Englisch.

Die Itementwicklung auf der Basis der Bildungsstandards folgt weitgehend dem *State of the Art* (vgl. z.B. Rost 2004), hat allerdings einige Besonderheiten zu berücksichtigen. Hierzu zählt unter anderem in der Sekundarstufe I, dass Aufgaben für unterschiedliche Jahrgangsstufen (8, 9 und 10) und Abschlüsse (Hauptschulabschluss und Mittlerer Schulabschluss) generiert werden müssen, die sich dennoch auf einer gemeinsamen Metrik abbilden lassen. Im Wesentlichen vollzieht sich der Prozess in sieben Schritten:

1. *Fachdidaktisch und lernpsychologisch fundierte Konkretisierung der Kompetenzen in den jeweiligen Fächern*: Mit Blick auf die Definition von Messmodellen ist es unumgänglich zu präzisieren, was unter den jeweiligen Kompetenzen (Konstrukten) zu verstehen ist, konkret: Wie ist beispielsweise das Hör- und Leseverstehen in der Fremdsprache theoretisch verankert? Die dafür notwendigen konzeptionellen Arbeiten am IQB finden in enger Kooperation mit Fachdidaktikerinnen und Fachdidaktikern statt.
2. *Erarbeitung von Richtlinien zur Konstruktion von Testaufgaben/Items (Item- und Testspezifikationen)*: Hier geht es um die Erstellung von Trainingsmaterial, das bei der Schulung von Aufgabenentwicklern eingesetzt werden kann. In diesen Materialien werden die Konstrukte/Kompetenzen erläutert, noch wichtiger sind allerdings die Ausführungen, wie Tests üblicherweise konstruiert werden und welche Fehler man bei der Itementwicklung vermeiden sollte. Dies erfolgt im Einklang mit gängigen Werken zur Testkonstruktion (vgl. z.B. Rost 2004).
3. *Initiierung und Aufrechterhaltung der Itementwicklung durch erfahrene Lehrkräfte*: Mit Hilfe der in Schritt 2 erarbeiteten Materialien werden im dritten Schritt Lehrkräfte aller 16 Länder geschult. Anschließend generieren diese Lehrkräfte unter fachdidaktischer Betreuung die Testaufgaben in regionalen Arbeitsgruppen. Lehrkräfte im Bereich der Aufgabenentwicklung zu professionalisieren, erfüllt hier auch den Zweck, die so erreichte Expertise später in die jeweiligen Länder zu tragen, um sie dort bei Arbeiten zur Qualitätssicherung anzuwenden.
4. *Begutachtung der Items durch Experten*: Bildungsforscher, Psychometriker und Fachdidaktiker beurteilen und kommentierten in diesem Schritt die Items hinsichtlich ihrer psychometrischen und fachdidaktischen Güte. Dazu zählt auch die Klassifikation der Aufgaben hinsichtlich der durch sie gemessenen Kompetenzen und Anforderungsbereiche.
5. *Empirische Erprobung der entwickelten Aufgaben in großen Schülerstichproben*: Sämtliche Items werden in großen Stichproben erprobt. Die Stichprobengröße wird dabei so gewählt, dass hinreichend robuste Schätzungen der Itemparameter auf der Grundlage von Modellen der Item-Response-Theory möglich sind. Zusätzlich werden kleinere empirische Untersuchungen zur Validierung der Items durchgeführt.
6. *Normierung der Aufgaben*: Auf der Basis national repräsentativer Stichproben von Schülerinnen und Schülern werden die Items kalibriert und nationale Skalen wer-

den definiert. Neben den neu entwickelten Instrumenten werden auch etablierte Leistungstests eingesetzt, um so zusätzliche Validitätshinweise zu erhalten.

7. *Definition von Kompetenzstufen (Standard-Setting)*: Im Konsens zwischen Fachdidaktikern, Erziehungswissenschaftlern, Lehrkräften und Vertretern der Politik und unter Einbeziehung von empirische Befunden aus den Pilotierungs- und Normierungsstudien werden Niveaustufen definiert und festgestellt, bei welchen Werten auf den nationalen Skalen davon ausgegangen werden kann, dass die entsprechenden Schülerinnen und Schüler die in den Standards beschriebenen Leistungserwartungen erfüllt haben. Dieses konsensuelle Verfahren wird dabei einem psychometrischen Vorgehen, bei dem schwierigkeitsgenerierende Merkmale von Aufgaben zur Definition von Niveaustufen bzw. deren Grenzen verwendet werden (vgl. Hartig 2007), vorgezogen.

Die Abbildung 1 zeigt am Beispiel Deutsch und Mathematik in der Grundschule, mit welchen Zeiträumen bei der Abarbeitung der Schritte 1 bis 7 zu rechnen ist. *Cum grano salis* vergehen vier Jahre vom Beginn der Aufgabenentwicklung bis zum Vorliegen nationaler Skalen.

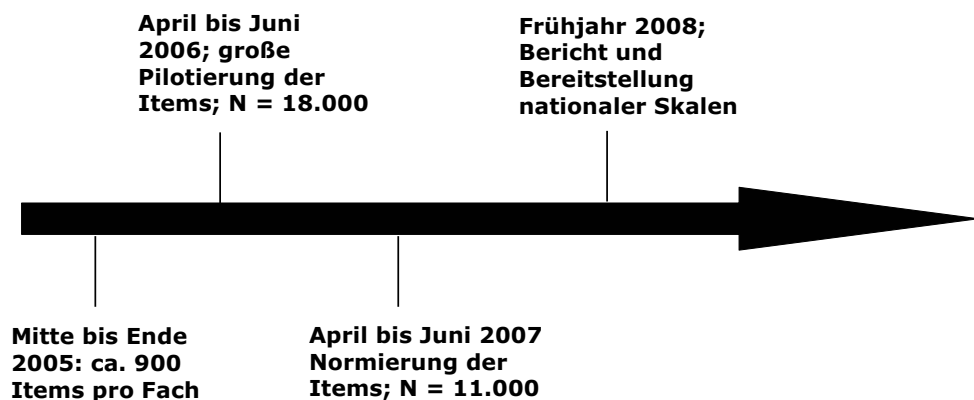


Abb. 1: Zeitlicher Ablauf bei der Standard-basierten Aufgabenentwicklung für die Fächer Deutsch und Mathematik in der Grundschule

Im Prozess der Itementwicklung wurden pro Fach ca. 900 Items für die Jahrgangsstufen 3 und 4 erstellt, die jeweils den allgemeinen und inhaltlichen Kompetenzen sowie den Anforderungsniveaus zugeordnet wurden. Die Aufgabenentwicklung für die beiden nebeneinander liegenden Jahrgänge war nötig, da die Standards zum einen für das Ende der vierten Jahrgangsstufe definiert sind, zukünftige standard-basierte Vergleiche allerdings in der dritten Jahrgangsstufe durchgeführt werden sollen (vgl. 2.2).

Die Aufgaben wurden im Frühjahr 2006 im Rahmen von IGLU (vgl. Bos u.a. 2007) in einer national repräsentativen Stichprobe der 3. und 4. Jahrgangsstufe erprobt. Inse-

samt bearbeiteten rund 18.000 Schülerinnen und Schüler in einem äußerst komplexen *Multi-Matrix-Design* die rund 900 Items pro Fach. Konkret wurde ein Untersuchungsplan umgesetzt, bei dem jede Schülerin / jeder Schüler nur eine Teilmenge der Aufgaben vorgelegt wurde. Die Testung dauerte zwei Schulstunden. Das Ziel dieser Pilotstudie bestand allein darin, solche Teilaufgaben zu identifizieren, die für Erstellung einer nationalen Metrik und sich daran anschließende Vergleichsstudien (s.u.) geeignet sind.

Die Normierungsstudie mit rund 12.000 Schülern wurde im Jahr 2007 im Rahmen der *Trends in Mathematics and Science Study* (TIMSS) durchgeführt mit der Zielsetzung, nationale Skalen und Kompetenzstufenmodelle bereitzustellen. Dies wird in der Tat im Jahr 2008 der Fall sein (zu weiteren Details vgl. IQB im Druck a, b). Für Mathematik in der Sekundarstufe I werden die nationalen Skalen mit den dazu gehörigen Kompetenzstufenmodellen ebenfalls im Jahr 2008 vorliegen, in den Sprachen Deutsch, Englisch und Französisch werden sie 2009 folgen.

## 2.2 *Qualitätssicherung auf der Basis der Bildungsstandards – die Plöner Beschlüsse der KMK vom Juni 2006*

Die 16 Länder haben im Juni 2006 eine Gesamtstrategie der Qualitätssicherung im Bildungssystem verabschiedet, in der neben der nationalen Bildungsberichterstattung (vgl. Konsortium Bildungsberichterstattung 2006) Schulleistungsvergleiche auf der Basis der Bildungsstandards zukünftig eine herausragende Stellung einnehmen sollen. Hierauf wird im Folgenden eingegangen.

### *Stichprobenbasierte internationale und nationale Vergleiche:*

Bereits in den „Konstanzer Beschlüssen“ von 1997 hatte die KMK festgelegt, dass sich Deutschland regelmäßig an internationalen Schulleistungsstudien beteiligt, um auf Systemebene Informationen über Erträge schulischer Bildungsprozesse zu gewinnen. Die Folge war die Beteiligung an der PISA-Studie der OECD und an PIRLS/IGLU der *International Association for the Evaluation of Educational Achievement* (IEA). Die in beiden Programmen realisierten internationalen Vergleiche sollen Optimierungs- bzw. Steuerbedarf im allgemeinbildenden System Deutschlands aufdecken.

Mit dem Vergleich zwischen den Bundesländern, wie er erstmalig im Rahmen von PISA-E 2000 realisiert worden ist (vgl. Deutsches PISA-Konsortium 2002), ist dieses Systemmonitoring um eine nationale Komponente ergänzt worden. Der Ländervergleich trägt dem föderalen System Rechnung und schafft Transparenz hinsichtlich der Bildungserträge in den 16 Ländern der Bundesrepublik Deutschland. Auf der Basis der international entwickelten standardisierten Tests für die Bereiche Mathematik, Lesen und Naturwissenschaften werden Leistungsvergleiche durchgeführt, die teilweise um nationale Instrumente erweitert werden. Die Ländervergleiche wurden in PISA 2003 und PISA 2006 wiederholt.

So wichtig der auf internationalen Instrumenten beruhende Ländervergleich ist, weist er doch auch Beschränkungen auf (vgl. Köller 2007):

- Durch die internationalen Vorgaben bei der Testentwicklung können die Besonderheiten der Fächer in den einzelnen Teilnehmerstaaten kaum berücksichtigt werden.
- Die fremdsprachlichen Kompetenzen bleiben bei dieser Strategie bislang unberücksichtigt.
- Es gibt in den Naturwissenschaften keine Auffächerung in die drei Domänen Physik, Chemie und Biologie.
- Im Bereich der deutschen Sprache beschränken sich die internationalen Tests auf den Bereich Leseverständnis, weitere zentrale Teilkompetenzen (Hörverstehen, Schreiben, Sprachbetrachtung) werden ausgeklammert.

Unter anderem als Folge dieser Beschränkungen hat die KMK in ihren „Plöner Beschlüssen“ vom Juni 2006 eine Gesamtstrategie zur Qualitätssicherung im allgemeinbildenden Schulsystem vorgelegt, in der das nationale Bildungsmonitoring auf der Grundlage der Bildungsstandards durchgeführt werden soll. Auf der Basis von landesweit repräsentativen Stichproben soll in den Ländern festgestellt werden, welche Anteile der Schülerpopulation die Standards erreichen bzw. überschreiten. Das entsprechende Programm ist in Tabelle 2 im Überblick wiedergegeben.

<b>Tab 2: Internationale Vergleiche und Ländervergleiche auf der Basis der Bildungsstandards in den nächsten Jahren als Folge der Plöner Beschlüsse der KMK</b>			
<i>Jahr</i>	<i>Stufe</i>	<i>Internationale Studie</i>	<i>Fächer im Ländervergleich</i>
2009	Sekundarstufe I	PISA	Deutsch, Englisch, Französisch
2011	Primarstufe	IGLU/TIMSS	Deutsch, Mathematik
2012	Sekundarstufe I	PISA	Mathematik, Biologie, Chemie, Physik
2015	Sekundarstufe I	PISA	Deutsch, Englisch, Französisch
2016	Primarstufe	IGLU	Deutsch, Mathematik
2018	Sekundarstufe I	PISA	Mathematik, Biologie, Chemie, Physik

PISA: Programme for International Student Assessment; IGLU: Internationale Grundschul-Lese-Untersuchung; TIMSS; Trends in Mathematics and Science Study

Die Beschlüsse sehen im Grundschulbereich vor, dass die Überprüfung der Bildungsstandards im Fünf-Jahres-Rhythmus in der 3. Jahrgangsstufe geschehen soll, und zwar zeitlich gekoppelt an IGLU/PIRLS (vgl. Bos u.a. 2003), welche ebenfalls einem Fünf-Jahres-Rhythmus folgt. Neben diesem nationalen Monitoring wird sich Deutschland weiterhin an den internationalen Studien (PIRLS sowie TIMSS) beteiligen. Letztmalig in IGLU/PIRLS 2006 wurde der Ländervergleich auf der Basis internationaler Instrumente durchgeführt, der Bericht zu den Länderunterschieden wird Ende 2008 vorliegen. Erstmals in 2011 mit der nächsten PIRLS- und TIMSS-Erhebung wird es dann auf der Basis der Bildungsstandards zu einem Ländervergleich in den Fächern Deutsch und

Mathematik kommen. Dies wiederholt sich dann passend zum IGLU/PIRLS-Rhythmus alle fünf Jahre.

Im Bereich der Sekundarstufe I sehen die Plöner Beschlüsse vor, dass letztmalig 2006 der nationale Vergleich zwischen den Ländern auf der Basis der PISA-Instrumente erfolgte. Auch hier wird im Spätherbst 2008 der Bericht über Länderunterschiede erscheinen. Ab PISA 2009 wird dann der nationale Vergleich zwischen den Ländern auf Grundlage der Bildungsstandards durchgeführt, in 2009 für die Sprachen Deutsch, Englisch und Französisch, 2012 folgt der Vergleich in Mathematik und den Naturwissenschaften, 2015 folgen wieder die Sprachen, 2018 Mathematik und die Naturwissenschaften. Die Ländervergleiche in der Sekundarstufe I passen sich damit einem sechsjährigen Rhythmus an, getrennt für die Sprachen und Mathematik und die Naturwissenschaften. Berücksichtigt werden 8. Jahrgangsstufen (Bildungsgänge, die zum Hauptschulabschluss führen) und 9. Jahrgangsstufen (Bildungsgänge, die zum Mittleren Schulabschluss führen). Die für die ländervergleichenden Studien notwendigen Itempools werden aktuell unter der Ägide des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) aufgebaut, die entsprechenden Schritte wurden bereits beschrieben. Zentrale Vorarbeiten für das Fach Mathematik in der Sekundarstufe I wurden durch das Deutsche PISA-Konsortium geleistet (vgl. Ehmke u.a. 2006).

Der Gewinn gegenüber den bisherigen internationalen Studien wird vor allem auch darin liegen, dass in den Sprachen keine Beschränkung mehr auf das Leseverständnis stattfindet, sondern die Bereiche freies Schreiben und Hören (Deutsch, Englisch, Französisch) sowie Sprachbetrachtung und Orthografie (Deutsch) berücksichtigt werden. Inwieweit es möglich sein wird, mündliche Sprachproduktion und (in den Fremdsprachen) interkulturellen Kompetenzen im Large-Scale-Assessment zu testen, ist Gegenstand der aktuellen Forschungsarbeiten am IQB (vgl. Köller 2005).

#### *Flächendeckende Vergleichsarbeiten auf der Basis der Bildungsstandards:*

Über die Ländervergleiche hinaus hat sich in allen 16 Ländern Deutschlands ein System etabliert, in dem zu festen Zeitpunkten im Jahr ausgewählte Jahrgänge in Kernfächern in einer Gesamterhebung getestet werden. Dies gilt für den Grundschulbereich ebenso wie für die Sekundarstufe I. Das hier prominenteste Projekt ist ohne Frage „Vergleichsarbeiten in der Grundschule“ (VERA; vgl. Hosenfeld/Groß Ophoff/Bittins 2006; Helmke u.a. 2005), die mittlerweile für alle 16 Länder in den Fächern Mathematik und Deutsch durch die Universität Koblenz-Landau koordiniert und zukünftig am Ende der 3. Jahrgangsstufe durchgeführt wird. Im Fall von VERA ist es gelungen, länderübergreifend die Erträge schulischer Lehr-Lernprozesse in der Grundschule zu testen. Systematische Rückmeldungen gekoppelt mit didaktischen Kommentaren zu den Aufgaben sollen zur Unterrichtsentwicklung und Förderung der Schülerinnen und Schüler beitragen, wobei vielfach offen bleibt, wie dies genau geschehen soll.

Weit weniger einheitlich stellte sich bislang die Situation von flächendeckenden Vergleichsarbeiten in der Sekundarstufe I dar. Je nach Land wurden unterschiedliche Fächer in variierenden Jahrgängen und Schulformen (Bildungsgängen) einmal im Jahr überprüft. In den letzten Jahren ist es teilweise zu Länderverbänden gekommen, bei

denen die Aufgabenentwicklung zentralisiert, die jeweilige Durchführung aber in den Händen der einzelnen Länder liegt. Beispiele hierfür sind die länderübergreifenden Vergleichsarbeiten in den Fächern Deutsch, Mathematik und Englisch in der 6. Jahrgangsstufe (vgl. [www.vera.lernnetz.de](http://www.vera.lernnetz.de)), an denen sich sieben Länder beteiligen sowie die Lernstandserhebungen in der 8. Jahrgangsstufe im Fach Mathematik, an denen sich im Jahr 2008 12 Länder beteiligt haben. Inwieweit Schulen hier verpflichtet sind, an den Vergleichsarbeiten teilzunehmen, variiert je nach Land.

Die eingesetzten Tests sollen sich an den Standards bzw. daran angelehnten Lehrplänen orientieren. Die Items sollen hinsichtlich der jeweils getesteten Kompetenzen spezifiziert werden und flankierende didaktische Kommentare sollen analog zum Vorgehen in der Grundschule (s.o.) zur Unterrichtsentwicklung beitragen, wobei auch hier offen bleibt, wie dies genau geschehen soll. Hinsichtlich ihrer inhaltlichen und psychometrischen Güte variieren die Vergleichsarbeiten/Lernstandserhebungen erheblich. Beispielsweise werden in den Vergleichsarbeiten in der Grundschule nur pilotierte Aufgaben, deren psychometrische Kriterien bekannt sind, verwendet. In anderen Programmen (z.B. Orientierungsarbeiten in Sachsen) wird auf die Pilotierung der Aufgaben zugunsten größerer Unterrichtsnähe verzichtet. Gemeinsam ist allen diesen Testuntersuchungen, dass sie – im Gegensatz zu den internationalen Studien und den vom IQB verantworteten Ländervergleichen – nicht durch eigens ausgebildetes Personal administriert und ausgewertet werden, sondern vor Ort durch die Lehrkräfte der jeweiligen Schule. Anders wären flächendeckende Erhebungen gar nicht realisierbar. Dieses Vorgehen entspricht der Funktion der Vergleichsarbeiten als einer Art formativer Evaluation, impliziert aber, dass wissenschaftlich seriöse Vergleiche zwischen Schulen oder gar Ländern auf dieser Basis nicht vorgenommen werden können.

Das Kernproblem aller flächendeckenden Vergleichsarbeiten besteht darin, dass sie bislang keinen expliziten Bezug zu den Bildungsstandards in dem Sinne haben, dass auf einem nationalen Maßstab, der durch die operationalisierten Standards vorgegeben ist, Ergebnisse an die Schulen zurückgemeldet werden. Dies ist bislang unmöglich, da die entsprechenden Skalen und Kompetenzmodelle durch das IQB noch nicht bereitgestellt werden können. Stattdessen behilft man sich mit rudimentären Kompetenzstufenmodellen (im Falle der Vergleichsarbeiten in der Grundschule vgl. Helmke u.a. 2005) oder mit sozialen Vergleichen, in denen die einzelne Schule/Klasse am Landesdurchschnitt gemessen wird. Teilweise wird dabei versucht, die Eingangsselektivität der Schülerschaft zu korrigieren (sogenannte faire Vergleiche; vgl. Helmke u.a. 2005).

Mit den Plöner Beschlüssen der KMK von 2006 ist versucht worden, hier Abhilfe zu schaffen und die flächendeckenden Vergleichsarbeiten an die Bildungsstandards anzubinden. Die Beschlüsse sehen vor, dass ab 2010, wenn entsprechende normierte Items und die dazu gehörigen Kompetenzstufenmodelle auf der Basis der Bildungsstandards vorliegen, in allen Ländern flächendeckende Vergleichsarbeiten durchgeführt werden, die eine explizite Verortung der Leistungen auf den nationalen Skalen ermöglichen. In der Grundschule sollen die Vergleichsarbeiten am Ende der 3. Jahrgangsstufe in Deutsch und Mathematik stattfinden, in der Sekundarstufe I sollen sie am Ende der 8. Jahrgangsstufen in den Fächern Deutsch, Mathematik und erste Fremdsprache realisiert

werden. Damit sollen Schulen und Klassen die Gelegenheit erhalten zu überprüfen, wo die eigenen Schülerinnen und Schüler im Hinblick auf die Erreichung der in den Standards definierten Ziele stehen.

### **3. Herausforderungen und Risiken der Qualitätssicherung auf der Basis der Bildungsstandards**

Mit den Bildungsstandards und dem damit verbundenen System der stichprobenbasierten und flächendeckenden Leistungserhebungen ist die Hoffnung verbunden, sowohl auf Systemebene als auch auf der Ebene der Einzelschule oder Einzelklasse Informationen zur Verbesserung von institutionalisierten Bildungsprozessen zu gewinnen. Diese Hoffnung hat in der Vergangenheit insbesondere die Vergleichsarbeiten in der Grundschule und Sekundarstufe I getragen, ohne dass es allerdings bislang Evidenz dafür gibt, wonach sich erkennbare Veränderungen im Unterrichtsgeschehen als direkte Folge der Vergleichsarbeiten ergeben haben. Im Folgenden sollen daher Herausforderungen im Bereich der Professionalisierung von Lehrkräften kurz skizziert werden. Daran anschließend werden Probleme skizziert, die sich ergeben, wenn Bildungsstandards in ein Testsystem münden, das an den Schulen als Bedrohung wahrgenommen werden kann.

#### *3.1 Herausforderungen für die Schul- und Unterrichtsentwicklung*

In ihrer Expertise haben Klieme u.a. (2003) immer wieder betont, dass Bildungsstandards nur dann zur Unterrichtsentwicklung beitragen können, wenn auf ihrer Basis neben einem System der Qualitätssicherung auch eines der Qualitätsentwicklung im Bildungssystem etabliert wird. Der Glaube, allein die Testung würde zu einer stetigen Unterrichtsentwicklung führen – dieser Glaube legitimiert im Übrigen aktuell in vielen Ländern die Durchführung von Vergleichsarbeiten – ist unrealistisch, vielmehr können Standards und das an sie angebundene System der flächendeckenden Vergleichsarbeiten und stichprobenbasierten *Assessments* nur dann Wirkung entfalten, wenn sie durch entsprechende Unterstützungssysteme flankiert werden (vgl. z.B. Hosenfeld u.a. 2006). Und in der Tat sind mit den Bildungsstandards weit reichende Konsequenzen für die Etablierung einer neuen Aufgaben- und Unterrichtskultur verbunden. Blum (2006) schreibt hierzu: „Nur ein Unterricht, der den eigenaktiven Erwerb von Kompetenzen in lernförderlicher Arbeitsatmosphäre in den Mittelpunkt aller Lehr-/Lernanstrengungen stellt, wird Lernenden überhaupt die Chance bieten, die in den Standards formulierten Kompetenzerwartungen auch tatsächlich zu erfüllen ... Etwas konkreter bedeutet ‚*standardorientiertes Unterrichten*‘: Jede einzelne Unterrichtsstunde und jede Unterrichtseinheit muss sich daran messen lassen, inwieweit sie zur Weiterentwicklung inhaltsbezogener und allgemeiner Schüler-Kompetenzen beiträgt... Die wichtigste Frage ist nicht ‚Was haben wir durchgenommen?‘, sondern ‚Welche Vorstellungen, Fähigkeiten und Einstellungen sind entwickelt worden?‘“ (ebd. S. 15ff.). Die Entwicklung von Vorstel-

lungen, Fähigkeiten und Einstellungen ist dabei unmittelbar an den intelligenten Einsatz von Aufgaben im Unterricht gebunden. Das Erkennen, welches Potenzial eine Aufgabe beim Kompetenzaufbau, bei der Konsolidierung von Kompetenzen und beim Evaluieren von Kompetenzen hat, stellt eine zentrale professionelle Aufgabe für die Fachgruppen an den Schulen dar. Sie lässt sich eigentlich nur kooperativ erfolgreich lösen, indem Lehrkräfte, beispielsweise in den Fachkonferenzen gemeinsam über kompetenzorientierte Aufgaben und deren Einsatz im Unterricht reflektieren.

Im Sinne dieser Kompetenzorientierung ist in Abbildung 2 eine Mathematikaufgabe illustriert, bei der die allgemeine mathematische Kompetenz „Modellieren“ im Vordergrund steht (vgl. Leiß/Blum 2006):

## Tanken

Herr Stein wohnt in Trier, 20 km von der Grenze zu Luxemburg entfernt. Er fährt mit seinem VW Golf zum Tanken nach Luxemburg, wo sich direkt hinter der Grenze eine Tankstelle befindet. Dort kostet der Liter Benzin nur 1,05 €, im Gegensatz zu 1,30 € in Trier. Lohnt sich die Fahrt für Herrn Stein? Begründe deine Antwort.

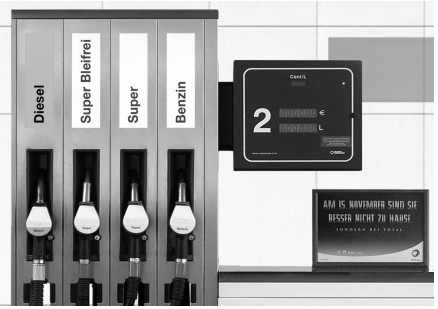


Abb 2: *Beispielaufgabe zum kompetenzorientierten Unterricht im Fach Mathematik: Allgemeine Kompetenz: Modellieren (vgl. Leiß & Blum, 2006, S. 42)*

Fünf Schritte führen bei dieser Aufgabe zur richtigen Lösung (s. ausführlicher in Leiß/Blum 2006, S. 42): (1) Zunächst muss die Problemsituation verstanden werden, (2) anschließend muss das Problem strukturiert und präzisiert werden. (3) Es folgt die Mathematisierung, d.h. die Transformation in mathematische Terme, die (4) ausgerechnet werden müssen. (5) Schließlich muss das Ergebnis interpretiert und überprüft werden. Deutlich wird, dass es sich hierbei um eine anwendungsbezogene, das mathematische Verständnis fördernde Aufgabe handelt. Um diese Chancen nutzen können, sind verschiedene Schritte der Lehrerprofessionalität notwendig. Hierzu zählen:

Eigene Professionalisierung der Lehrkräfte im Hinblick auf die Philosophie und die Inhalte der Standards. Hier liegen mittlerweile vielfältige Materialien vor (vgl. z.B. Artelt/Riecke-Baulecke 2004; Blum u.a. 2006), die ergänzt werden müssen durch entsprechende Fortbildungsveranstaltungen, was im Übrigen in vielen Ländern auch der Fall ist.

Flächendeckende Angebote der dezentralen Lehrerfortbildung, wie sie im Rahmen des Modellversuchs Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts (SINUS) bzw. SINUS-Transfer umgesetzt wurden (vgl. Ostermeyer u.a. 2004). In SINUS/SINUS-Transfer wurden kooperative Arbeitsformen entwickelt und



besonderer Wert auf eine kognitiv aktivierende, kompetenzorientierte Unterrichts gelegt. Am SINUS-Transfer-Projekt, das seit August 2004 auch in Grundschulen durchgeführt wird, beteiligen sich immerhin 13 Länder, sodass für die meisten Schulleitungen auch Ansprechpartnerinnen und Ansprechpartner im jeweiligen Land vorliegen.

Forcierung der Auseinandersetzung in den Fachgruppen/Fachkonferenzen mit den Bildungsstandards. Die Standards werden nur dort in den Unterricht einfließen, wo permanent die Beschäftigung mit der Kompetenzorientierung vom Kollegium eingefordert wird. Auch dies muss in ein System der dezentralen Professionalisierung eingebunden sein.

Aufforderung an die Lehrkräfte, Eltern kompetent über die Standards und die damit verbundenen Implikationen für Qualitätssicherung und -entwicklung zu informieren. Mit den Bildungsstandards ist die explizite Aufforderung verbunden, Rechenschaft über das Erreichte abzulegen. Dies soll zukünftig in flächendeckende Vergleichsarbeiten münden, welche sich an den Standards orientieren. Solche Vergleichsarbeiten sorgen jedoch häufig auf Lehrkraft- wie Elternseite zur Verunsicherung. Eine Einbettung dieser Arbeiten in die Philosophie der Qualitätssicherung auf der Basis von Bildungsstandards muss dementsprechend in die Elternschaft kommuniziert werden, damit über die Sinnhaftigkeit des Unterfangens reflektiert werden kann.

Gelingt die breite Umsetzung dieser Maßnahmen, so steigen die Chancen, dass die mit den Bildungsstandards gekoppelten Visionen für einen erfolgreichen Unterricht in den Schulen Realität werden. Bildungsstandards werden aber weitgehend wirkungslos bleiben, wenn der Glaube, allein ihre Überprüfung steigere die Qualität von Unterricht, fortwährt.

### 3.2 *Risiken*

In diesem abschließenden Kapitel soll noch auf ein erhebliches Risiko aufmerksam gemacht werden, das bereits in der Expertise von Klieme u.a. (2003) zur Einführung von Bildungsstandards aufgegriffen wurde und dort zu dem expliziten Hinweis führte, dass auf der Basis der Bildungsstandards kein breites System des *High Stakes Testings* etabliert werden sollte. Die Folgen eines solchen Systems könnten zum Phänomen der so genannten *Inflating Gain Scores* (vgl. u.a. Green/Winters/Forster 2003; Koretz 2002) führen, die bei Vergleichsarbeiten/Lernstandserhebungen in den USA auftreten. Bei diesem Phänomen steigen die Leistungen in den flächendeckenden Vergleichsarbeiten innerhalb der jeweiligen Bundesstaaten von Jahr zu Jahr an, ohne dass sich auch entsprechende Zuwächse im stichprobenbasierten *National Assessment of Educational Progress* (NAEP) zeigen. Wo *Score Inflation* auftritt, muss die Validität der Vergleichsarbeiten/Lernstandserhebungen in Zweifel gezogen werden.

Mit der Unterzeichnung des *No-Child-Left-Behind*-Gesetzes im Jahre 2001 ist in den USA eine Grundlage geschaffen worden, die es erlaubt Schulen auf Grund der Leistungen ihrer Schülerinnen und Schüler in flächendeckenden Vergleichsarbeiten zu belohnen oder zu bestrafen. In diesem System des so genannten *High Stakes Testings* ist es

demnach für Schulen existenziell, erfolgreich in den jeweiligen Testungen abzuschneiden. In den USA hat dies (a) zu einer Welle des *Teaching to the Test* und zu Betrugsversuchen geführt (vgl. Green u.a. 2003) und (b) zu einer systematischen Erforschung der Effekte, die solche Maßnahmen haben, bei denen Schulen die Ergebnisse weniger für die eigene professionelle Entwicklung nutzen können, sondern vielmehr Sanktionen erwarten müssen.

Sehr prominent sind hier die Arbeiten von Koretz (unter anderem 2002), in denen die Leistungen in *High Stakes Tests* mit denen im stichprobenbasierten NAEP (*Low Stakes Test*), das keine Rückwirkungen auf die Einzelschule hat, auf Bundesstaatenebene verglichen werden. Sofern die erfassten Testleitungen realistische Abbilder der erreichten Kompetenzniveaus auf Seiten der Schülerinnen und Schüler sind, sollten sich (a) vergleichbare Trends und (b) hohe Korrelationen auf Schulebene zwischen den Leistungen in *Low* und *High Stakes Tests* ergeben. Dagegen sollten sich deutliche Abweichungen ergeben, d.h. divergierende Trends und geringe Korrelationen, wenn es verschiedentlich in Schulen zu Praktiken kommt, mit denen versucht wird, die Werte der Schülerinnen und Schüler in *High Stakes Tests* zu maximieren. Koretz und Barron (1998) haben zur Illustration dieser Effekte Trendanalysen für Kentucky aus dem *Kentucky Instructional Results Information System (KIRIS) Testing Program (High Stakes Tests)* mit solchen aus dem NAEP (*Low Stakes Tests*) verglichen. Berücksichtigt wurden die Kompetenzentwicklungen von Viert- und Achtklässlern in Mathematik zwischen 1992 und 1996. In KIRIS ergaben sich im untersuchten Zeitraum Leistungszuwächse, die in der 4. Jahrgangsstufe einer Effektstärke von  $d = .61$  entsprachen, in der 8. Jahrgangsstufe einer Effektstärke von  $d = .52$ . In NAEP lagen die entsprechenden Effektstärken bei  $d = .17$  und  $d = .13$ .

Green u.a. (2003) haben in verschiedenen US-Bundesstaaten auf Schulebene die Korrelationen zwischen den Leistungstrends in *Low* und *High Stakes Tests* analysiert. Im optimalen Szenario, in dem es auf Seiten der Schulen differenzielle Fördereffekte gäbe, sollten diese Korrelationen nahe 1 liegen, d.h. erfolgreiche Schulen sollten sowohl in *Low* wie in *High Stakes Tests* günstige Trends aufweisen. Die Analysen von Green u.a. (2003) ergaben je nach Bundesstaat und Distrikt sehr unterschiedliche Ergebnisse. Während in Florida die Trends in *Low* und *High Stakes Tests* sehr hoch korrelierten ( $r = .71$ ), lag die Korrelation in Virginia nur bei  $r = .17$ , im Chicago School Distrikt ergab sich gar eine Nullkorrelation, d.h. die Leistungstrends in *Low* und *High Stakes Tests* waren vollständig entkoppelt.

Koretz (2002) nennt verschiedene Ursachen, welche solche unerwarteten Befunde, die jegliche Validität von *Assessment*-Programmen in Frage stellen, erklären können:

*Umwidmung von Unterrichtszeit und -material (Teaching to the Test)*: Um Schülerinnen und Schüler auf die *High Stakes* Testung optimal vorzubereiten, werden im Unterricht primär Übungssituationen hergestellt, in denen das Aufgabenmaterial eine hohe Ähnlichkeit zu den im Test verwendeten Aufgaben aufweist. Darüber hinaus kann es auch zu einer systematischen Erhöhung der Unterrichtsstunden genau in den Fächern kommen, für die in absehbarer Zeit eine *High Stakes* Testung ansteht. Stecher und Barron (1999) untersuchten im Bundesstaat Kentucky Unterrichtszeiten in verschiedenen

Fächern und Jahrgangsstufen. Dabei zeigte sich, dass viele Schulen in der vierten Jahrgangsstufe, in der die Naturwissenschaften im *High Stakes Testing* überprüft wurden, die Wochenstundenzahlen in den Naturwissenschaften deutlich erhöhten (5.2 Wochenstunden verglichen zu 3.5 Wochenstunden in der 5. Jahrgangsstufe). In Mathematik, wo das *High Stakes Testing* in der 5. Jahrgangsstufe durchgeführt wurde, stieg die Zahl der Wochenstunden von 4.9 (Klasse 4) auf 6.4 (Klasse 5). Bei diesem Vorgehen ist zu erwarten, dass die Schülerinnen und Schüler in der Tat in den Tests höhere Kompetenzniveaus zeigen, diese aber nicht nachhaltig eingehalten werden, da Zeit und Material nach dem *High Stakes Testing* wieder zurückgefahren werden.

*Betrug (Cheating)*: In den USA besteht mittlerweile hinreichend Evidenz (vgl. Koretz 2002), wonach Lehrerinnen und Lehrer in *High Stakes Testings* Einfluss auf die Testbearbeitung nehmen, in dem sie (a) Lösungen vorsagen, (b) Lösungen verbessern, (c) Aufgaben, die später im Test auftauchen, üben lassen und (d) Aufgaben umformulieren, um die Lösungen zu erleichtern.

*Coaching*: Hierunter versteht man eine Anpassung des Unterrichts an die kognitiven Anforderungen der Aufgaben in *Assessment*-Programmen. Dies ist etwas anderes als bloßes *Teaching to the Test*, will es doch Kompetenzen fördern und kann daher durchaus wünschenswert sein. Bezogen beispielsweise auf die Bildungsstandards in der ersten Fremdsprache ist es sicherlich zu begrüßen, wenn mehr Unterrichtszeit für kommunikative Fertigkeiten in den Bereichen des Hörens und Sprechens reserviert wird. Gleichzeitig wächst dadurch die Gefahr, dass Bereiche, die später nicht Gegenstand der Testung sind, vernachlässigt werden.

Ohne Frage stellen solche Effekte Systeme der testbezogenen Qualitätssicherung in Frage und wann immer man plant, Maßnahmen des *High Stakes Testing* zu etablieren, wird man mit entsprechenden Nebenwirkungen rechnen müssen. In Deutschland beginnen einige Länder wie Nordrhein-Westfalen und Bayern damit, die Schulen, die in den flächendeckenden Vergleichsarbeiten am besten abgeschnitten haben, zu prämiieren bzw. zu veröffentlichen. Andere Länder (Brandenburg) lassen die Ergebnisse aus Vergleichsarbeiten mit einem erheblichen Gewicht in die Halbjahreszeugnisnote einfließen. Solche Maßnahmen erhöhen vermutlich die Chance für das Auftreten von *Inflating Gain Scores* und es stellt sich die Frage, ob man mit der Etablierung eines Qualitätssicherungssystems zur Förderung von Kompetenzen auf Seiten der Schülerinnen und Schüler beitragen möchte oder vor Ort in den Schulen eine Kultur der Testwertmaximierung aufbauen möchte.

#### 4. Abschluss

Auf Grundlage der in 2003 und 2004 verabschiedeten Bildungsstandards ist in Deutschland eine Gesamtstrategie der Qualitätssicherung im allgemeinbildenden Schulsystem entwickelt worden, die aktuell umgesetzt wird, indem Aufgabenpools generiert werden, welche stichprobenbasierte und flächendeckende Assessments in Schulen der Primarstufe und Sekundarstufe I erlauben werden. Längerfristig wird die gymnasiale Oberstufe in dieses System der Qualitätssicherung eingebunden werden. Ein qualitätssicherndes System im Bildungswesen auf der Basis von Bildungsstandards, dies sollte deutlich geworden sein, führt nicht automatisch zu einer Steigerung der Unterrichtsqualität und als deren Folge zu höheren Lernerträgen auf Seiten der Schülerinnen und Schüler. Vielmehr müssen Maßnahmen der Qualitätssicherung durch Unterstützungsmaßnahmen flankiert werden. Dies umfasst die dezentrale Professionalisierung von Schulleitungen und Lehrkräften auf der Basis der Standards ebenso wie zusätzliche, keineswegs kostenneutrale Angebote für benachteiligte Kinder und Jugendliche. Will man Bildungsstandards und die daran angebundene Qualitätssicherung für die Schul- und Unterrichtsentwicklung konstruktiv nutzen, so sind zum jetzigen Zeitpunkt alle Maßnahmen in Richtung *High Stakes Testing* eher kontraindiziert. Wünschenswert ist es vielmehr, Schulen Instrumente der Selbstevaluation an die Hand zu geben, mit denen sie standardbasiert Leistungsstände ihrer Schülerinnen und Schüler feststellen und bei Bedarf Fördermaßnahmen einleiten können.

#### Literatur:

- Artelt, C./Riecke-Baulecke, T. (2004): Bildungsstandards: Fakten, Hintergründe, Praxistipps. München: Oldenbourg.
- Blum, W. (2006): Bildungsstandards Mathematik: Einführung. In: Blum, W./Drüke-Noe, C./Hartung, R./Köller, O. (Hrsg.): Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen. Berlin: Cornelsen-Scriptor, S. 14–32.
- Blum, W./Drüke-Noe, C./Hartung, R./Köller, O. (Hrsg.) (2006): Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsideen und Fortbildungsmöglichkeiten. Berlin: Cornelsen-Scriptor.
- Bos, W./Lankes, E.-M./Prenzel, M./Schwippert, K./Walther, G./Valtin, R. (Hrsg.) (2003): Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Münster: Waxmann.
- Bos, W./Hornberg, S./Arnold, K.-H./Faust, G./Fried, L./Lankes, E.-M./Schwippert, K./Valtin, R. (Hrsg.) (2007): IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Münster: Waxmann.
- Bremerich-Vos, A. (2003): Bildungsstandards, Kompetenzstufen, Kernlehrpläne, Parallel- und Vergleichsarbeiten. In: Deutschunterricht 5, S. 4–10.
- Deutsches PISA-Konsortium (Hrsg.) (2001): PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske + Budrich.
- Deutsches PISA-Konsortium (Hrsg.) (2002): PISA 2000. Die Länder der Bundesrepublik Deutschland im Vergleich. Opladen: Leske + Budrich.

- Deutsches PISA-Konsortium (Hrsg.) (2007): PISA 2006: Der Bildungsstand der Jugendlichen in Deutschland – Die Ergebnisse der dritten internationalen Vergleichsstudie. Münster: Waxmann.
- Ehmke, T./Leiß, D./Blum, W./Prenzel, M. (2006): Entwicklung von Testverfahren für die Bildungsstandards Mathematik. In: *Unterrichtswissenschaft* 34, S. 220–238.
- Green, J.P./Winters, M.A./Forster, G. (2003): Testing high stakes tests: Can we believe the results of accountability tests? Civic Report. New York: Center for Civic Innovation.
- Hartig, J. (2007): Skalierung und Definition von Kompetenzniveaus. In: Beck, B./Klieme, E. (Hrsg.): *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Ergebnisse Band 1*. Weinheim: Beltz, S. 83–99.
- Helmke, A./Hosenfeld, I./Ophoff, J. G./Hochweber, J./Isaac, K./Koch, U./Scherthan, F./Wagner, S. (2005): VERA 2004: Erste Ergebnisse des Ländervergleichs. Landau: Universität Koblenz-Landau.
- Hosenfeld, I./Groß-Ophoff, J./Bittins, P. (2006): Vergleichsarbeiten und Schulentwicklung. *Schulmanagement-Handbuch Band 118*. München: Oldenbourg.
- IQB (im Druck a): Bildungsstandards: Kompetenzen überprüfen: Grundschule Mathematik 3./4. Schuljahr. Berlin: Cornelsen.
- IQB (im Druck b): Bildungsstandards: Kompetenzen überprüfen: Grundschule Deutsch 3./4. Schuljahr. Berlin: Cornelsen.
- Klieme, E./Avenarius, H./Blum, W./Döbrich, P./Gruber, H./Prenzel, M./Reiss, K./Riquarts, K./Rost, J./Tenorth, H.E./Vollmer, J. (2003): Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Reihe Bildungsreform, Band 1. Bonn: Bundesministerium für Bildung und Forschung.
- Klieme, E./Avenarius, H./Baethge, M./Döbert, H./Hetmeier, H.-W./Meister-Scheufelen, G./Rauschenbach, T./Wolter, A. (2007): Grundkonzeption der Bildungsberichterstattung für Deutschland. In: Krüger, H.-H./Rauschenbach, T./Sander, U. (Hrsg.): *Bildungs- und Sozialberichterstattung*. 6. Beiheft der Zeitschrift für Erziehungswissenschaft. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 129–145.
- KMK (2004): Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss. München: Luchterhand.
- KMK (2006): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. München: Luchterhand.
- Köller, O./Watermann, R./Trautwein, U./Lüdtke, O. (Hrsg.) (2004): Wege zur Hochschulreife in Baden-Württemberg. TOSCA – Eine Untersuchung an allgemeinbildenden und beruflichen Gymnasien. Opladen: Leske+Budrich.
- Köller, O. (2005): Neue Besen kehren gut: Das Institut zur Qualitätsentwicklung im Bildungswesen soll die nationalen Bildungsstandards in Deutschland überprüfen. In: *Psychologie in Erziehung und Unterricht* 52, S. 281–286.
- Köller, O. (2007): Bildungsmonitoring und Vergleichsarbeiten in Deutschland. In: Labudde, P. (Hrsg.): *Bildungsstandards am Gymnasium*. Bern: h.e.p., S. 65–73.
- Konsortium Bildungsberichterstattung (2006): *Bildung in Deutschland. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Bielefeld: Bertelsmann.
- Koretz, D. (2002): Limitations in the use of achievement tests as measures of educators' productivity. In: *Journal of Human Resources* 37, S. 752–777.
- Koretz, D./Barron, S.I. (1998): *The Validity of Gains on the Kentucky Instructional Results Information System (KIRIS)*. Santa Monica: RAND.
- Leiß, D./Blum, W. (2006): Beschreibung zentraler mathematischer Kompetenzen. In: Blum, W./Drückeno, C./Hartung, R./Köller, O. (Hrsg.): *Bildungsstandards Mathematik: konkret*. Berlin: Cornelsen-Scriptor, S. 33–50.
- Ostermeier, C./Carstensen, C.H./Prenzel, M./Geiser, H. (2004): Kooperative unterrichtsbezogene Qualitätsentwicklung in Netzwerken. Ausgangsbedingungen für die Implementation im BLK-Modellversuchsprogramm SINUS. In: *Unterrichtswissenschaft* 32, S. 215–237.
- Rost, J. (2004): *Lehrbuch Testtheorie/Testkonstruktion*. Bern: Huber.

- Rupp, A.A./Vock, M./Harsch, C./Köller, O. (2008): Developing standards-based assessment tasks for English as a first foreign language – Context, processes, and outcomes in Germany. Münster: Waxmann
- Spinner, K.H. (2005): Der standardisierte Schüler. In: Didaktik Deutsch 18, S. 4–13.
- Stecher, B.M./Barron, S. I. (1999): Quadrennial Milepost Accountability Testing in Kentucky. CSE Technical Report No. 505. Los Angeles: Center for the Study of Evaluation, University of California.
- Weinert, F.E. (2001): Vergleichende Leistungsmessung in Schulen – Eine umstrittene Selbstverständlichkeit. In: Weinert F.E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim: Beltz, S. 17–31.

*Anschrift des Autors:*

Prof. Dr. Olaf Köller, Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im Bildungswesen, Unter den Linden 6, 10099 Berlin, Email: iqboffice@iqb.hu-berlin.de