

Köller, Olaf

## **Bildungsstandards - Verfahren und Kriterien bei der Entwicklung von Messinstrumenten**

*Zeitschrift für Pädagogik 54 (2008) 2, S. 163-173*

urn:nbn:de:0111-opus-43418

in Kooperation mit / in cooperation with:

# **BELTZ**

<http://www.beltz.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)

## Inhaltsverzeichnis

### Thementeil: Bildungsstandards außerhalb der „Kernfächer“

*Heinz-Elmar Tenorth*

Bildungsstandards außerhalb der „Kernfächer“. Herausforderungen für den Unterricht und die fachdidaktische Forschung. Zur Einleitung in den Thementeil ..... 159

*Olaf Köller*

Bildungsstandards – Verfahren und Kriterien bei der Entwicklung von Messinstrumenten ..... 163

*Sabine Krause/Roumiana Nikolova/Henning Schluß/Thomas Weiß/Joachim Willems*

Kompetenzerwerb im evangelischen Religionsunterricht. Ergebnisse der Konstruktvalidierungsstudie der DFG-Projekte RU-Bi-Qua/KERK ..... 174

*Detlev Leutner*

Metamorphose eines Forschungsprojektes. Ein Kommentar zum Beitrag von Krause et al. über den „Kompetenzerwerb im evangelischen Religionsunterricht – Ergebnisse der Konstruktvalidierungsstudie der DFG-Projekte RU-Bi-Qua / KERK“ ..... 189

*Martin Rothgangel*

Bildungsstandards für den Religionsunterricht. Zur fachdidaktischen Konsistenz des Berliner Forschungsprojekts ..... 194

*Waltraud Schreiber*

Ein Kompetenz-Strukturmodell historischen Denkens ..... 198

*Michael Sauer*

Historisches Denken und Geschichtsunterricht. Ein Kommentar zum Beitrag von Waltraud Schreiber ..... 213

*Bernd Schönemann*

Bildungsstandards und Geschichtsunterricht. Ein Kommentar zu Waltraud Schreiber und Michael Sauer ..... 218

*Eckhard Klieme/Katrin Rakoczy*

Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts ..... 222

<i>Deutscher Bildungsserver</i>	
Linktipps zum Thema „Bildungsstandards in der geisteswissenschaftlichen Fächergruppe an Sekundarschulen“ .....	238
 <i>Allgemeiner Teil</i>	
<i>Sarah Hitzler/Heinz Messmer</i>	
Gespräche als Forschungsgegenstand in der Sozialen Arbeit .....	244
<i>Hildegard Macha/Monika Witzke</i>	
Familie und Gender. Rollenmuster und segmentierte gesellschaftliche Chancen ....	261
 <i>Besprechungen</i>	
<i>Peter Faulstich</i>	
Michael Göhlich/Jörg Zirfas: Lernen. Ein pädagogischer Grundbegriff .....	279
<i>Alfred Schäfer</i>	
Norbert Ricken: Die Ordnung der Bildung. Beiträge zu einer Genealogie der Bildung .....	280
<i>Karsten Kenklies</i>	
Karl Ernst Nipkow: Der schwere Weg zum Frieden. Geschichte und Theorie der Friedenspädagogik von Erasmus bis zur Gegenwart .....	284
<i>Martin Rothland</i>	
Uwe Schaarschmidt/Ulf Kieschke (Hrsg.): Gerüstet für den Schulalltag. Psychologische Unterstützungsangebote für Lehrerinnen und Lehrer .....	287
<i>Rudi Heidemann</i>	
Joachim Bauer: Lob der Schule. Sieben Perspektiven für Schüler, Lehrer und Eltern .....	291
<i>Marcelo Caruso</i>	
Susanne Weber/Susanne Maurer (Hrsg.): Gouvernamentalität und Erziehungswissenschaft. Wissen – Macht – Transformation .....	294
 <i>Dokumentation</i>	
Pädagogische Neuerscheinungen .....	297

Olaf Köller

## Bildungsstandards – Verfahren und Kriterien bei der Entwicklung von Messinstrumenten

### 1. Thema und Problemstellung

Die ständige Konferenz der Kultusminister in der Bundesrepublik Deutschland (KMK) hat als Reaktion auf die enttäuschenden Ergebnisse deutscher Schülerinnen und Schüler in PISA-2000 (Deutsches PISA-Konsortium 2001) 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. Anfang 2003 lag mit der durch das Bundesministerium für Bildung und Forschung (BMBF) geförderten Expertise „Zur Einführung nationaler Bildungsstandards“ (Klieme u.a. 2003) ein Dokument vor, auf dessen Grundlage sich Bildungsstandards entwickeln ließen. Bis Ende 2004 wurden auf dieser Basis für den Grundschulbereich KMK-Standards in den Fächern Deutsch und Mathematik erarbeitet. In der Sekundarstufe I wurden als weitere Fächer die erste Fremdsprache (Englisch oder Französisch) und die Naturwissenschaften (Biologie, Chemie, Physik) berücksichtigt. Im Oktober 2007 wurde schließlich im Plenum der KMK beschlossen, Standards für das Ende der gymnasialen Oberstufe – erneut für die Fächer Deutsch, Mathematik, Englisch/Französisch, Biologie, Chemie und Physik zu erarbeiten.

Nicht ganz zu Unrecht ist als Folge der Beschränkung auf diese sieben Fächer kritisiert worden, dass sie damit zu „Kernfächern“ geadelt, auf der anderen Seite Fächer, die nicht in den Katalog der Kernfächer aufgenommen wurden, zu randständigen, unwichtigen, „weichen“ oder auch „unnützen“ Fächern abgestuft [wurden]“ (Benner u.a. 2007, S. 141). Nur folgerichtig haben Fachverbände, Fachdidaktiken und Erziehungswissenschaften selbst mit der Erarbeitung von Bildungsstandards in den „unnützen“ Fächern begonnen. Gelungene Beispiele hierfür sind die Standards für die Fächer Religion (Benner u.a. 2007; vgl. auch den Beitrag von Krause u.a. in diesem Heft) und Geographie (Deutsche Gesellschaft für Geographie 2007).

Teilweise unreflektiert ist in den Entwürfen zu Bildungsstandards geblieben – dies gilt natürlich auch für die KMK-Standards – welche Funktionen Bildungsstandards erfüllen sollen, mit anderen Worten wie sie als Steuerungsinstrumente im allgemein bildenden Schulsystem wirksam werden können, sei es in Form von Lernstandserhebungen, welche die Erreichung der Standards überprüfen, oder sei es in Form von Visionen, besser: konkreten Vorschlägen für einen standard- bzw. kompetenzorientierten Unterricht. Aus einer stärker theoretischen Perspektive ist zudem zu bemängeln, auch diese Kritik gilt in weiten Teilen für die KMK-Standards, dass die hinter den Bildungsstan-

dards liegenden Kompetenzmodelle nicht hinreichend elaboriert sind und ungeklärt bleibt, wie die Kompetenzen in didaktischen, erziehungswissenschaftlichen und psychologischen Modellen zu verorten sind. Ohne solch eine theoretische Verankerung und die exakte Beschreibung dessen, was man unter Kompetenzen versteht, ist aber die valide Messung bzw. Operationalisierung der in den Standards beschriebenen Leistungserwartungen unmöglich.

Der vorliegende Beitrag möchte diese Kritik konstruktiv wenden und Anregungen geben, welche Arbeiten bei der Entwicklung und Operationalisierung von Bildungsstandards nötig sind, um empirisch gesicherte Kompetenzmodelle zu etablieren und mit Hilfe der generierten Testinstrumente die Erreichung der Standards zu überprüfen. Viele dieser Gedanken wurden bereits in der Expertise von Klieme u.a. (2003) geäußert aber in späteren Arbeiten an Bildungsstandards nicht umgesetzt. Insgesamt wurden dort und werden hier vor allem Erwartungen an die Fachdidaktiken formuliert. Sie erfahren angesichts der Arbeit an Bildungsstandards, welche Analyse-, Konstruktions- und empirische Forschungskompetenz sie besitzen.

Beginnen werden soll mit dem Versuch (2.), das Beziehungsgefüge zwischen Bildungsstandards, Kompetenzen und wissenschaftliche Konstrukten zu klären. Anschließend (3.) soll am Beispiel der KMK-Standards für das Fach Englisch am Ende der Sekundarstufe I erläutert werden, unter welchen Voraussetzungen Standards als Instrumente der Qualitätssicherung geeignet sind, aber auch, unter welchen Umständen sie Ziele formulieren, die sich der Messung dadurch entziehen, dass die Kompetenzen selbst theoretisch zu unpräzise bzw. zu breit gefasst sind. Danach soll kurz auf die Setzung von Niveaustufen (Standard-Setting) eingegangen und deutlich gemacht werden, dass die Setzung solcher Stufen ein auf die Erzeugung von Konsens zielender, aber kein theoretisch fundierter Akt ist. Abschließend (4.) werden drei Schlussfolgerungen für die Erarbeitung von Standards, Kompetenzmodellen und Messinstrumenten gegeben.

## **2. Bildungsstandards, Kompetenzen und Konstrukte**

In ihrer Expertise von 2003 haben Klieme u.a. in verschiedenen Kapiteln ausgeführt, wie sich Bildungsstandards und Kompetenzen zueinander verhalten. Danach benennen Bildungsstandards fachspezifisch „die wesentlichen Ziele pädagogischer Arbeit, ausgedrückt als erwünschte Lernergebnisse bzw. Kompetenzen der Schülerinnen und Schüler“ (Klieme u.a. 2003, S. 13) zu bestimmten Zeitpunkten ihrer Bildungsbiographien. In den vorliegenden Arbeiten zu Bildungsstandards schlägt sich diese Charakterisierung darin nieder, dass Standards als so genannte *Can do Statements* formuliert sind. Beispielsweise heißt es in den Standards für den Mittleren Abschluss im Fach Mathematik im Kompetenzbereich Modellieren: „den Bereich oder die Situation, die modelliert werden soll, in mathematische Begriffe, Strukturen und Relationen übersetzen“ (KMK 2004a, S. 12). Im Bereich der ersten Fremdsprache für den Mittleren Schulabschluss heißt es für den Kompetenzbereich Lesen u.a.: „Die Schülerinnen und Schüler können Korrespondenz lesen, die sich auf das eigene Interessengebiet bezieht und die wesent-

liche Aussage erfassen (B2); klar formulierte Anweisungen, unkomplizierte Anleitungen, Hinweise und Vorschriften verstehen (B1/B2) ...“ (KMK 2004b, S. 14). Schließlich heißt es in den Bildungsstandards für das Fach Geographie im Bereich Erkenntnisgewinn/Methoden (Deutsche Gesellschaft für Geographie 2007, S. 20): „Die Schülerinnen und Schüler können problem-, sach- und zielgemäß Informationen aus Karten Texten, Bildern, Statistiken, Diagrammen usw. auswählen.“

Die Standards sind hier als erfolgreich gezeigtes Verhalten definiert und man unterliegt einem Missverständnis, wenn man dieses Verhalten (die Performanz) mit Kompetenzen gleichsetzt. Lehnt man sich an einen stärker psychologisch geprägten Kompetenzbegriff an (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 hypothetischer Konstrukte. Solche Konstrukte (latente Merkmale) können erst mit Hilfe von Messinstrumenten der Beobachtung zugänglich gemacht werden (Operationalisierung) und die Messinstrumente erlauben es dann, Annahmen über die Eigenschaften der Kompetenzen (latenten Variablen) zu überprüfen.

Mit Bezug auf Weinert (2001) wird der Kompetenzbegriff von Klieme u.a. (2003) um motivationale, volitionale und soziale Aspekte erweitert. Dies wird teilweise dahingehend missverstanden, dass Aufgaben zur Überprüfung von domänenspezifischen Kompetenzen nicht allein kognitive, sondern gleichzeitig motivationale, volitionale und soziale Komponenten erfassen sollten. Die neueste PISA-Publikation (Deutsches PISA-Konsortium 2007) illustriert eindrucksvoll im Bereich der Naturwissenschaften, dass der Bezug auf die breite Fassung des Kompetenzbegriffs notwendigerweise darin mündet, dass neben kognitiven Leistungstests zusätzliche Fragebögen zu motivationalen Aspekten eingesetzt werden, z.B. solche zum domänenspezifischen Interesse. Nur die Trennung bei der Operationalisierung der Teilfacetten eines Kompetenzbereichs erlaubt die valide Messung jeder Facette.

Ein letzter Punkt, der bei der Konzeptionalisierung des Kompetenzbegriffes zentral ist, ist die Einbettung in ein so genanntes nomologisches Netzwerk, in dem andere Konstrukte in ihren Beziehungen zu den in den Standards aufgeführten Kompetenzen spezifiziert werden. Welche Rolle spielen beispielsweise die Variablen Vorwissen, schlussfolgerndes Denken, Genrewissen und Dekodierungsgeschwindigkeit beim Lese- und Hörverstehen? Was haben rezeptive Kompetenzen in der Verkehrs- und Fremdsprache gemeinsam? Worin unterscheiden sie sich? Worin besteht das Gemeinsame zwischen Bewerten und Beurteilen in der Biologie und argumentativem Schreiben im Fach Deutsch? Wer also fundierte Standards für sein Fach haben möchte, muss die domänenspezifischen Kompetenzen zu denen aus anderen Fächern (Domänen) konzeptionell abgrenzen und dies idealer Weise auch empirisch zeigen können.

Akzeptiert man, dass hinter dem Kompetenzkonzept im Wesentlichen hypothetische Konstrukte stehen, so stellt sich die Frage nach der theoretischen Fundierung der Konstrukte. Solche Fundierungen zu erarbeiten, ist eine genuine, meistens nur kooperativ zu bewältigende Aufgabe der Fachdidaktik und Erziehungswissenschaft/Psychologie, die

der Definition von Standards zeitlich vorgeordnet sein sollte. Erst wenn man elaboriert hat, was beispielsweise unter Lesekompetenz zu verstehen ist, wird man aus der Konstruktdefinition Verhaltensweisen ableiten können, die bei hohen oder geringen Ausprägungen auf dem Konstrukt beobachtbar sein sollte.

Ein positives Beispiel dafür, dass Kompetenzen vor ihrer Messung theoretisch solide fundiert wurden, ist ohne Frage die Large-scale Studie Deutsch-Englisch-Schülerleistungen-International (vgl. Beck/Klieme 2007; DESI-Konsortium 2006). Dort wurde beispielsweise die Lesekompetenz in drei Subdimensionen zerlegt (vgl. Nold/Willenberg 2007), und zwar (1) Wörter und Sätze möglichst schnell zu erkennen und zu verarbeiten (Dekodieren), (2) sprachliche Kompetenzen bei der Konstruktion von Bedeutung textueller Zusammenhänge einzusetzen und (3) diese mit passendem Weltwissen zu verknüpfen, um ein mentales Modell eines Textes im Detail und Gesamtzusammenhang zu entwickeln. Damit wurde an die internationalen Arbeiten in diesem Bereich angeknüpft und Items entsprechend dieser Konzeption generiert (vgl. Kintsch 1998; Alderson 2000).

Ein negatives Beispiel für den Fall, dass eine theoretische Fundierung ausbleibt, sind einige Formulierungen in den Bildungsstandards für das Fach Deutsch in der Grundschule (KMK 2005). Dort findet sich u.a. für die *Can do Statements* im Bereich Leseverstehen die Formulierung „bei der Beschäftigung mit literarischen Texten Sensibilität und Verständnis für Gedanken und Gefühle und zwischenmenschliche Beziehungen zeigen (...)“ (S. 12). Aus solchen Formulierungen ergeben sich sofort Fragen wie: Was ist mit Sensibilität und Verständnis für Gedanken und Gefühle gemeint? Wie können Schülerinnen und Schüler Sensibilität für Gefühle bei der Beschäftigung mit dem Text zeigen? Wodurch wird dies deutlich und wie lässt sich dies messen? Da das beschriebene Konstrukt selbst unklar bleibt, ist auch die Messung nicht möglich (vgl. hierzu beispielsweise Rost 2004). Das oben aufgeführte Statement beschreibt also eher eine didaktische Position – z.B. die Verwendung von produktiv-kreativen Aufgaben im Literaturunterricht wie das Verfassen von inneren Monologen als Gedankenwiedergabe von Figuren (Spinner 2003) – und ist primär als ein Appell an Lehrkräfte zu verstehen. Das Spannungsfeld zwischen theoretischer Unterfütterung von Standards und ihrer Messung soll nachfolgend am Beispiel Englisch illustriert werden.

### **3. Möglichkeiten und Grenzen bei der Messung von Kompetenzen und bei der Überprüfung von Bildungsstandards: Das Beispiel Englisch**

#### *3.1 Bildungsstandards und Europäischer Referenzrahmen*

Bildungsstandards können als Instrument der Qualitätssicherung nur dann wirksam werden, wenn sich auf der Basis der hinter ihnen liegenden Kompetenzen entsprechende Leistungstests entwickeln lassen. Im Fach Englisch werden in den Bildungsstandards der KMK für die Sekundarstufe I vier zentrale Kompetenzbereiche genannt, die sich an dem Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER; Europarat 2001)

orientieren. Der GER beschreibt ausführlich, was Lernende leisten müssen, „um eine Sprache für kommunikative Zwecke zu benutzen, und welche Kenntnisse und Fertigkeiten sie entwickeln müssen, um in der Lage zu sein, kommunikativ erfolgreich zu handeln“ (Europarat 2001, S. 14). Die dazu notwendigen Kompetenzen, wie sie in den Bildungsstandards formuliert sind, enthält die Tabelle 1: (1) kommunikative Fertigkeiten, (2) Verfügung über sprachliche Mittel, (3) interkulturelle und (4) methodische Kompetenzen.

Tab. 1: <b>Kompetenzbereiche in den Bildungsstandards für die erste Fremdsprache</b>	
<i>Funktionale Kommunikative Kompetenzen</i>	
Kommunikative Fertigkeiten: Leseverstehen Hör- und Hör-/Sehverstehen Sprechen An Gesprächen teilnehmen Zusammenhängendes Sprechen Schreiben Sprachmittlung	Verfügung über die sprachlichen Mittel: Wortschatz Grammatik Aussprache und Intonation Orthographie
<i>Interkulturelle Kompetenzen</i>	
Soziokulturelles Orientierungswissen Verständnisvoller Umgang mit kultureller Differenz Praktische Bewältigung interkultureller Begegnungssituationen	
<i>Methodische Kompetenzen</i>	
Textrezeption (Leseverstehen und Hörverstehen) Interaktion Textproduktion (Sprechen und Schreiben) Lernstrategien Präsentation und Mediennutzung Lernbewusstheit und Lernorganisation	

Kommunikative Fertigkeiten umfassen die fünf Bereiche Leseverstehen, Hör- und Hör-/Sehverstehen, Sprechen (in den Formen an Gesprächen teilnehmen und zusammenhängendes Sprechen), Schreiben sowie Sprachvermittlung. Kommunikative Fertigkeiten sind gebunden an die Verfügung über die sprachlichen Mittel Wortschatz, Grammatik, Aussprache und Intonation sowie Orthografie. Im Bereich der interkulturellen Kompetenzen werden vor allem Haltungen erfasst, die „ihren Ausdruck gleichermaßen im Denken, Fühlen und Handeln und ihre Verankerung in entsprechenden Lebenserfahrungen und ethischen Prinzipien haben“ (vgl. etwa KMK 2004b, S. 16). Schließlich greifen methodische Kompetenzen Teilaspekte allgemeiner sowie kommunikativer Kompetenzen (pragmatische Kompetenzen) auf, die sich vor allem Fragen der Selbstregulation beim Agieren in der Fremdsprache beziehen und sich theoretisch gut in der psychologischen Forschung zum selbstregulierten Lernen verankern lassen (vgl. z.B. Zimmerman/Schunk 2001).



Im GER werden drei Grundniveaus – A, B und C – unterschieden, die in je zwei Teilniveaus aufgliedert sind. Die Niveaus A1 und A2 stehen für elementare Sprachverwendung, die Niveaus B1 und B2 für selbstständige Sprachverwendung und die Niveaus C1 und C2 für eine kompetente Sprachverwendung.

Die KMK-Standards gehen davon aus, dass mit Erreichen des Hauptschulabschlusses im Mittel Leistungen auf GER-Niveau A2, mit Erreichen des Mittleren Schulabschlusses im Durchschnitt Leistungen auf GER-Niveau B1, teilweise auf B2, erbracht werden können. Konkret liest sich dies beim Mittleren Abschluss für den Bereich Lesen (vgl. KMK 2004b, S. 12):

„Die Schülerinnen und Schüler können

- weitgehend selbstständig verschiedene Texte aus Themenfeldern ihres Interessen- und Erfahrungsbereiches lesen und verstehen (B1+),
- 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),
- längere Texte nach gewünschten Informationen durchsuchen und Informationen aus verschiedenen Texten zusammentragen, um eine bestimmte Aufgabe zu lösen (B1+),
- in kürzeren literarischen Texten (z.B. Short Stories) die wesentlichen Aussagen erfassen und diese zusammentragen, um eine bestimmte Aufgabe zu lösen (B1),
- die Aussagen einfacher literarischer Texte verstehen,
- in klar geschriebenen argumentativen Texten zu vertrauten Themen die wesentlichen Schlussfolgerungen erkennen, z.B. in Zeitungsartikeln (B1/ B1+).“

Erkennbar ist dabei, dass sich die Leistungserwartungen im Wesentlichen auf die Teilkompetenzen (1) Identifikation von Detailinformationen, (2) Verknüpfen von Informationen zum Aufbau eines Situationsmodells (Kintsch 1998), sowie (3) Schlussfolgerungen beziehen und dabei auf ein breites Genre von Sach- und literarischen Texten anwendbar sein sollten, ohne dass die Inhalte weiter spezifiziert werden (Orientierung an Kompetenzen anstelle von Inhalten). Durch die klare Konstruktfassung, die internationale Anbindung (z.B. Alderson 2000) und die einfachen *Can do Statements* wird es möglich, eine Vielzahl von Aufgaben zur Überprüfung der Standards zu generieren. Dasselbe gilt im Übrigen für das Hörverstehen und Schreiben in der Fremdsprache Englisch (zu Details vgl. Rupp/Vock/Harsch/Köller, im Druck).

### 3.2 *Probleme der Operationalisierung am Beispiel der interkulturellen Kompetenz*

In den Bildungsstandards selbst (vgl. KMK 2004b), aber auch in der nationalen und internationalen Literatur wird die interkulturelle Kompetenz mehrdimensional als Zusammenspiel von kognitiven, affektiven und behavioralen Merkmalen verstanden

(vgl. z.B. Hesse/Göbel 2007). Konkret heißt es hierzu in den Standards (KMK 2004b, S. 16):

„Die Schülerinnen und Schüler

- kennen elementare spezifische Kommunikations- und Interaktionsregeln ausgewählter englisch- bzw. französischsprachiger Länder und verfügen über ein entsprechendes Sprachregister, das sie in vertrauten Situationen anwenden können,
- sind neugierig auf Fremdes, aufgeschlossen für andere Kulturen und akzeptieren kulturelle Vielfalt,
- sind bereit, sich auf fremde Situationen einzustellen und sich in Situationen des Alltagslebens angemessen zu verhalten,
- sind in der Lage, ungewohnte Erfahrungen auszuhalten, mit ihnen sinnvoll und angemessen umzugehen und das Fremde nicht als etwas wahrzunehmen, das Angst macht,
- können sich in Bezug auf die Befindlichkeiten und Denkweisen in den fremdkulturellen Partner hineinversetzen,
- kennen gängige Sicht- und Wahrnehmungsweisen, Vorurteile und Stereotype des eigenen und des fremdkulturellen Landes und setzen sich mit ihnen auseinander,
- können kulturelle Differenzen, Missverständnisse und Konfliktsituationen bewusst wahrnehmen, sich darüber verständigen und gegebenenfalls gemeinsam handeln.“

Man erkennt hier die Vermischung von deklarativem Wissen (Spiegelstriche 1 und 6), Ungewissheitsorientierung und Ethnozentrismus (Spiegelstriche 2, 3 und 4), Empathie (Spiegelstrich 5) und sozialer Kompetenz (letzter Spiegelstrich). Zu einem ähnlichen Resultat bei der Vielfalt der Teilkompetenzen kommen auch Hesse und Göbel (2007) auf Grund einer Curriculumsanalyse in den 16 Ländern der Bundesrepublik Deutschland.

Während unbestritten ist, dass sich die aufgezählten Kompetenzen grundsätzlich messen lassen, vielfach existieren hierzu auch Instrumente (vgl. Hesse/Göbel 2007), ergeben sich durch die Mehrdimensionalität viele Probleme, u.a. die Frage nach typischen Profilen. Besonders prekär ist die Frage nach der Definition von operationalen Standards und der Festsetzung, bei welchem Empathieausmaß beispielsweise ein Schüler die Standards erreicht hat. Deutlich wird hier, dass sich nicht alle in den Bildungsstandards definierten Ziele ohne größeren Forschungsaufwand operationalisieren lassen. Man könnte argumentieren, Bildungsziele entzögen sich hier der einfachen Messung und Überprüfung.

### *3.3 Grenzen bei der großflächigen Normierung von Aufgaben und bei der Überprüfung der Erreichung der Standards*

Bildungsstandards erfüllen ihre Steuerungsfunktion im Bereich der Qualitätssicherung, wenn die zu ihrer Überprüfung notwendigen Instrumente großflächig und ökonomisch

eingesetzt werden können. Am einfachsten umsetzbar ist dabei die Messung mit Hilfe von Paper-Pencil-Instrumenten im intakten Klassenverband und mit wenig anschließendem Kodieraufwand. Dies gelingt sehr gut im Bereich der Mathematik oder bei den rezeptiven Sprachkompetenzen, findet aber Grenzen bei der Feststellung sprachproduktiver Kompetenzen, hier in erster Linie beim Sprechen. Dies gilt insbesondere für dialogische Sprechsituationen, die sich kaum im Klassenverband realisieren lassen und oft großen Kodieraufwand erfordern. Letzterer ergibt sich daraus, dass Beobachtungs- bzw. Beurteilungsrichtlinien für Beobachter erarbeitet und Vorkehrungen getroffen werden müssen, die eine objektive Bewertung der Sprechkompetenz erlauben. Im *Large-scale Assessment* können solche Kompetenzbereiche nur ansatzweise erfasst werden, beispielsweise in einer computerbasierten Testung, in der über ein *Headset* ein Frage-Antwort-Test realisiert wird.

Zusammenfassend sollte deutlich geworden sein, dass die großflächige Erfassung von gewissen Teilkompetenzen nur schwer möglich ist und sich Teile der Bildungsstandards der Messung entziehen. Die Annahme, Standards ließen sich generell in Messinstrumente und *Assessments* überführen, ist danach nicht haltbar. Es wird umso schwerer, sie in Testaufgaben umzusetzen, je weniger die hinter ihnen liegenden Kompetenzen theoretisch fundiert sind.

### 3.4 *Operationale Festlegung von Standards*

In Papier gegossene, mehr oder weniger präzisierete Leistungserwartungen, wie sie in Bildungsstandards festgehalten sind, lassen offen, welche konkreten Leistungen bei der Bewältigung von Aufgaben gezeigt werden müssen, damit die Standards erfüllt werden. Als Folge sind die häufig geführten Diskussionen über Mindest-, Regel- und Idealstandards obsolet, so lange keine empirischen Befunde zu tatsächlichen Leistungen von Schülerinnen und Schülern in standardisierten Tests vorliegen. Erst die Bereitstellung einer Leistungsskala (Metrik) und die Definition von Kompetenzstufen, wie sie beispielsweise in PISA stattfinden (Deutsches PISA-Konsortium 2007), erlauben die finale Festlegung von *Cut Scores*, bei denen Standards – egal ob Mindest-, Regel- oder Idealstandards – erreicht sind. In diesem Zusammenhang ist es wichtig festzuhalten, dass die definierten Niveaustufen im Gegensatz zu den Kompetenzmodellen, in denen die Dimensionen präzisiert werden, in der Regel nicht wirklich theoriebasiert sind. Sie stützen sich vielmehr auf *post hoc* Analysen der gewonnenen Daten und sind an einen Standard-Setting-Prozess gekoppelt (vgl. Hambleton/Jaeger/Plake/Mills 2000), bei dem im Zusammenspiel psychometrischer Analysen, fachdidaktischer Erwägungen und politischer Verträglichkeit gut interpretierbare Bereiche auf dem Kompetenzkontinuum definiert werden (vgl. auch Hartig 2007). Es handelt sich dabei um konsensuelle Festlegungen, welche die Interpretation von Testwerten erleichtern, die aber weit davon entfernt sind, grundlagenwissenschaftlich fundiert zu sein. Sie gar in entwicklungspsychologischen Modellen verankern zu wollen, stellt einen Rückfall in überholte theoretische Ansätze dar. Wie es Adams und Wu (2002, S. 197) ausdrücken: „Dividing (...) these continua in-

to levels, though useful for communication about students' development, is essentially arbitrary.“ (Adams/Wu 2002, S. 197).

Konsensuell gewonnene Niveaustufen auf einem empiriebasierten Leistungskontinuum können dann in der Tat verwendet werden um Bereich zu definieren, in denen Mindest-, Regel- und Idealstandards erreicht werden.

#### **4. Schlussfolgerungen: Was heißt dies für die Erarbeitung und Messung von Bildungsstandards?**

Bildungsstandards können ohne Frage das Fundament für qualitätssichernde und qualitätsentwickelnde Maßnahmen bilden. Im vorliegenden Beitrag wurde versucht zu zeigen, dass sie ihre Funktion im Bereich der Qualitätssicherung aber nur unter gewissen Rahmenbedingungen erfüllen können. Dies soll noch einmal in drei Schlussfolgerungen zusammengefasst werden.

- *Schlussfolgerung 1:* Ohne fundierte Kompetenzmodelle keine guten Standards

Jedes Fach ist gut beraten, Bildungsstandards fachdidaktisch und pädagogisch-psychologisch zu verorten. Erst dann, wenn das den Leistungen hinterliegende Kompetenzkonstrukt mit all seinen Facetten und Antezedenzen präzise beschrieben ist, wird es möglich sein, ein tragfähiges Messmodell zu spezifizieren, auf dessen Basis die Entwicklung von Items zur Messung der Kompetenzen gelingen kann. Zu einem fundierten Kompetenzmodell gehört notwendigerweise ein nomologisches Netzwerk, das die Beziehung der jeweiligen Kompetenz zu anderen Konstrukten spezifiziert. Nomologische Netzwerke helfen zu klären, ob die im jeweiligen Fachkontext entwickelten Standards überhaupt domänenspezifisch sind.

- *Schlussfolgerung 2:* Ohne fundierte, aber gleichzeitig sparsame Kompetenzmodelle keine guten Messinstrumente

Am Beispiel der interkulturellen Kompetenz in den Fremdsprachen wurde deutlich gemacht, wie schwierig es ist, Instrumente zu entwickeln, wenn Kompetenzbereiche sehr facettenreich sind und neben kognitiven auch motivationale, emotionale und soziale Aspekte enthalten. Einmal abgesehen davon, dass die Vielfältigkeit der Facetten die Entwicklung von Messinstrumenten überhaupt erschwert, führt sie auch dazu, dass die operationale Definition, wann Standards – egal ob Mindest-, Regel- oder Idealstandards – erreicht sind, unmöglich wird.

Am Beispiel der jüngsten PISA-Publikation (Deutsches PISA-Konsortium 2007) wurde weiterhin illustriert, dass eine breite Fassung des Kompetenzkonstrukts unter Berücksichtigung motivationaler, volitionaler und selbstregulativer Aspekte nur in der Entwicklung getrennter Instrumente für die verschiedenen Aspekte münden kann.

- *Schlussfolgerung 3*: Ohne konsensuell gewonnene Niveaustufen keine substanzielle Diskussion über Mindest-, Regel- und Idealstandards

Die Festlegung der Niveaustufen auf der kontinuierlichen Kompetenzskala erfolgt zu- meist mit Hilfe einer empiriebasierten und auf Plausibilität orientierten Konsensfindung zwischen Vertretern verschiedener Fachdisziplinen in Abstimmung mit Vertretern der Bildungspolitik. Dass Forschungsprojekte, die nach einer a priori erfolgenden, entwicklungspsychologischen und fachdidaktischen Verankerung der Kompetenzstufen streben, erfolgreich sein und in absehbarer Zeit empirisch belastbare Ergebnisse hervorbringen können, scheint derzeit zweifelhaft. Die in den Standards beschriebenen Kompetenzen sind durchgängige kontinuierliche Personenmerkmale und Veränderungen finden auf einem Kontinuum statt, auf dem sich die Schülerinnen und Schüler langsam „hocharbeiten“. Die Annahme, es gäbe von Zeit zu Zeit qualitative Sprünge von einer zur nächsten Kompetenzstufe und dies sei entwicklungspsychologisch oder fachdidaktisch plausibel begründbar, steht dabei zumindest teilweise im Gegensatz zu den neueren Erkenntnissen der Lernpsychologie, in der Stufenmodelle keine Rolle spielen.

Die konsensuell gewonnenen und empirisch fundierten Stufen auf dem Kontinuum können aber den Ausgangspunkt für eine datenbasierte Definition von Mindest-, Regel- und Idealstandards sein, die politisch und pädagogisch verträglich ist. Ohne belastbare Daten, die zeigen, wo auf dem Leistungskontinuum Schülerinnen und Schüler tatsächlich liegen, bleiben Diskussionen über Mindest-, Regel- und Idealstandards substanzlos.

## Literatur

- Adams, R./Wu, M. (2002): PISA 2000 technical report. Paris: OECD.
- Alderson, J.C. (2000): *Assessing Reading*. Cambridge: Cambridge University Press.
- Beck, B./Klieme, E. (Hrsg.) (2007): *Sprachliche Kompetenzen – Konzepte und Messung – DESI-Studie*. Weinheim: Beltz.
- Benner, D./Krause, S./Nikolova, R./Pilger, T./Schluß, H./Schieder, R./Weiß, T./Willems, J. (2007): Ein Modell domänenspezifischer religiöser Kompetenz. Erste Ergebnisse aus dem DFG-Projekt RU-Bi-Qua. In: Benner, D. (Hrsg.): *Bildungsstandards: Chancen und Grenzen, Beispiele und Perspektiven*. Paderborn: Schöningh, S. 141–156.
- DESI-Konsortium (2006): *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch Englisch Schülerleistungen International*. Frankfurt: Deutsches Institut für Internationale Pädagogische Forschung (dipf).
- Deutsche Gesellschaft für Geographie (2007): *Bildungsstandards im Fach Geographie für den Mittleren Schulabschluss mit Aufgabenbeispielen*. Berlin: DGfG.
- Deutsches PISA-Konsortium (Hrsg.) (2001): *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Deutsches PISA-Konsortium (Hrsg.) (2007): *PISA '06. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Europarat. (2001): *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin: Langenscheidt.
- Hambleton, R.K./Jaeger, R.M./Plake, B.S./Mills, C. (2000): *Setting performance standards on complex educational assessments*. In: *Applied Psychological Measurement* 24, S. 355–366.

- Hartig, J. (2007): Skalierung und Definition Kompetenzniveaus. In: Beck, B./Klieme, E. (Hrsg.), S. 83–99.
- Hesse, H.G./Göbel, K. (2007): Interkulturelle Kompetenz. In: Beck, B./Klieme, E. (Hrsg.), S. 253–269.
- 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.
- Kintsch, W. (1998): Comprehension. Cambridge: Cambridge University Press.
- KMK (2004a): Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. München: Luchterhand.
- KMK (2004b): Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss. München: Luchterhand.
- KMK (2005): Bildungsstandards im Fach Deutsch für den Primarbereich (Jahrgangsstufe 4). München: Luchterhand.
- Nold, G./Willenberg, H. (2007): Lesefähigkeit. In: Beck, B./Klieme, E. (Hrsg.), S. 23–41
- Rost, J. (2004). Lehrbuch Testtheorie – Testkonstruktion. Bern: Huber.
- Rupp, A.A./Vock, M./Harsch, C./Köller, O. (im Druck): The development, calibration, and validation of standards-based tests for English as a first foreign language. Münster: Waxmann.
- Spinner, K.-H. (2003): Handlungs- und produktionsorientierte Verfahren im Literaturunterricht. In: Kämper-van den Boogart, M. (Hrsg.): Deutschdidaktik. Leitfaden für die Sekundarstufe I und II. Berlin: Cornelsen-Scriptor.
- Weinert, F.E. (2001): Vergleichende Leistungsmessung in Schulen – Eine umstrittene Selbstverständlichkeit. In: Weinert, F.E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim: Beltz, S. 17–31.
- Zimmerman, B.J./Schunk, D.H. (Hrsg.) (2001): Self-regulated learning and academic achievement. Mahwah, NJ: Erlbaum.

*Anschrift des Autors:*

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