



Ehmke, Timo: Leiß, Dominik: Blum, Werner: Prenzel, Manfred

Entwicklung von Testverfahren für die Bildungsstandards Mathematik.

Rahmenkonzeption, Aufgabenentwicklung, Feld- und Haupttest

Unterrichtswissenschaft 34 (2006) 3. S. 220-238



Quellenangabe/ Reference:

Ehmke, Timo; Leiß, Dominik; Blum, Werner; Prenzel, Manfred: Entwicklung von Testverfahren für die Bildungsstandards Mathematik. Rahmenkonzeption, Aufgabenentwicklung, Feld- und Haupttest - In: Unterrichtswissenschaft 34 (2006) 3, S. 220-238 - URN: urn:nbn:de:0111-opus-55160 - DOI: 10.25656/01:5516

https://nbn-resolving.org/urn:nbn:de:0111-opus-55160 https://doi.org/10.25656/01:5516

in Kooperation mit / in cooperation with:



http://www.juventa.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 ausschießlich für den persönlichen, nich-kommerzeilein Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Einschientsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Köplien dieses Dokuments müssen alle Uhreberrechtshirwisels und sonstigen Hinweise auf gesetzlichten Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in grendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen

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 Inis document solely intended for four personain, 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.

Digitalisiert Kontakt / Contact:

pedocs

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung E-Mail: pedocs@dipf.de

Internet: www.pedocs.de



An (05) Wat

# Unterrichtswissenschaft

Zeitschrift für Lernforschung 34. Jahrgang / 2006 / Heft 3

Thema:

Kompetenzmodelle	
Verantwortliche Herausgeber: Olaf Köller, Manfred Prenzel	
Erratum	194
André A. Rupp, Michael Leucht, Ralph Hartung ,Die Kompetenzbrille aufsetzen' – Verfahren zur multiplen Klassifikation von Lernenden für Kompetenzdiagnostik in Unterricht und Testung	195
Timo Ehmke, Dominik Leiß, Werner Blum, Manfred Prenzel Entwicklung von Testverfahren für die Bildungsstandards Mathematik. Rahmenkonzeption, Aufgabenentwicklung, Feld- und Haupttest	220
Olaf Köller, Ulrich Trautwein, Kai S. Cortina, Jürgen Baumert: Rezeptive Kompetenzen in Englisch am Ende der gymnasialen Oberstufe. Verankerung deutscher Abiturienten am Gemeinsamen Europäischen Referenzrahmen für die Sprachen	239
Allgemeiner Teil	
Jörg Zumbach, Johann Pixner  Textverständlichkeit und kognitive Belastung beim Lernen mit Text und Hypertext	25 <i>6</i>
Andreas Hartinger Interesse durch Öffnung des Unterrichts – wodurch?	272

Timo Ehmke, Dominik Leiß, Werner Blum, Manfred Prenzel

# Entwicklung von Testverfahren für die Bildungsstandards Mathematik

Rahmenkonzeption, Aufgabenentwicklung, Feld- und Haupttest

Um die schulische Bildung in Deutschland zu verbessern, hat die Ständige Konferenz der Kultusminister der Länder der Bundesrepublik (KMK) beschlossen, Bildungsstandards in Deutschland einzuführen. Ziel dieses Beitrags ist es, die wichtigsten Schritte in dem Verlauf zu beschreiben, wie man - ausgehend von einer theoretischen Rahmenkonzeption für Bildungsstandards - hin zu empirisch messbaren Kompetenzen kommt. Die Rahmenkonzention der Bildungsstandards in Mathematik unterscheidet dabei verschiedene Dimensionen: mathematische Leitideen, mathematische Kompetenzen und gestufte Anforderungsbereiche. Sie dient als theoretischer Konstruktionsrahmen für die Entwicklung von Testaufgaben für die Bildungsstandards. Konzipiert, entwickelt und pilotiert wurden die Aufgaben in vier Regionalgruppen unter der Aufsicht einer Steuerungs- und einer Bewertungsgruppe. Die Arbeit in diesen Fachgruppen wurde erfolgreich abgeschlossen. Dieser Prozess kann daher auch als paradigmatische Vorlage für die Umsetzung von Bildungsstandards in anderen Fächern (etwa Naturwissenschaften, Deutsch oder Fremdsprachen) dienen. Der Beitrag beschreibt ferner das Testdesign des Feldtests für die Bildungsstandards sowie die empirischen Kennwerte, nach denen die Aufgaben für den Haupttest ausgewählt werden. Der Beitrag schließt mit einer kurzen Bewertung des Aufgabenentwicklungsprozesses und gibt einen Ausblick auf ausstehende Arbeitsschritte zur Implementation der Bildungsstandards.

The introduction of education standards in Germany was a resolution of the conference of ministers of education of the federal states (Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik, KMK) in order to improve school education in Germany. The aim of this paper is to describe the most important steps in the process of how - based on a theoretical framework for education standards - to gain competencies that can be measured empirically. The framework of education standards in mathematics distinguishes between different dimensions: overarching ideas, mathematical competencies and task areas at different levels. It served as a theoretical framework for the development of tasks for education standards. The

tasks for education standards were conceived, developed and pretested in four regional groups under the supervision of an evaluation team.

The work in these groups was successfully completed. Thus, this process can also serve as a paradigmatic template for the implementation of education standards in other subjects (science, German, or foreign languages, for instance). This paper also describes the test design of the field trial for education standards, as well as the empirical values according to which the tasks for the main study were selected. The paper ends with a short evaluation of the task development process and suggests the future work that is necessary for the implementation of education standards.

#### 1. Einleitung

Mit der Veröffentlichung der Ergebnisse der TIMS-Studie (Baumert & Lehmann, 1997) begann in Fachkreisen eine Diskussion über die Notwendigkeit einer Bildungsreform in Deutschland. Nach der Bekanntgabe der Ergebnisse von PISA 2000 wurde diese Debatte auch von der breiten Öffentlichkeit aufgegriffen. Wie kann die Qualität der schulischen Bildung in Deutschland gesichert werden? Mit Blick auf die Bildungssysteme in Staaten, deren Jugendliche bei TIMSS und PISA hohe durchschnittliche Kompetenzen erreichen, wurde über regelmäßige vergleichende Lernstandserhebungen und zentrale Abschlussprüfungen diskutiert. Eines der Handlungsfelder von der Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik (KMK) betrifft die Einführung von Bildungsstandards.

Das Anliegen dieses Beitrags ist es, die wichtigsten Schritte in dem Prozess zu beschreiben, wie aus den Bildungsstandards für Mathematik Messinstrumente entwickelt und erprobt werden, um letztlich das Erreichen von Bildungsstandards überprüfen zu können. Dabei wird von einer vorliegenden Rahmenkonzeption ausgegangen, die inhaltlich zu erreichende Regelstandards für den Mittleren Abschluss in Mathematik definiert. Dieser Ablauf ist in mehrere Stufen unterteilt und wurde zunächst paradigmatisch vom PISA-Konsortium 2006 für die Mathematik durchgeführt.

In diesem Beitrag werden rückblickend die wichtigsten Beschlüsse der KMK zur Einführung der Bildungsstandards zusammengefasst (Abschnitt 2) und die inhaltliche Rahmenkonzeption für die Bildungsstandards in Mathematik skizziert (Abschnitt 3). Anschließend wird der eigentliche Prozess der Aufgabenentwicklung dargelegt (Abschnitt 4), der Voraussetzung ist für die Umsetzung in ein Feldtestinstrument (Abschnitt 5). Kriterien zur Auswahl von Aufgaben für den Haupttest und ein Ausblick auf mögliche methodische Auswertungsstrategien werden in Abschnitt 6 und 7 beschrieben. Der Beitrag schließt mit einer kurzen Bewertung des Prozesses und einem Ausblick auf ausstehende Arbeitsschritte (Abschnitt 8).

## 2. KMK-Beschlüsse zu Bildungsstandards in Deutschland

Als eine Reaktion auf die bildungspolitischen Befunde von PISA 2000 hat die Ständige Konferenz der Kultusminister der Länder der Bundesrepublik am 24.05.2002 beschlossen, nationale Bildungsstandards in den Fächern Deutsch und Mathematik sowie für die erste Fremdsprache (Englisch, Französisch) für den Mittleren Abschluss und für den Hauptschulabschluss zu erarbeiten. Sie sollen in landesweiten Vergleichs- und Orientierungsarbeiten sowie in bundesweiten Vergleichsstudien eingesetzt werden.

Die inhaltliche Diskussion um die Einführung der Bildungsstandards wurde maßgeblich von der BMBF-Expertise "Zur Einführung nationaler Bildungsstandards" (Klieme et al., 2003) vorangebracht. Sie beschreibt die Funktionen, konzeptuellen Grundlagen und die Entwicklung von Bildungsstandards. Mit der Einführung von Bildungsstandards soll ein weiterer Schritt beim Wechsel von der Input- hin zur Output-Orientierung vollzogen werden. Qualitätssicherung im Bildungssystem wird dann nicht mehr allein durch Bildungsprogramme (Lehrpläne) gesteuert, sondern auch über zu erzielende Lernergebnisse (Bildungsstandards). Zusätzlich liefern die regelmäßig durchgeführten Schulleistungsstudien wie PISA, IGLU oder TIMSS Rückmeldung über Kompetenzniveaus im internationalen Vergleich.

Am 4.12.2003 hat die KMK die Einführung von Bildungsstandards im Fach Mathematik beschlossen (Kultusministerkonferenz, 2003). Für den Mittleren Schulabschluss und für den Hauptschulabschluss werden damit Bildungsstandards im Sinne einer Output-Kontrolle definiert und verbindliche Ziele und Kriterien festgelegt, anhand derer das Erreichen eines Standards erfasst und gemessen werden kann (Kultusministerkonferenz, 2004). Die Bildungsstandards bieten allen Lehrkräften, Eltern, Schülerinnen und Schülern, der Bildungsadministration und der Öffentlichkeit eine Orientierung über verbindliche Anforderungen. Sie stellen außerdem eine Basis für Evaluationen auf mehreren Ebenen dar: Bildungssystem, Schule, Klasse, Die wichtigste Intention der Bildungsstandards zielt aber auf die Verbesserung der unterrichtlichen Qualität, um die fachliche Bildung von Kindern und Jugendlichen zu steigern (Blum, Drüke-Noe, Leiß, Wiegand & Jordan, 2005; Prenzel, 2005). Ansätze und Beispiele für Maßnahmen zur Qualitätsverbesserung im mathematisch-naturwissenschaftlichen Unterricht zeigen etwa Blum & Leiß (2005), Biermann & Blum (2001) oder Prenzel & Ostermeier (2003). Bildungsstandards sind somit Bestandteil eines umfassenden Systems der Qualitätssicherung, das Fortschritte in der Schul- und Unterrichtsentwicklung einschließen und Parameter für die interne und externe Evaluation bereitstellen soll (Klieme & Steinert, 2004).

Die KMK-Bildungsstandards stellen abschlussbezogene Regelstandards im Unterschied zu Mindeststandards dar. Sie

- "greifen die Grundprinzipien des jeweiligen Unterrichtsfachs auf,

- beschreiben die fachbezogenen Kompetenzen und zugrunde liegenden Wissensbestände, die Schülerinnen und Schüler bis zu einem bestimmten Zeitpunkt ihres Bildungsgangs erreicht haben sollen,
- zielen auf systematisches und vernetztes Lernen und folgen so dem Prinzip des kumulativen Kompetenzerwerbs,
- beschreiben die erwarteten Leistungen im Rahmen von Anforderungsbereichen,
- beziehen sich auf den Kernbereich eines Faches und geben den Schulen Gestaltungsräume für ihre pädagogische Arbeit,
- weisen ein mittleres Anforderungsniveau aus,
- werden durch Aufgabenbeispiele veranschaulicht" (Kultusministerkonferenz, 2003).

Die Kultusministerkonferenz hat sich damit für ein pragmatisches Vorgehen entschieden, weil notwendige Mindeststandards erst nach einem längeren Prozess der Erfahrung im Umgang mit Bildungsstandards formuliert werden können (Kultusministerkonferenz, 2004, S. 14). Mindeststandards setzen voraus, dass die empirischen Schwierigkeiten von Aufgaben bereits bekannt sind, dass Niveaustufen präzisiert und insgesamt die Standards und Aufgabenbeispiele validiert wurden. Bildungsstandards in einem ersten Schritt sofort als Mindeststandards zu definieren birgt die Gefahr, einerseits Schülerinnen und Schüler massiv zu unterfordern, aber andererseits auch größere Teile der Schülerschaft durch überzogene Bildungsstandards zu überfordern. Aus diesem Grund hat sich die Kultusministerkonferenz darauf verständigt, zunächst Regelstandards zu definieren. Damit ist zunächst ein mittleres Anforderungsniveau definiert worden, das auf der Basis von Einschätzungen von Experten aus Schule und Unterricht beruht.

# 3. Rahmenkonzeption für Bildungsstandards in Mathematik

Die Konzeption der Bildungsstandards im Fach Mathematik orientiert sich an den Vorarbeiten der Standards, die vom National Counsil of Teachers of Mathematics (NCTM, 2000) entwickelt wurden. Der US-Amerikanische Verband der Mathematiklehrkräfte begann bereits vor mehr als fünfzehn Jahren mit der Entwicklung von entsprechenden Standards für Unterrichtsprozesse. Eine weitere fachspezifische Vorlage stellt das Rahmenkonzept für die Mathematik dar (OECD, 2003). Beide Rahmenkonzeptionen beschreiben die Mathematik-Kompetenz aus einer inhaltlichen und einer prozessorientierten Sicht. Erstere umfasst verschiedene inhaltliche Themenbereiche des Mathematikunterrichts, die zweite bezieht sich auf mathematikbezogene Denkprozesse und Tätigkeiten. Die KMK-Bildungsstandards greifen diese Differenzierung auf und beschreiben die Bereiche als "mathematische Leitideen" und "mathematische Kompetenzen" (OECD, 2003). Die Bildungsstandards grenzen sich jedoch von den NCTM-Standards

ab, weil sie keine Unterrichtsstandards definieren, sondern auf Kompetenzen ausgerichtet sind.

Tab. 1: Beschreibung der fünf mathematischen Leitideen

Mathematische Leitidee	Beschreibung								
Zahl	<ul> <li>Alle Arten von Quantifizierungen, in denen Zahlen verwendet werden, um Situationen zu beschreiben oder zu organisieren.</li> </ul>								
Messen	<ul> <li>Alle Arten von Größen in verschiedenen Maßeinheiten zu erfassen und mathematisch zu verarbeiten.</li> </ul>								
Raum und Form	<ul> <li>Alle Arten von ebenen oder r\u00e4umlichen Konfigurationen, Gestalten und Mustern.</li> </ul>								
Funktionaler	<ul> <li>Alle Arten von relationalen und funktionalen</li> </ul>								
Zusammenhang	Beziehungen zwischen mathematischen Objekten Alle Arten von Phänomenen und Situationen, die								
Daten und Zufall	statistische Daten beinhalten oder bei denen der Zu- fall eine Rolle spielt								

Insgesamt werden fünf mathematische Leitideen unterschieden (Tabelle 1). Im Vergleich zu den vier "Overarching Ideas" in PISA 2003 (Blum et al., 2004, S. 49f) überschneiden sich "Raum und Form" mit "Space & Shape", "Funktionaler Zusammenhang" mit "Change & Relationship", "Daten und Zufall" mit "Uncertainty". Der vierte Inhaltsbereich in PISA "Quantity" wird bei den Bildungsstandards stärker durch die beiden Bereiche "Zahl" und "Messen" differenziert.

Das Auseinandersetzen mit konkreten Inhalten, den mathematischen Leitideen, erfordert verschiedene mathematische Kompetenzen. Im Rahmen der KMK-Bildungsstandards werden sechs mathematische Kompetenzen als wichtig erachtet (Tabelle 2), die für alle Ebenen des mathematischen Arbeitens relevant sind. Diese Kompetenzen werden immer anhand eines konkreten Inhaltsbereiches erworben und angewendet.

Mathematikaufgaben lassen sich nach den mathematischen Leitideen und den für ihre Lösung erforderlichen mathematischen Kompetenzen klassifizieren. Zur Lösung von Mathematikaufgaben kommen die oben genannten sechs mathematischen Kompetenzen in unterschiedlicher Ausprägung zum Einsatz. Anspruch und Komplexität werden durch drei Anforderungsbereiche eingestuft (Tabelle 3). Beispiele, die die unterschiedlichen Zuordnungen von Mathematikaufgaben zu mathematischen Leitideen, Kompetenzen und Anforderungsniveaus illustrieren, finden sich in KMK (2004) und in Blum, Drüke-Noe, Hartung & Köller (in Druck).

Durch die Definition von mathematischen Inhaltsbereichen (Leitideen), von relevanten Tätigkeiten (Kompetenzen) sowie durch die Unterscheidung von Anforderungsbereichen wird eine mehrdimensionale Rahmenstruktur festgelegt. Diese Rahmenstruktur kann durch Testaufgaben ausgefüllt werden,

Tab. 2: Beschreibung der sechs mathematischen Kompetenzen

Mathematische Kompetenz	Beschreibung
Mathematisch argumentieren	<ul> <li>Fragen stellen, die für die Mathematik charakteristisch sind und Vermutungen begründet äußern.</li> <li>Mathematische Argumentationen entwickeln (wie Erläuterungen, Begründungen, Beweise).</li> <li>Einen Lösungsweg beschreiben und begründen.</li> </ul>
Probleme mathematisch lösen	<ul> <li>Vorgegebene und selbst formulierte Probleme bearbeiten.</li> <li>Geeignete heuristische Hilfsmittel, Strategien und Prinzipien zum Problemlösen auswählen und anwenden.</li> <li>Die Plausibilität der Ergebnisse überprüfen, Lösungsideen finden und Lösungswege reflektieren.</li> </ul>
Mathematisch modellieren	<ul> <li>Den Bereich oder die Situation, die modelliert werden soll, in mathematische Begriffe, Strukturen und Relationen übersetzen.</li> <li>In dem jeweiligen mathematischen Modell arbeiten.</li> <li>Ergebnisse in dem entsprechenden Bereich oder der entsprechenden Situation interpretieren und prüfen.</li> </ul>
Mathematische Darstellungen verwenden	<ul> <li>Verschiedene Formen der Darstellung von mathematischen Objekten und Situationen anwenden, interpretieren und unterscheiden.</li> <li>Beziehungen zwischen Darstellungsformen erkennen.</li> <li>Unterschiedliche Darstellungsformen je nach Situation und Zweck auswählen und zwischen ihnen wechseln.</li> </ul>
Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen	<ul> <li>Mit Variablen, Termen, Gleichungen, Funktionen, Diagrammen, Tabellen arbeiten.</li> <li>Symbolische und formale Sprache in natürliche Sprache übersetzen und umgekehrt.</li> <li>Lösungs- und Kontrollverfahren ausführen.</li> <li>Mathematische Werkzeuge (wie Formelsammlung, Taschenrechner, Software) sinnvoll und verständig einsetzen.</li> </ul>
Mathematisch kommunizieren	<ul> <li>Überlegungen, Lösungswege bzw. Ergebnisse dokumentieren, verständlich darstellen und präsentieren, auch unter Nutzung geeigneter Medien.</li> <li>Die Fachsprache adressatengerecht verwenden.</li> <li>Äußerungen von anderen und Texte zu mathematischen Inhalten verstehen und überprüfen.</li> </ul>

die verschiedene Facetten dieser Struktur erfassen. Auf diese Weise lassen sich Testinstrumente entwickeln, die die Erfüllung von Standards auch empirisch messbar machen. Voraussetzung hierfür ist wiederum ein umfassender Aufgabenpool. Auf den Ablauf der Aufgabenentwicklung gehen wir im folgenden Abschnitt näher ein.

Tab. 3: Beschreibung von mathematischen Anforderungsbereichen

Anforderungsbereich	Beschreibung							
1. "Reproduzieren"	Grundlegende Begriffe, Sätze und Verfahren in einem abgegrenzten Gebiet und einem wiederholenden Zusammenhang wiedergeben oder direkt anwenden.							
2. "Zusammenhänge herstellen"	Bekannte Sachverhalte bearbeiten, indem Kenntnisse, Fertigkeiten und Fähigkeiten verknüpft werden, die in der Auseinandersetzung mit Mathematik auf verschiedenen Gebieten erworben wurden.							
3 "Verallgemeinern und Reflektieren"	Komplexe Gegebenheiten mit dem Ziel bearbeiten, zu eigenen Problemformulierungen, Lösungen, Begründungen, Folgerungen, Interpretationen oder Wertungen zu gelangen.							

# 4. Der Prozess der Aufgabenentwicklung für die Bildungsstandards in Mathematik

Die Entwicklung eines Aufgabenpools für die Bildungsstandards in Mathematik stellt eine didaktische und methodische Herausforderung auf hohem Niveau dar. In Deutschland gibt es bislang allerdings keine professionellen Einrichtungen wie etwa in den Niederlanden das Testinstitut Cito (2006), die mit der deutschen Schulsituation umfassend vertraut sind und die entsprechende Aufgaben professionell entwickeln könnten. Hier stellt sich die Frage: wer kann die notwendigen Testaufgaben zur Erfassung der Bildungsstandards ausarbeiten? Da das von der Kultusministerkonferenz (KMK) geplante Institut zur Qualitätsentwicklung im Bildungswesen (IQB) in 2004 noch nicht eingerichtet war, erklärte sich das PISA-Konsortium 2006 bereit, eine Konzeption und ein Erhebungsverfahren zu erarbeiten und prototypisch für die Mathematik das Verfahren der Testentwicklung durchzuführen und zu erproben. Die Lösung bestand darin, eine Struktur von mehreren Fachgruppen zu etablieren, die sich jeweils aus Fachexperten und mit der Schullandschaft vertrauten Lehrkräften zusammensetzten. Das mit der Aufgabenentwicklung betraute PISA-Konsortium schlug ein Verfahren vor, das von den Auftraggebern aufgegriffen wurde. Die Abbildung 1 zeigt ein Organigramm der beteiligten Arbeitsgruppen.

Die vier regionalen Aufgabenentwicklungsgruppen hatten den Auftrag, ca. 800 Aufgaben nach den Vorgaben der theoretischen Rahmenkonzeption (Kultusministerkonferenz, 2004) zu entwickeln, zu überarbeiten und eine Vor-Klassifikation der Aufgaben vorzunehmen. Dabei sollten in jeder Regionalgruppe insgesamt jeweils 200 Aufgaben entwickelt werden, verteilt auf vier Abgabetermine innerhalb eines Jahres. Die Mitglieder in diesen Gruppen waren unterrichtserfahrene Lehrkräfte, die meist selbst auch aktuell unterrichten. Diese Voraussetzung ermöglichte es, dass die Lehrkräfte die entwickelten Aufgaben in der Schule pilotieren konnten. Ziel der Aufgabenpilotierung war es, die Tragfähigkeit der Aufgabenformulierung zu

prüfen, die Vielfalt potentieller Schülerlösungen festzustellen sowie verschiedene Lösungswege und typische Fehllösungen zu sammeln. Die Zusammensetzung der Regionalgruppen berücksichtigte neben den Lehrkräften außerdem Vertreterinnen und Vertreter aus Landesinstituten. Die Federführung in den vier regionalen Gruppen übernahm jeweils eine Leiterin oder ein Leiter mit Erfahrung in der Personalführung, die als Ansprechperson im engen Austausch mit der Koordinationsgruppe stand und die Rückmeldungen aus der Aufgabenbewertungsgruppe entgegennahm. Zusätzlich gehörte jeder Aufgabenentwicklungsgruppe eine Person aus der Fachdidaktik zur wissenschaftlichen Beratung an.

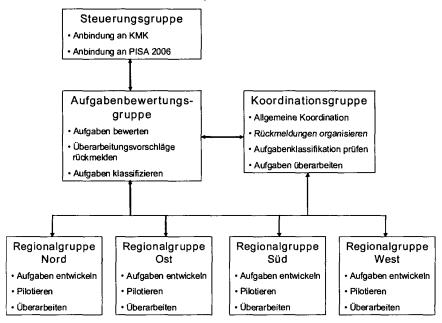


Abb. 1: Organigramm der Arbeitsgruppen für die Aufgabenentwicklung

Ziel der Arbeit in der Bewertungsgruppe war es, die in den Regionalgruppen entwickelten Aufgaben unter inhaltlichen und methodischen Gesichtspunkten zu prüfen und zu bewerten, ggf. gezielte Rückmeldung an die Aufgabenentwicklungsgruppen zur Überarbeitung zu geben, die Aufgaben nach den mathematischen Kompetenzen, Leitideen und Anforderungsbereichen zu klassifizieren sowie letztlich die Aufgaben für den Feldtest auszuwählen. Die Mitglieder dieser Gruppe sind Fachdidaktiker und Erziehungswissenschaftler, die sowohl mit der Analyse und Konstruktion von Aufgaben als auch mit der Konstruktion und Auswertung von Kompetenztests vertraut sind.

Die Steuerungsgruppe stellt durch den Vorsitzenden die Anbindung an die KMK-Fachkommission her und gewährleistet durch Prof. Dr. Manfred Prenzel (IPN, Kiel) und Prof. Dr. Werner Blum (Universität Kassel) die

Anknüpfung an PISA 2006. Der Steuerungsgruppe obliegt die Federführung des gesamten Prozesses der Aufgabenentwicklung für die Bildungsstandards in Mathematik.

Die Kasseler Koordinationsgruppe (Christina Drüke-Noe, Alexander Jordan, Dominik Leiß, Dr. Bernd Wiegand) begleitete die Aufgabenentwicklung und organisierte Bewertungs- und Rückmeldeprozesse an die Regionalgruppen. Darüber hinaus führte diese Gruppe auch Überarbeitungen an den Aufgaben durch und überprüfte die Aufgabenklassifikationen. Letztlich wurden hier alle entwickelten Ergebnisse gesammelt und stilistisch überarbeitet. Die Mitglieder dieser Gruppe zeichnen sich ebenfalls durch einen fachdidaktischen Hintergrund aus und verfügen über Kenntnisse in der Erstellung, Durchführung und Auswertung vergleichender Tests. Die Gruppe diente außerdem als permanenter Ansprechpartner für alle Beteiligten während des Entwicklungsprozesses.

Ergebnis des Arbeitsverlaufs, der in wenigen Monaten abgeschlossen werden musste, waren insgesamt ca. 1150 Aufgaben, die sich gemäß der Vorgaben über die Leitideen, Kompetenzen und Anforderungsbereiche verteilten. Nach Abschluss des Aufgabenentwicklungsprozesses wurden nun durch die Bewertungsgruppe aus diesem vorläufigen Aufgabenpool Items für den Feldtest ausgewählt. Die Auswahl erfolgte durch eine kriteriengeleitete Begutachtung. Danach wurden alle Aufgaben hinsichtlich des Grades ihrer Eignung als Testaufgabe auf einer vierstufigen Skala eingeschätzt. Die Antwortkategorien waren:

- 1. voll geeignet,
- 2. mit leichten Modifikationen geeignet,
- 3. nur nach umfangreicher Überarbeitung geeignet und
- 4. nicht geeignet.

Diese Ratings wurden von allen Mitgliedern der Bewertungsgruppe und der Kasseler Koordinationsgruppe abgegeben. In den Pool der Testitems wurden danach alle Aufgaben aufgenommen, die in der Mehrzahl mit "voll geeignet" eingeschätzt wurden. Aufgaben mit der Einschätzung "mit leichten Modifikationen geeignet" wurden nach der Einarbeitung konkreter Verbesserungsvorschläge durch die Bewertungsgruppe und die Kasseler Koordinationsgruppe als Testitems zugelassen. Aufgaben, die in der Mehrzahl mit "nur nach umfangreicher Überarbeitung geeignet" eingeschätzt wurden, wurden nur im Einzelfall geprüft und überarbeitet. Kriterium war, dass sie einen interessanten Aufgabenkontext enthielten oder Lücken in der mehrdimensionalen Rahmenstruktur auffüllen konnten. Aufgaben, die in der Mehrzahl als "nicht geeignet" eingeschätzt wurden, wurden nicht in den Pool der Testitems aufgenommen.

Insgesamt verlief der Prozess der Aufgabenentwicklung erfolgreich. Die eng gesteckten zeitlichen Vorgaben konnten im wesentlichen eingehalten werden. Die geforderte Aufgabenzahl wurde deutlich überschritten. Dennoch offenbarten sich während des Verlaufs einige Problembereiche. Die Regionalgruppen setzten sich im wesentlichen aus Lehrkräften zusammen, die eher beim Entwickeln von Aufgaben im Klassen- und Schulkontext Erfahrungen hatten. Dies schlug sich u.a. auch darin nieder, dass nicht alle Aufgaben als Testaufgaben geeignet waren. Viele Aufgaben waren zwar aus didaktischer Sicht als ausgesprochen wertvoll zu erachten, eigneten sich iedoch nur bedingt für den Einsatz in einem Testinstrument. Dies ist insbesondere der Fall, wenn die Kodierung der Antwort unverhältnismäßig aufwändig ist oder wenn die Aufgabenschwierigkeit so hoch ist, dass nur sehr wenige Schülerinnen und Schüler sie erfolgreich in einer Testsituation (Zeitdruck, keine Hilfestellung) lösen können. Defizite zeigten sich auch hinsichtlich der Ausgewogenheit von Leitideen, Kompetenzen und Anforderungsbereichen (vgl. Abschnitt 5). Hier spiegelte sich auch ein generelles Problem des Mathematikunterrichts wider, dass bestimmte Inhaltsbereiche und geforderte Kompetenzen deutlich stärker gewichtet werden als andere. Ferner zeigte sich, dass Mathematikaufgaben, die ein offenes und kein geschlossenes Antwortformat aufwiesen, aus test-pragmatischer Sicht zu häufig vertreten waren. Offene Antwortformate sind erheblich aufwändiger auszuwerten und können daher nur zu einem gewissen Anteil in einer Studie mit großen Fallzahlen berücksichtigt werden. Nicht unproblematisch wirkte sich auch der hohe Zeitdruck aus, unter dem die Aufgaben konzipiert worden sind.

Als Fazit zeigt sich dennoch, dass diese Entwicklungsstruktur geeignet ist und auch für andere Unterrichtsfächer angewandt werden könnte. Nötig bleiben jedoch Schulungen der Aufgabenentwickler, um ein adäquates Verständnis von Bildungsstandards und von Prinzipien der Konstruktion entsprechender Aufgaben zu erreichen. Die abschließende Überarbeitung der Aufgaben durch eine Bewertungsgruppe bleibt davon jedoch unberührt.

# 5. Von den Aufgaben zu einem Feldtestinstrument

Von den – durch die Bewertungsgruppe geprüften – Aufgaben wurden insgesamt 504 Items in den PISA 2006 Feldtest aufgenommen. Die Tabelle 4 zeigt die Verteilung der Aufgabenzahlen nach mathematischen Leitideen und Kompetenzen. Alle fünf Leitideen werden jeweils durch etwa 70-140 Aufgaben abgedeckt. Wie zu erwarten, sind die Leitideen "Funktionaler Zusammenhang" und "Messen" am stärksten vertreten. Um eine Aufgabe zu lösen, sind in der Regel mehrere mathematische Kompetenzen erforderlich. Deshalb wird jede Aufgabe mehreren Kompetenzen zugeordnet.

Am häufigsten berücksichtigt ist der Umgang mit symbolischen, formalen und technischen Elementen (K5). Am wenigsten gefordert werden die Kompetenzen "mathematisch argumentieren" (K1) und "mathematisch kommunizieren" (K6). Die Anzahl von Aufgaben pro Zelle reichen von 6

(L1 – Zahl und K6 – Mathematisch Kommunizieren) bis 61 (L4 – Funktionaler Zusammenhang und K5 – Mit Mathematik symbolisch/technisch umgehen). Ein vollständiges ausgeglichenes Facettendesign (Leitideen x Kompetenzen) wurde nicht systematisch angestrebt. Einzelne Zellen der Matrix sind nur mit wenigen Aufgaben besetzt. Die Anteile der Aufgaben in den einzelnen Anforderungsbereichen "Reproduzieren", "Zusammenhänge herstellen" und "Verallgemeinern und Reflektieren" verteilen sich etwa im Verhältnis 25/60/15 Prozent.

Tab. 4: Aufgabenzahlen nach Leitideen und Kompetenzen (im Feldtest)

	L1 - Zahl	L2 - Messen	L3 - Raum und Form	L4 - Funktionaler Zusammenhang	L5 - Daten und Zufall	Items pro Kompetenz
Items pro Leitidee	95	115	83	138	69	
K1 - Mathematisch argumentieren	10	13	11	12	14	60
K2 - Probleme mathematisch lösen	23	23	28	13	9	96
K3 - Mathematisch modellieren	20	35	8	38	22	123
K4 - Mathematische Darstellungen verwenden	9	38	42	48	35	172
K5 - Mit Mathematik symbolisch/technisch umgehen	43	47	18	61	9	178
K6 - Mathematisch kommunizieren	6	11	9	11	7	44

Nach der Entwicklung von Aufgaben bestand der nächste Schritt darin, die Mathematikaufgaben für die Bildungsstandards, angekoppelt an den Feldtest von PISA 2006, zu erproben (1. Testtag: PISA 2006, 2. Testtag: Bildungsstandards). Die Zusammenstellung der Aufgaben zu einzelnen Testheften ist vergleichbar mit der anderer groß angelegter Schulleistungsstudien.

Nicht alle Schülerinnen und Schüler können alle 504 Mathematikaufgaben innerhalb eines vertretbaren Zeitrahmens bearbeiten, da sich bei einer durchschnittlichen Bearbeitungsdauer von zwei Minuten pro Aufgabe eine rechnerische Testbearbeitungsdauer von rund 17 Stunden ergibt. Um die benötigte Testzeit zu verringern, aber gleichzeitig alle Aufgaben einsetzen zu können, wird analog zu PISA als Testdesign ein sogenanntes "Multi-Matrix-Design" gewählt (Prenzel, Drechsel, Carstensen & Ramm, 2004, S. 28). Dabei werden nach bestimmten Prinzipien unterschiedliche Testhefte mit systematisch variierten Aufgabenpacken erstellt und zufällig den Schülerinnen und Schüler zugewiesen. Das Multi-Matrix-Design des Tests gestattet es, empirische Kennwerte für die eingesetzten Aufgaben zu erhalten und Aussagen über die Kompetenzen in der Stichprobe zu treffen. Die Datentabelle, die man aufgrund des Multi-Matrix-Designs erhält, ist systematisch unvollständig. Die Auswertung benötigt daher ein spezielles Verfahren, das ein solches Testdesign berücksichtigt. Analog zum Auswertungsverfahren bei PISA wird auch zur Skalierung der Bildungsstandards auf die Verfahren der probabilistischen Item-Response-Theorie zurückgegriffen (OECD, 2005, S. 119; Rost, 2004b; Rost & Walter, 2004).

Um vergleichbare Aufgabenkennwerte im Rahmen eines Multi-Matrix-Designs zu erhalten, d.h. die Aufgabenschwierigkeiten der einzelnen Testhefte in der gleichen Metrik abbilden zu können, müssen die Testhefte durch gemeinsame Aufgaben "verlinkt" sein. Unter "Linking" versteht man in diesem Zusammenhang die Verbindung zwischen zwei Testheften durch Aufgaben, die in beiden Testheften enthalten sind. Aus diesem Grund berücksichtigt das Multi-Matrix-Design jede Aufgabe (jeden Aufgabenpacken) genau zwei Mal.

Im Feldtest für die Bildungsstandards wurden insgesamt 21 Testheftvarianten eingesetzt. In jedem Testheft befinden sich zwölf Packen mit jeweils im Durchschnitt vier Aufgaben. Jeder Packen kommt zweimal, an unterschiedlichen Positionen vor. Insgesamt ergibt sich demnach: 12 Packen x 4 Aufgaben x 21 Testhefte = 1008 = 2 x 504 Aufgaben. Die Testzeit beträgt insgesamt 120 Minuten, pro Packen im Durchschnitt 10 Minuten, pro Aufgabe 2,5 Minuten. Das Design des Feldtests veranschaulicht Abbildung 2. Darin ist die Zuordnung der Packen zu den 21 Testheften angegeben. Die Bearbeitung eines Testheftes teilt sich in zwei Phasen à 60 Minuten, die durch eine Pause geteilt werden.

Bei der Konzeption des Testdesigns ist es wichtig, mögliche Positionseffekte zu kontrollieren, denn auch die Position einer Aufgabe im Testheft beeinflusst ihre empirische Schwierigkeit. Aufgaben, die am Ende eines Testheftes stehen, werden möglicherweise aus Zeitgründen nicht mehr von allen Schülerinnen und Schülern bearbeitet. Die gemessene Aufgabenschwierigkeit erhöht sich so künstlich. Um solche Effekte zu kontrollieren und zu minimieren, kommen gleiche Aufgaben an mehreren Positionen in verschiedenen Testheften vor. Im vorliegenden Testdesign wurden die Packen innerhalb der fünf Leitideen systematisch auf alle sechs Packenpositionen (vor und nach der Pause) verteilt.

Ferner wurde berücksichtigt, dass die Schülerinnen und Schüler möglichst vier Leitideen bearbeiten. Das ist in allen Testheften der Fall, mit Ausnahme von Testheft 13, wo es nur drei Leitideen sind.

Durchgeführt wurde der Feldtest im Frühjahr 2005. Die Stichprobe bestand dabei aus 100 Schulen in den fünf Bundesländern Schleswig-Holstein, Hamburg, Nordrhein-Westfalen, Bayern und Thüringen. Berücksichtigt wurden bei den Schularten: 24 Hauptschulen, 15 Integrierte Gesamtschulen, 12 Schulen mit mehreren Bildungsgängen, 21 Realschulen sowie 28 Gymnasien. Die Bildungsstandards wurden jeweils in zwei kompletten 9. Klassen getestet. Insgesamt liegen damit für etwa 5000 Neuntklässler Daten vor. Der Feldtest fand an jeweils zwei Testtagen statt. Am ersten Testtag bearbeiteten die Schülerinnen und Schüler die internationalen PI-SA-Instrumente, am zweiten Testtag ging es um die Aufgaben für die Bildungsstandards. An dieser Stelle sei noch einmal angemerkt, dass das Ziel

des Feldtests ist, die Aufgaben und Instrumente zu erproben, nicht die Kompetenz von Jugendlichen zu erfassen.

											Testher										
Zelt	1	2	3	4	5	- 6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
10	L1-1	L1-7	L1-7	L1-13	L1-13	L1-19	L1-19	L1-1	L4-1	L4-7	L4-7	L4-13	L4-13	L4-19	L4-19	L4-25	L4-25	L4-31	L5-1	L5-7	L5-7
10	L1-2	L1-6	L1-8	L1-12	L1-14	L1-18	L1-20	L1-5	L4-2	L4-6	L4-8	L4-12	L4-14	L4-18	L4-20			L4-30	L5-2	L5-6	L5-8
10	L1-3	L1-6	L1-9	L1-11	L1-15	L1-17	L1-21	L1-23	L4-3	L4-6	L4-9	L4-11	L4-15	L4-17	L4-21		L4-27	L4-29	1.5-3	L5-5	L5-9
10	L1-4	L1-4	L1-10	L1-10	L1-16	L1-16	L1-22	L1-22	L4-4	L4-4	L4-10	L4-10	L4-18	L4-16	L4-22	L4-22	L4-28	L4-28	L5-4	L5-4	L5-10
10	L4-2	L4-6	L4-8	L4-12	L4-14	L4-16	L4-20	L4-24	L1-2	L1-6	L1-8	L1-12	L1-14	L1-18	L1-20	L2-32	L5-2	L5-6	L4-1	L4-26	L4-30
10	L4-27	L4-29	L4-3	L4-5	L4-9	L4-11	L4-15	L4-17	L1-9	L1-11	L1-15	L1-17	L1-21	L1-23	L1-3	L2-5	L5-9	L5-11	L4-31	L4-21	L4-23
10											Pause										
10	L5-13	L5-13	L5-1	12-1	L2-7	12-7	L2-13	L2-13	L2-19	L2-19	L2-25	L2-25	L2-31	L2-31	L3-1	L3-7	L3-7	L3-13	L3-13	L3-19	L3-19
10	L5-12	L5-14	L5-18	L2-2	L2-6	L2-8	L2-12	L2-14	L2-16	L2-20	L2-24	1.2-26	L2-30	12-2	L3-2	L3-8	L 3-8	L3-12	L3-14	L3-18	L3-2
10	L5-11	L5-15	L5-17	L2-3	L2-5	L2-9	L2-11	L2-15	L2-17	L2-21	L2-23	L2-27	L2-29	L2-32	L3-3	L3-5	L 3-9	L3-11	L3-15	L3-17	L3-20
10	L5-10	L5-16	L5-16	1.2-4	L2-4	L2-10	L2-10	L2-16	L2-18	L2-22	L2-22	L2-28	L2-28	12-33	L3-4	L3-4	L3-10	L3-10	L3-15	L3-16	L3-21
10	L2-1	L2-6	L2-8	L5-8	L5-12	L5-14	L5-18	L3-1	L3-6	L3-8	L3-12	L3-14	L3-18	L3-20	L2-12	L2-14	L2-18	L2-20	L2-24	L2-26	L 2-30
10	L2-9	L2-11	L2-15	L5-15	L5-17	L5-3	L5-5	L3-9	L3-11	L3-15	L3-17	L3-21	L3-3	L3-5	L2-17	L2-21	L2-23	L2-27	L2-29	L2-33	L2-3

Abb. 2: Verteilung der Aufgabenpacken nach Leitideen im Multi-Matrix-Design (Feldtest)

Abk.: Leitideen: L1 (Zahl, 23 Packen), L2 (Messen, 33 Packen), L3 (Raum und Form, 21 Packen), L4 (Funktionaler Zusammenhang, 31 Packen), L5 (Daten und Zufall, 18 Packen).

# 6. Vom Feldtest zum Haupttest

Welche Kriterien sind jetzt für die Auswahl der Aufgaben für den Haupttest entscheidend? Die Schülerantworten vom Feldtest werden analog zum methodischen Vorgehen bei PISA, nach Verfahren der Item-Response-Theory skaliert (Carstensen, Knoll, Rost & Prenzel, 2004; OECD, 2005). Die dabei eingesetzte Software "ConQuest" (Wu, Adams & Wilson, 1997) liefert eine Reihe von Aufgabenkennwerten, die für die Bewertung der empirischen Qualität relevant sind. Sie lassen sich in zwei Bereiche zusammenfassen:

#### 6.1. "Klassische" Kriterien zur Itemauswahl

Die beiden wichtigsten Kennwerte für die Itemauswahl sind die Aufgabenschwierigkeit und die Trennschärfe. Die Aufgabenschwierigkeit wird durch die absolute oder prozentuale Lösungshäufigkeit ausgedrückt. Wie viel Prozent der Schülerinnen und Schüler können eine Aufgabe lösen? Die Trennschärfe ist die punkt-biseriale Korrelation einer Aufgabe mit der Gesamtleistung (Item-Test-Korrelation). Wie hängt die Lösungshäufigkeit einer Aufgabe mit der Gesamttestleistung zusammen? Wie gut differenziert ein Item zwischen verschiedenen Eigenschaftsausprägungen von Personen?

## 6.2. Itemauswahlkriterien der Item-Response-Theory

Im Rahmen der Item-Response-Theory werden Modelle beschrieben, die einen funktionalen Zusammenhang zwischen der Personenfähigkeit und der Schwierigkeit einer Aufgabe modellieren. Personenfähigkeit und Aufgabenschwierigkeit werden dabei gleichzeitig als Modellparameter geschätzt, wobei im Rasch-Modell gleiche Trennschärfe für alle Aufgaben angenommen wird. Abweichungen von der beobachteten Trennschärfe einzelner Items können durch verschiedene Fit-Statistiken untersucht und beurteilt werden (OECD, 2005; Rost, 2004a).

Diese Aufgabenkennwerte wurden für alle Feldtest-Aufgaben berechnet und ausgewertet. Für den Haupttest wurden insgesamt 313 Items aufge-

nommen, die sich wie folgt auf die fünf Leitideen verteilen: L1: 60 Aufgaben, L2: 66 Aufgaben, L3: 54 Aufgaben, L4: 87 Aufgaben und L5: 46 Aufgaben.

	l														es the	Ř													
Zeit	□-	2	3	4	5	_6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
20	1.2-1	L2-2	123	L2-4	L2-5	L3-1	L3-2	L3-3	L3-4	L3-5	L5-1	L5-2	L5-3	L5-4	15-5	L4-1	L4-2	L4-3	L4-4	L4-5	L4-6	L4-7	L1-1	L1-2	L1-3	L1-4	L1-5	L1-€	L1-7
20	L1-7	1,2-1	L2-2	L2-3	L2-4	L2-5	L3-1	L3-2	L3-3	L3-4	L3-5	L5-1	L5-2	L5-3	L5-4	L5-5	L4-1	L4-2	L4-3	14-4	L4-5	L4-6	L4-7	L1-1	L1-2	L1-3	L1-4	L1-5	L1-6
20	L1-4	L1-5	L1-6	L1-7	L2-1	L2-2	L2-3	L2-4	L2-5	L3-1	L3-2	L3-3	L3-4	L3-5	L5-1	L5-2	L5-3	L5-4	L5-5	L4-1	L4-2	L4-3	L4-4	L4-5	L4-6	L4-7	L1-1	L1-2	L1-3
10															ause														
20	L3-2	L3-3	134	L3-5	L5-1	L5-2	L5-3	L5-4	L5-5	L4-1	L4-2	L4-3	L4-4	L4-5	L4-6	L4-7	L1-1	L1-2	L1-3	L1-4	L 1-5	L1-6	L1-7	L2-1	L2-2	L2-3	L2-4	12.5	L3-1
20	L5-4	L5-5	L4-1	L4-2	L4-3	14-4	L4-5	L4-6	L4-7	L1-1	L1-2	L1-3	L1-4	L1-5	L1-6	L1-7	L2-1	L2-2	L2-3	L2-4	L2-5	L3-1	L3-2	เมง	L3-4	L3-5	L5-1	L5-2	L5-3
20	L4-6	L4-7	L1-1	L1-2	L1-3	L1-4	L1-5	L1-6	L1-7	L2-1	L2-2	L2-3	L2-4	L2-5	L3-1	L3-2	L3-3	L3-4	L3-5	L5-1	L5-2	L5-3	L5-4	L5-5	L4-1	L4-2	L4-3	14-4	L4-5

Abb. 3: Verteilung der Aufgabenpacken nach Leitideen im Multi-Matrix-Design (Haupttest)

Abk.: Leitideen: L1 (Zahl, 7 Packen), L2 (Messen, 5 Packen), L3 (Raum und Form, 5 Packen), L4 (Funktionaler Zusammenhang, 7 Packen), L5 (Daten und Zufall, 5 Packen).

Die Abbildung 3 zeigt das Testdesign für den Haupttest am zweiten Testtag von PISA 2006. Wie im Feldtest sind die Testaufgaben in einzelne Packen zusammengefasst. Jeder Packen enthält etwa 10 Aufgaben aus einer Leitidee und beansprucht 20 Minuten Testzeit. Priorität bei der Festlegung des Multi-Matrix-Designs für den Haupttest hatten folgende Punkte:

- Die Aufgaben innerhalb der Leitideen sollten möglichst gut verlinkt sein, um eine hohe Messqualität in den einzelnen Inhaltsbereichen zu gewährleisten.
- Zwischen den Leitideen sollte es ebenfalls eine gute Verlinkung geben, um die Leitideen als einzelne Dimensionen möglichst gut abzubilden.
- Innerhalb jedes einzelnen Testheftes sollten möglichst alle fünf Leitideen enthalten sein.
- Durch vollständige Rotation der Packen sollten Positionseffekte kontrolliert werden.

#### 7. Auswertungsstrategien

In der Diskussion über die methodische Auswertung von Bildungsstandards werden verschiedene Vorgehensweisen thematisiert. Rost (2004b) unterscheidet bei der Auswertung von Kompetenzmodellen für die psychometrische Erfassung von Bildungsstandards drei Arten von Modellen: Kompetenzstufen, Kompetenzdimensionen und Kompetenzmuster.

Bei Kompetenzstufenmodellen wird das gemessene Fähigkeitskontinuum in Intervalle unterteilt, die mit unterschiedlichen Leistungsniveaus korrespondieren und deshalb auch als Kompetenzstufen bezeichnet werden. Dieses Verfahren wurde beispielsweise in PISA 2003 angewendet. Die Kompetenzskala in Mathematik unterscheidet hier zwischen sechs Kompetenzstufen (OECD, 2004).

Mehrdimensionale Modelle berücksichtigen mehrere Dimensionen von Kompetenzen gleichzeitig. Es lassen sich dadurch Kompetenzprofile erfas-

sen, die Stärken und Schwächen in einzelnen Dimensionen von Subgruppen aufzeigen können. Der nationale Naturwissenschaftstest in PISA 2003 ist ein aktuelles Anwendungsbeispiel für ein zweidimensionales Facettenmodell (Rost, Walter, Carstensen, Senkbeil & Prenzel, 2004; Rost, Walter, Carstensen, Senkbeil & Prenzel, 2005; Senkbeil, Rost, Carstensen & Walter, 2005). Das Facettendesign kreuzt systematisch sieben kognitive Komponenten (z.B. Umgang mit Grafiken) und drei Inhaltsbereiche (Biologie, Physik, Chemie).

Um Kompetenzmuster von Teilkompetenzen zu identifizieren, können so genannte Mischverteilungsmodelle (Mixed-Rasch-Model) angewendet werden (Rost, 2004a). Sie teilen Personen in Gruppen mit gleichem Kompetenzprofil und schätzen gleichzeitig das quantitative Kompetenzniveau innerhalb der Gruppen. Einen ersten Schritt in diese Richtung unternimmt Kleine (2004; 2005) am Beispiel der Bildung von Kompetenzstufen mit Hilfe einer Latent-Class-Analyse.

Im Gegensatz zu den beiden ersten methodischen Ansätzen erlaubt die Methode der Mischverteilungsmodelle eine qualitative Unterscheidung von Personengruppen. Die ersten beiden Ansätze beruhen hingegen auf einer quantitativen Leistungsmessung.

Für die geplante Auswertung der Bildungsstandards in Mathematik ist vorgesehen, ein Kompetenzstufenmodell für die fünf Leitideen zu konstruieren. Damit lassen sich Kompetenzprofile über die fünf Inhaltsbereiche abbilden. In einem zweiten Schritt ist zu prüfen, ob es möglich ist, mit einem zweidimensionalen Facettenmodell Profile für Leitideen und mathematische Kompetenzen zu erfassen.

#### 8. Diskussion und Ausblick

Der vorliegende Beitrag zeigte die wichtigsten Schritte zur Umsetzung der Bildungsstandards in der Mathematik auf, angefangen mit dem KMK-Beschluss von 2003 bis hin zum Testdesign der Hauptuntersuchung zu den Bildungsstandards am zweiten Testtag von PISA 2006. Die Erfahrungen mit der organisatorischen Struktur zur Aufgabenentwicklung sowie die Vorgaben einer mehrdimensionalen Rahmenstruktur für die Bildungsstandards sind überwiegend positiv. Während des Verlaufs der Aufgabenentwicklung wurde jedoch auch deutlich, wie schwierig es sich gestaltet. Aufgaben zu entwickeln, die auch empirisch die Konzepte eines theoretischen Rahmenmodells abdecken. Die Entwicklung von Testitems ist eine höchst anspruchsvolle Aufgabe, die in anderen Staaten von professionellen Instituten übernommen wird. Letztlich wurde aber mit der – aus vielen Fachexperten bestehende - Gruppenstruktur ein paradigmatisches Verfahren erfolgreich erprobt, das auch für die Entwicklung von Aufgaben für andere Fächer (z. B. aus den Naturwissenschaften) übertragen werden kann. Mit der Entwicklung von Testinstrumenten für die Bildungsstandards wurde so ein wichtiger Schritt unternommen, sich stärker am Output von Bildungssystemen zu orientieren.

Das nächste Vorgehen besteht darin, mit Hilfe des entwickelten, erprobten und normierten Aufgabenpools, die Einhaltung der Standards sowohl landesweit als auch länderübergreifend zu überprüfen und zurückzumelden. Für die Organisation und Unterstützung der Länder bei dieser Aufgabe ist das bundesweit tätige, von den Ländern gemeinsam getragene "Institut zur Qualitätsentwicklung im Bildungswesen – Wissenschaftliche Einrichtung der Länder an der Humboldt-Universität zu Berlin" (IQB) gegründet worden (Köller, 2005).

Neben diesen eher technischen und organisatorischen Schritten wird die Implementation der Bildungsstandards vor allem eine große Herausforderung in den Schulen sein. Hier gilt es die Perspektiven zu nutzen, die sich mit der Einführung der Bildungsstandards für die Schul- und Unterrichtsentwicklung eröffnen. Durch die Bildungsstandards werden Anforderungen an das Lehren und Lernen in der Schule konkretisiert und erwünschte Lernergebnisse bei den Schülerinnen und Schülern definiert. Durch die alleinige Rückmeldung der erreichten Kompetenzen von Jugendlichen werden jedoch Leistungen nicht automatisch verbessert. Hierfür bedarf es einer systematischen Schulentwicklung mit interner und externer Evaluation (Kultusministerkonferenz, 2004, S. 10f). Die Bildungsstandards stellen hier quasi eine externe Evaluation dar, die sich durch klare Maßstäbe und eine standardisierte Rückmeldung zu Unterrichtsergebnissen auszeichnet. Hiermit verbunden ist die Möglichkeit, auch auf die interne Evaluation innerhalb einer Schule zu wirken (vgl. dazu das ausführliche Modell von der Leistungsevaluation zur Unterrichtsentwicklung in Helmke & Schrader, 2001). Die Bildungsstandards fördern den Blick auf langfristige, kumulativ angelegte Lernprozesse und Lernergebnisse. Innerhalb der Schulen sollen sie dazu beitragen, die Kommunikation und die Kooperation der Lehrkräfte zu fördern und ihr professionelles Handeln zu verbessern. Die in den Bildungsstandards systematisch und fachdidaktisch begründeten Kompetenzmodelle (Klieme & Steinert, 2004, S. 137) bieten Anlässe und Möglichkeiten für professionelles Lernen in der Schule. Diskussionen und regelmäßiger Austausch über Unterrichtspraktiken können dazu beitragen, dass erfolgreiche Unterrichtsmodelle schulintern weiterentwickelt werden und in Richtung einer neuen Unterrichts-, Aufgaben- und Kooperationskultur wirken (Blum et al., 2005).

In diesem Zusammenhang wird sich auch die Frage stellen, wie innerhalb der Bildungssysteme der Länder geeignete Unterstützungsmaßnahmen für die Schulen aussehen können. Wichtig wird sein, die Autonomie und Eigenverantwortung von Schulen zu verstärken. Auch müssen Schulen mit besonderen Rahmenbedingungen (z. B ein hoher Migrantenanteil) in die Lage versetzt werden, auf problematische Befunde etwa durch verstärkte

Fördermaßnahmen reagieren zu können. Inwieweit diese Herausforderungen innerhalb und außerhalb der Schulen umgesetzt werden, wird auch an den Ergebnissen eines regelmäßigen Bildungsmonitoring abzulesen sein.

#### Literatur

- Baumert, J. & Lehmann, R. (Hrsg.). (1997). TIMSS Mathematischnaturwissenschaftlicher Unterricht im internationalen Vergleich. Opladen: Leske & Budrich.
- Biermann, M. & Blum, W. (2001). Eine ganz normale Mathe-Stunde? Was "Unterrichtsqualität" konkret bedeuten kann. Mathematik lehren, 108, 52-54.
- Blum, W., Drüke-Noe, C., Leiß, D., Wiegand, B. & Jordan, A. (2005). Zur Rolle von Bildungsstandards für die Qualitätsentwicklung im Mathematikunterricht. Zentralblatt für Didaktik der Mathematik, 37(4), 267-274.
- Blum, W., Drüke-Noe, K., Hartung, R. & Köller, O. (Hrsg.). (in Druck). Bildungsstandards Mathematik konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsideen und Fortbildungsmöglichkeiten. Berlin: Cornelsen/Scriptor.
- Blum, W. & Leiß, D. (2005). Modellieren im Unterricht mit der "Tanken"-Aufgabe, Mathematik lehren, 128, 18-21.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. & Carstensen, H. C. (2004). Mathematische Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost, U. Schiefele & P.-K. Deutschland) (Hrsg.), PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland Ergebnisse des zweiten internationalen Vergleichs (S. 47-92). Münster: Waxmann.
- Carstensen, C. H., Knoll, S., Rost, J. & Prenzel, M. (2004). Technische Grundlagen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland Ergebnisse des zweiten internationalen Vergleichs (S. 371-387). Münster: Waxmann.
- Cito. (2006). [Homepage]. Internet: http://www.cito.nl/ [15.05.2006].
- Helmke, A. & Schrader, A. (2001). Von der Leistungsevaluation zur Unterrichtsentwicklung. In R. Silbereisen & M. Reitzle (Hrsg.), Psychologie 2000.
  Bericht über den 42. Kongress der Deutschen Gesellschaft für Psychologie in Jena (S. 594-606). Lengerich: Pabst.
- Kleine, M. (2004). Quantitative Erfassung von mathematischen Leistungsverläufen in der Sekundarstufe I. Hildesheim: Franzbecker.
- Kleine, M. (2005). Latent-Class-Analyse: Ein Bindeglied zwischen Empirie und Theorie zur quantitativen Erfassung mathematischer Leistungen. Journal für Mathematikdidaktik, 26(2), 97-113.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, H., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H. E. & Vollmer, J. (2003). Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Bonn: Bundesministerium für Bildung und Forschung.
- Klieme, E. & Steinert, B. (2004). Einführung der KMK-Bildungsstandards. Zielsetzungen, Konzeptionen und Einführung in den Schulen am Beispiel der Mathematik. Der mathematische und naturwissenschaftliche Unterricht. MNU, 57(3), 132-137.

- Köller, O. (2005). Neue Besen kehren gut: Das Institut für Qualitätsentwicklung im Bildungswesen soll die nationalen Bildungsstandards in Deutschland überprüfen. Psychologie in Erziehung und Unterricht(4), 281-286.
- Kultusministerkonferenz. (2003). Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Beschluss vom 4.12.2003. Luchterhand.
- Kultusministerkonferenz. (2004). Bildungsstandards der Kultusministerkonferenz. Erläuterungen zur Konzeption und Entwicklung.
- NCTM. (2000). National Counsil of Teachers of Mathematics: Principles and standards for school mathematics. Reston, VA: NCTM.
- OECD. (2003). The PISA 2003 Assessment Framework. Mathematics, Reading, Science and Problem Solving Knowledge and Skills. Paris: OECD.
- OECD. (2004). Lernen für die Welt von morgen. Erste Ergebnisse von PISA 2003. Paris: OECD.
- OECD. (2005). PISA 2003 Technical Report. Paris: OECD.
- Prenzel, M. (2005). Bildungsstandards und die Weiterentwicklung des Mathematikunterrichts. In H.-W. Henn & G. Kaiser (Hrsg.), Mathematikunterricht im Spannungsfeld von Evolution und Evaluation. Festschrift für Werner Blum (S. 261-271). Hildesheim: Franzbecker.
- Prenzel, M., Drechsel, B., Carstensen, C. H. & Ramm, G. (2004). PISA 2003 eine Einführung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland Ergebnisse des zweiten internationalen Vergleichs (S. 13-46). Münster: Waxmann.
- Prenzel, M. & Ostermeier, C. (2003). Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts Ein unterrichtsbezogenes Qualitätsentwicklungsprogramm. Beiträge zur Lehrerbildung. Fachzeitschrift zu Theorie und Praxis der Aus- und Weiterbildung von Lehrerinnen und Lehrern, 21(2), 265-276.
- Rost, J. (2004a). Lehrbuch Testtheorie Testkonstruktion (2., vollst. überarb. u. erw. Aufl.). Bern: Hans Huber.
- Rost, J. (2004b). Psychometrische Modelle zur Überprüfung von Bildungsstandards anhand von Kompetenzmodellen. Zeitschrift für Pädagogik, 50(5), 662-678.
- Rost, J. & Walter, O. (2004). Multimethod Item Response Theory. In M. Eid & E. Diener (Hrsg.), Handbook of Psychological Measurement: A Multimethod Perspective (S. 249-268): American Psychological Association.
- Rost, J., Walter, O., Carstensen, C. H., Senkbeil, M. & Prenzel, M. (2004). Naturwissenschaftliche Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland Ergebnisse des zweiten internationalen Vergleichs (S. 111-146). Münster: Waxmann.
- Rost, J., Walter, O., Carstensen, C. H., Senkbeil, M. & Prenzel, M. (2005). Der nationale Naturwissenschaftstest PISA 2003. Der mathematische und naturwissenschaftliche Unterricht, 58(4), 196-204.
- Senkbeil, M., Rost, J., Carstensen, C. H. & Walter, O. (2005). Der nationale Naturwissenschaftstest PISA 2003. Entwicklung und empirische Überprü-

fung eines zweidimensionalen Facettendesigns. Empirische Pädagogik, 19, 1-24

Wu, M., Adams, R. J. & Wilson, M. (1997). ConQuest: Multi-Aspect Test Software. Camberwell: Australian Council for Educational Research.

#### Anschrift der Autoren:

Dr. Timo Ehmke, Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN), Universität Kiel, Olshausenstr. 62, 24098 Kiel, Tel.: 0431/880-4090, E-Mail: ehmke@ipn.uni-kiel.de

Dominik Leiß, Universität Kassel, Fachbereich Mathematik, Heinrich-Plett-Str. 40, 34109 Kassel, Tel.: 0561/8044308, E-Mail: dleiss@mathematik.uni-kassel.de

Prof. Dr. Werner Blum, Universität Kassel, Fachbereich Mathematik, Heinrich-Plett-Str. 40, 34109 Kassel, Tel.: 0561/8044623, E-Mail: blum@mathematik. uni-kassel.de

Prof. Dr. Manfred Prenzel, Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN), Universität Kiel, Olshausenstr. 62, 24098 Kiel, Tel.: 0431/880-3111, E-Mail: prenzel@ipn.uni-kiel.de