

Tätigkeitsbericht 2007–2013

Kompetenz in der Bildung.

Die Arbeit des Instituts zur Qualitätsentwicklung im Bildungswesen.



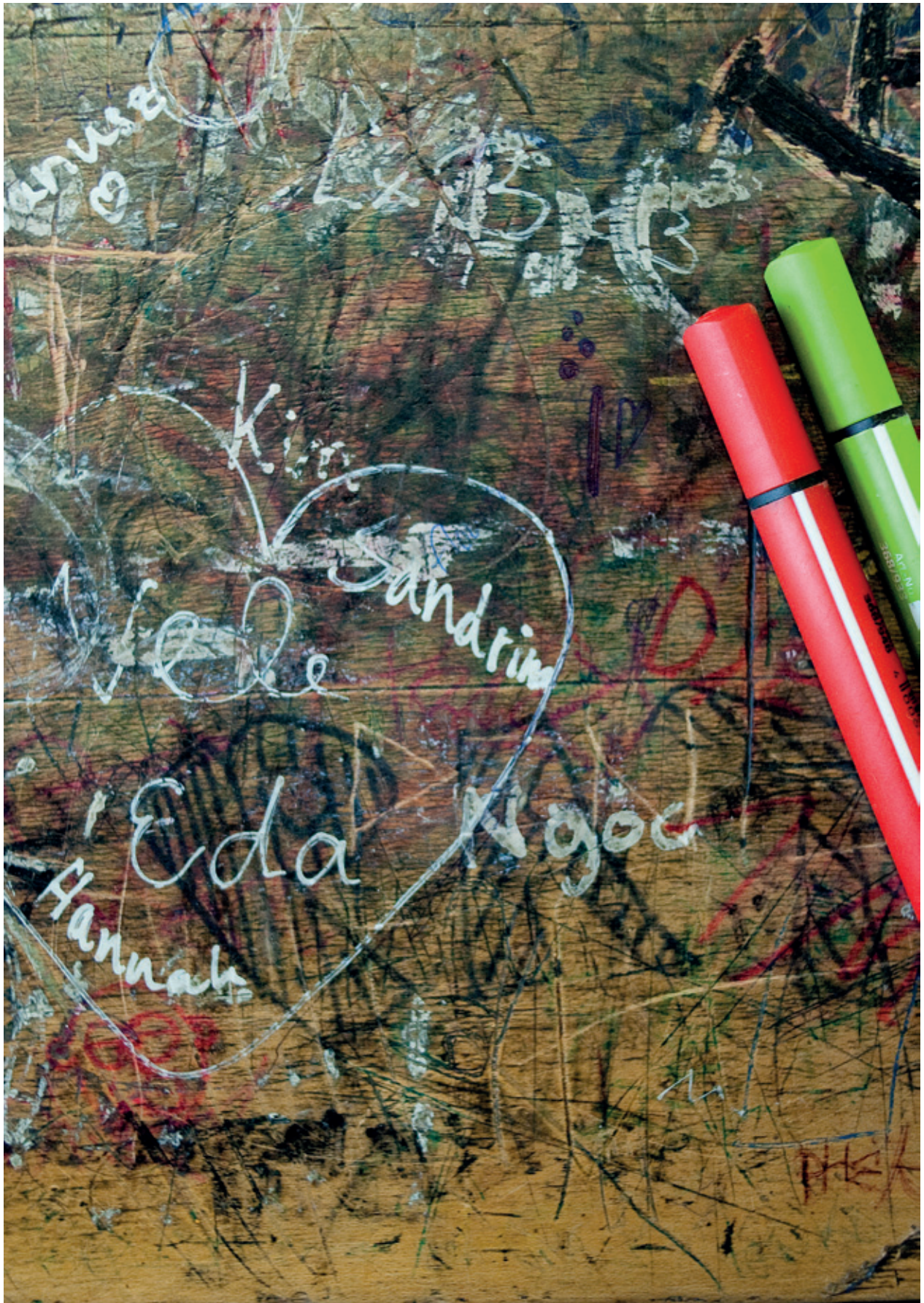


„Kompetenz in der Bildung“



Inhalt

Vorwort	7
Tätigkeitsbericht 2007–2013	
o1 // Das Institut zur Qualitätsentwicklung im Bildungswesen	11
o2 // Das Kerngeschäft des IQB	27
Bildungsstandards und ihre Operationalisierung mit Aufgaben	
o3 // Ländervergleichsstudien	47
Zentrale Überprüfung des Erreichens der Bildungsstandards	
o4 // Vergleichsarbeiten	57
Ein wichtiger Bestandteil der Gesamtstrategie	
der Kultusministerkonferenz zum Bildungsmonitoring	
o5 // Forschung	69
o6 // Der Beitrag des IQB zum wissenschaftlichen Austausch	85
Das FDZ und die Koordinierungsstelle	
o7 // Publikationen	93
o8 // Anhang	119
Impressum	130



Vorwort

Tätigkeitsbericht 2007 – 2013

„Kompetenz in der Bildung“ – unter diesem Leitspruch will der vorgelegte Tätigkeitsbericht auf die zurückliegenden sieben Jahre Arbeit des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) an der Humboldt-Universität zu Berlin zurückblicken. Mit der Doppelsinnigkeit von „Kompetenz in der Bildung“ wollen wir bewusst beide Aspekte würdigen: zum einen die Bedeutung, die der Kompetenzbegriff in den letzten Jahren in der wissenschaftlichen und schulpraktischen Arbeit erlangt hat, und zum anderen die Kompetenz der Mitarbeiterinnen und Mitarbeiter des IQB, die sie im Bemühen um eine bessere Schul- und Unterrichtsqualität entwickelt haben und einsetzen.



Nach den großen internationalen Schulleistungstudien im letzten Jahrzehnt – PISA, TIMSS und IGLU – und der 2013 veröffentlichten PIAAC-Studie bei Erwachsenen hat sich Kompetenz zu einer Art Schlüsselkonzept für Zielerwartungen im Bildungssystem und für die Evaluation der Ergebnisse von Bildungsprozessen auf allen Ebenen entwickelt. Das „Kompetent-Sein“ von Schülerinnen und Schülern gilt als *der* Qualitätsindikator für ganze Bildungssysteme – von Bundesländern wie Staaten – und gleichzeitig auch als ein wichtiger Beurteilungsmaßstab für die Arbeit von einzelnen Schulen und Lehrkräften.

Wenn ein Konzept mit derartigen Erwartungen an seine universelle Nutzbarkeit in der Bildungs-, Schul- und Unterrichtsforschung wie auch der schulischen Praxis aufgeladen wird, dann birgt dies politische und wissenschaftliche Chancen und Risiken gleichermaßen. Mit unserer Arbeit am IQB versuchen wir, die positiven Potenziale des Kompetenzkonzepts, d. h. dessen Klärungs- und Orientierungsfunktion zu nutzen, um die Debatte über Bildungsqualität aktiv mitzugestalten. Die Frage, was denn nun genau zu einem modernen Verständnis mathematischer oder fremdsprachlicher Kompetenzen gehören soll, hat zu ausgesprochen kontroversen, am

Ende aber produktiven Diskussionen zwischen Fachdidaktik, psychologisch-pädagogischer Bildungsforschung und bildungsadministrativer Praxis geführt. Das IQB hat diesen Diskussionsprozess in den letzten Jahren vielfach moderiert und inhaltlich mitgeprägt. Beispielhaft hierfür sind die im Zuge der Bildungsstandardentwicklung entstandenen Kompetenzstufenmodelle für Deutsch, Mathematik, Englisch, Französisch und die drei naturwissenschaftlichen Fächer. Hieran waren Fachdidaktikerinnen und Fachdidaktiker aus über 30 deutschsprachigen Universitätsinstituten gemeinsam mit Vertreterinnen und Vertretern aus Schulpraxis und Bildungsverwaltung beteiligt. Hinzu kommt, dass es in den letzten Jahren in der direkten Zusammenarbeit von Fachdidaktiken und psychologisch orientierter Messtheorie (Psychometrie) einen enormen Professionalisierungsschub in Deutschland bei der Entwicklung guter Testaufgaben gegeben hat. Im Zuge der Überprüfung, inwieweit die Bildungsstandards in den schulischen Kernfächern erreicht werden, hat das IQB bundesweit mehrere hundert Lehrkräfte geschult und in den Aufgabenentwicklungsprozess eingebunden.

Die Neuausrichtung von Kerncurricula, Rahmenlehrplänen, Lehreraus- und Fortbildungskonzepten, Schulqualitätsrahmen für Maßnahmen der Selbst- und Fremdevaluation, Lehrwerken und Unterrichtsmaterialien – alles unter dem Schlagwort Kompetenzorientierung – belegen die beispiellose Orientierungsfunktion, die dem Kompetenzbegriff auf fast allen Ebenen des Bildungssystems zugesprochen und zugemutet wird. In viele dieser Prozesse sind die Ergebnisse der IQB-Arbeit direkt oder indirekt eingeflossen. Im Jahr 2010 schließlich hat das IQB eine weitere Aufgabe übernommen, für die wir als Institut Neuland betreten mussten: die Entwicklung von Bildungsstandards für die Allgemeine Hochschulreife in den Fächern Deutsch, Mathematik und fortgeführte Fremdsprache (Englisch, Französisch), seit 2013 verbunden mit der Erstellung eines Pools von Abiturprüfungsaufgaben für ganz Deutschland. Damit sollen nun erstmals nicht nur Test- und Lernaufgaben, sondern auch Prüfungsaufgaben auf der Grundlage von Kompetenzen und Standards länderübergreifend entwickelt werden. Dieser Prozess wird vom IQB koordiniert.

In den Kapiteln dieses Berichts wird über die Aktivitäten des IQB in den Jahren 2007 bis 2013 in allgemein verständlicher Weise informiert. Dabei wird sichtbar, dass unsere Arbeit eine ausbalancierte Mischung aus wissenschaftlichen Serviceleistungen für die 16 Länder in der Bundesrepublik Deutschland sowie anspruchsvollen Forschungsleistungen im Austausch mit der *Scientific Community* erfordert. Das eine kann ohne das andere nicht auf hohem Niveau gewährleistet werden.

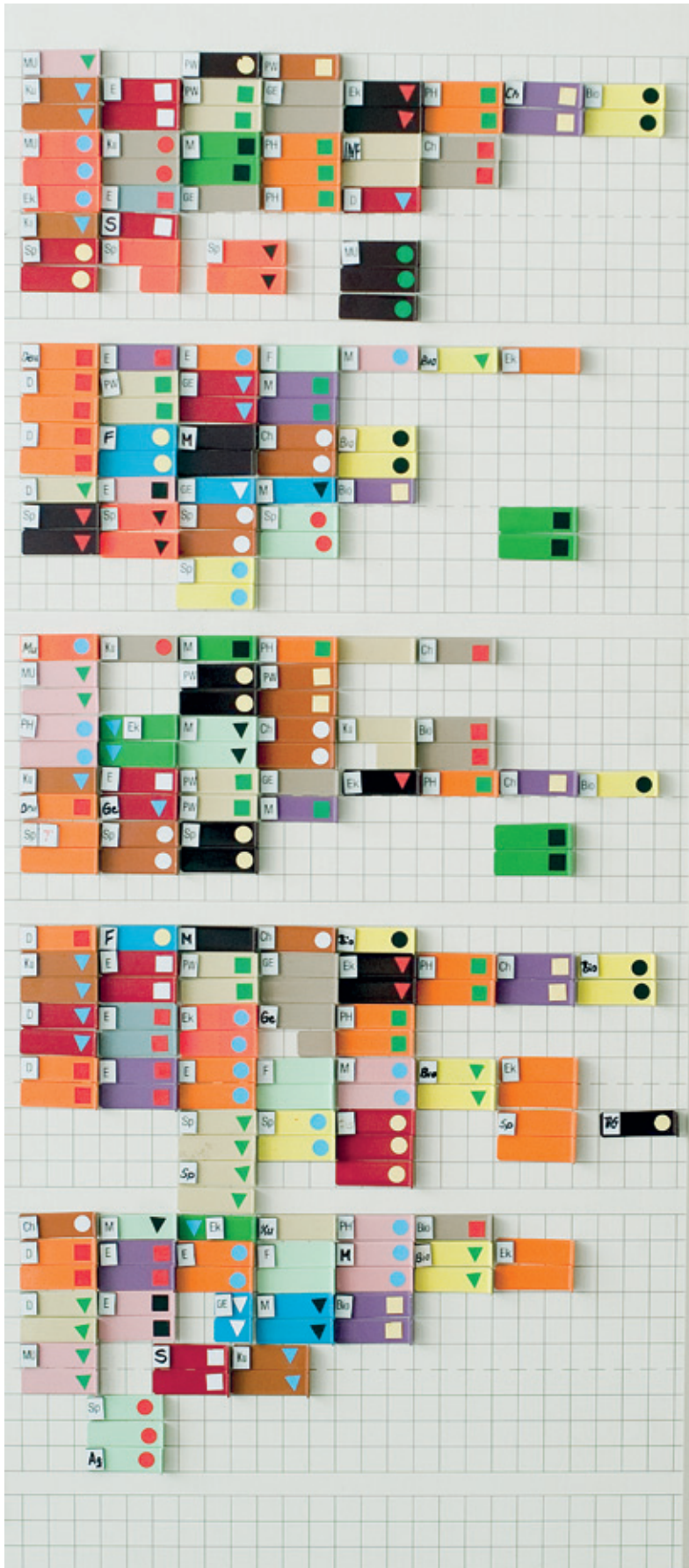
Natürlich hängt die Effizienz einer wissenschaftlichen Einrichtung wie dem IQB von zahlreichen Rahmenbedingungen und Personen und deren Zusammenwirken ab. Die Arbeit des IQB erfolgt in enger Abstimmung mit den Gremien und mit dem Sekretariat der Kultusministerkonferenz (KMK) sowie dem Kuratorium des Instituts, das sich aus dem Generalsekretär der KMK, drei Wissenschaftlern (Prof. Dr. Jürgen Baumert [seit Sommer 2014 Prof. Dr. Ewald Terhart], Prof. Dr. Detlev Leutner, Prof. Dr. Heinz-Elmar Tenorth) und den beiden Vorsitzenden der Amtschefscommission „Qualitätssicherung in Schulen“ (aktuell StS Burkhard Jungkamp, MD Dr. Peter Müller) zusammensetzt. Mit den Genannten ist eine vertrauensvolle und konstruktive, gelegentlich auch durch kontroverse Diskussionen geprägte Zusammenarbeit möglich, die das Arbeiten an der Schnittstelle von Kultusadministration, Politik und Wissenschaft interessant und lohnenswert macht. Auch die Unterstützung des Präsidiums und der Verwaltung der Humboldt-Universität sowie der Austausch mit den wissenschaftlichen Kolleginnen und Kollegen vor Ort zählen zu den wichtigen Bedingungen unserer Arbeit. Ihnen allen sei gedankt. Besonderer Dank gilt auch unseren zahlreichen Kooperationspartnerinnen und Kooperationspartnern in den Fachdidaktiken und in der empirischen Bildungsforschung andernorts für die gute Zusammenarbeit.

Am Ende sind es aber doch die Kompetenzen und das Engagement der eigenen Mitarbeiterinnen und Mitarbeiter – Wissenschaftler wie Nicht-Wissenschaftler –, die den Erfolg eines Instituts wie des IQB ausmachen. Wir als Leitung möchten ihnen an dieser Stelle ganz herzlich danken.

Berlin, im Sommer 2014

Prof. Dr. Hans Anand Pant
Direktor

Prof. Dr. Petra Stanat
Direktorin



01 // Das Institut zur Qualitätsentwicklung im Bildungswesen

Das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) ist eine wissenschaftliche Einrichtung, die die Länder in der Bundesrepublik Deutschland bei der Qualitätsentwicklung und Qualitätssicherung im allgemein bildenden Schulsystem unterstützt. Das Institut wird von den 16 Ländern finanziert, die im Vorstand (ab 1.1.2014 Kuratorium) des IQB durch die beiden Vorsitzenden der Amtschefskommission „Qualitätssicherung in Schulen“ und den Generalsekretär der Kultusministerkonferenz der Länder (KMK) vertreten sind. Weiterhin gehören dem Gremium drei renommierte Vertreter der Bildungsforschung an (Prof. Dr. Jürgen Baumert [seit Sommer 2014 Prof. Dr. Ewald Terhart], Prof. Dr. Detlev Leutner, Prof. Dr. Elmar Tenorth). Diese Zusammensetzung soll den Austausch zwischen Bildungspolitik und Bildungsforschung gewährleisten und sicherstellen, dass bei inhaltlichen und strukturellen Entscheidungen, die die Arbeit des IQB betreffen, beide Perspektiven berücksichtigt werden. Die wissenschaftliche Unabhängigkeit des IQB wird durch seinen Status als An-Institut der Humboldt-Universität zu Berlin gesetzlich garantiert. Das Institut hat im Dezember 2004 unter der Leitung von Prof. Dr. Olaf Köller seine Arbeit aufgenommen und wurde nach einer positiven Evaluation 2007 durch politischen Beschluss der KMK auf Dauer gestellt. Seit 2010 wird das IQB von Prof. Dr. Hans Anand Pant und Prof. Dr. Petra Stanat geleitet.

Auftrag und Ziele des Instituts

Der Ausgangspunkt und die Grundlage der Arbeit des IQB sind die Bildungsstandards, die von der KMK 2003/2004 verabschiedet worden sind. Sie definieren, welche Kompetenzen Schülerinnen und Schüler bis zu einem bestimmten Zeitpunkt in ihrer Schullaufbahn entwickelt haben sollen. Das IQB hat den Auftrag, die Bildungsstandards weiterzuentwickeln und regelmäßig zu überprüfen, inwieweit diese Ziele in deutschen Schulen erreicht werden (*Bildungsmonitoring*). Ferner unterstützt das IQB die Länder darin, die Bildungsstandards, die als zentraler Orientierungsrahmen für alle Akteure im Bildungssystem dienen sollen, umzusetzen (*Implementierung*). Darüber hinaus gehört das IQB zu den wichtigsten Instituten in Deutschland, die im Bereich der empirischen Bildungsforschung aktiv sind (*Forschung*).

Mit seiner Arbeit soll das IQB

- zur Verbesserung der schulischen Bildung in Deutschland beitragen,
- den Anschluss an das internationale Leistungsniveau befördern,
- die Bemühungen der Länder um eine höhere Qualität in Unterricht und Schule unterstützen,
- den länderübergreifenden Austausch über Maßnahmen der Qualitätsentwicklung und Qualitätssicherung fördern und
- zur Sicherung von Vergleichbarkeit und Durchlässigkeit im deutschen Bildungswesen beitragen.

Die Arbeit des IQB umfasst sechs Bereiche, die im Folgenden kurz skizziert werden. Sie werden in den weiteren Kapiteln des vorliegenden Tätigkeitsberichts ausführlicher dargestellt.

Bildungsstandards und ihre Operationalisierung durch Aufgaben

Das IQB präzisiert die Bildungsstandards der KMK inhaltlich und konkretisiert die Kompetenzziele anhand von Aufgaben. Unter der Federführung des IQB entwickeln Lehrkräfte *Testaufgaben*, mit denen die in den Bildungsstandards definierten Kompetenzen erfasst werden können. Diese werden aus fachdidaktischer Perspektive geprüft und empirisch erprobt. Weiterhin entwickelt das IQB fachdidaktisch und lernpsychologisch fundierte Kompetenzstufenmodelle, die zur inhaltli-

chen Interpretation von Testwerten dienen. Anhand der Modelle lässt sich beschreiben, welche Anforderungen Schülerinnen und Schüler bewältigen können, die ein bestimmtes Testergebnis erzielt haben. Die Kompetenzstufenmodelle des IQB sind so konstruiert, dass sie ein breites Leistungsspektrum auf wenigen Stufen beschreiben können, wobei jeweils eine Stufe dem von der KMK definierten Regelstandard entspricht. Ein Teil der Testaufgaben wird veröffentlicht, um die Kompetenzerwartungen der Bildungsstandards zu verdeutlichen und für alle Akteure im Bildungssystem transparent zu machen. Weiterhin entwickelt das IQB gemeinsam mit Kooperationspartnern aus den Fachdidaktiken *Lernaufgaben*, die zeigen sollen, welche Aufgabenstellungen im Unterricht eingesetzt werden können, um die angestrebten Kompetenzen zu entwickeln. Mit den Test- und Lernaufgaben, die im Rahmen von länderübergreifend einheitlichen Vergleichsarbeiten (VERA) und in Publikationen des IQB aus fachdidaktischer Sicht kommentiert werden, stellt das Institut den Ländern Material zur Verfügung, das zur Unterrichtsentwicklung genutzt werden kann.

Ländervergleichsstudien – Zentrale Überprüfung des Erreichens der Bildungsstandards

Das IQB prüft und berichtet im Rahmen seiner Ländervergleichsstudien regelmäßig, ob die Bildungsstandards in den Ländern erreicht werden. Ferner wird untersucht, inwieweit nicht nur die Heranwachsenden insgesamt, sondern auch verschiedene Schülergruppen (z. B. nach Geschlecht oder familiärem Hintergrund) die angezielten Kompetenzen entwickeln. Damit erhalten die Länder Rückmeldungen über Stärken und Schwächen ihrer schulischen Systeme, die sie zur Weiterentwicklung der Bildungsqualität nutzen können. Das IQB unterstützt die Länder bei der Interpretation der Ergebnismuster und beteiligt sich auf Wunsch auch an Diskussionen über mögliche Schlussfolgerungen. Weiterhin stellt das IQB mit seinen empirischen Studien Erkenntnisse zur Verfügung, die im Rahmen der nationalen Bildungsberichterstattung zur Beschreibung der Prozess- und Wirkungsqualität von Bildung beitragen können. Die Ländervergleichsstudien des IQB finden parallel zu den Erhebungen der großen internationalen Schulleistungsstudien (PIRLS/IGLU¹, TIMSS², PISA³) statt. In der Primarstufe werden sie seit 2011 alle fünf Jahre, in der Sekundarstufe I seit 2009 alle drei Jahre durchgeführt.

¹ Das Akronym PIRLS steht für *Progress in International Reading Literacy Study*; im deutschen Sprachraum wird diese Studie als Internationale Grundschul-Lese-Untersuchung (IGLU) bezeichnet.

² Das Akronym TIMSS steht für *Third International Mathematics and Science Study*. Seit 2003 wird es in der Bedeutung *Trends in International Mathematics and Science Study* verwendet.

³ Das Akronym PISA steht für *Programme for International Student Assessment*.

Vergleichsarbeiten

Für die IQB-Ländervergleichsstudien, die auf eine Evaluation von Bildungsqualität auf der Systemebene abzielen, werden die Daten an repräsentativen Stichproben erhoben. Im Unterschied dazu dienen Vergleichsarbeiten der landesweiten Untersuchung des Leistungsstands aller Schülerinnen und Schüler eines Jahrgangs. Die Länder führen die Vergleichsarbeiten jedes Jahr im Frühjahr durch. Das IQB koordiniert die Entwicklung und Erprobung von Testaufgaben sowie die Erstellung fachdidaktischer Materialien für die Vergleichsarbeiten in den Jahrgangsstufen 3 (Deutsch und Mathematik), 6 (Deutsch) und 8 (Deutsch, Mathematik sowie Englisch und Französisch als erste Fremdsprache). Mit den Ergebnissen der Vergleichsarbeiten erhalten Lehrkräfte auf die Bildungsstandards bezogene Informationen über die Stärken und Schwächen ihrer Klassen und Lerngruppen. Gemeinsam mit fachdidaktischen Hinweisen für die Weiterarbeit sollen diese Rückmeldungen die Schul- und Unterrichtsentwicklung der einzelnen Schulen unterstützen und zur Implementation der Bildungsstandards beitragen.

Pool mit Abiturprüfungsaufgaben

Seit Beginn des Schuljahres 2013/2014 koordiniert das IQB die Entwicklung eines Pools von Abiturprüfungsaufgaben, die auf den 2012 verabschiedeten Bildungsstandards der KMK für die Allgemeine Hochschulreife basieren. Auch die Entwicklung der Bildungsstandards für die Allgemeine Hochschulreife wurde vom IQB koordiniert. Mit der Einrichtung des Aufgabenpools wird das Ziel verfolgt, die Vergleichbarkeit der Anforderungen in den Abiturprüfungen der Länder zu gewährleisten sowie den Standardbezug und die Qualität der Aufgaben zu sichern. Der Aufgabenpool soll den Ländern ab dem Schuljahr 2016/2017 für den möglichen Einsatz im Abitur zur Verfügung stehen.

Forschung

Als wissenschaftliche Einrichtung führt das IQB zahlreiche Forschungsvorhaben durch. Diese betreffen zum einen Fragestellungen, die sich auf die Optimierung der Kernaufgaben des IQB im Bereich des Bildungsmonitorings beziehen. Zum anderen werden die Daten des IQB aber auch genutzt, um zu bestimmen, welche Faktoren den Lern- und Bildungserfolg von Kindern und Jugendlichen beeinflussen und wie Bildungsprozesse optimiert werden können. Im Einzelnen werden im IQB die folgenden Forschungsfragen bearbeitet:

- Welche Strukturen weisen schulbezogene Kompetenzen auf und wie lassen sich diese möglichst präzise messen?

- Wie lassen sich die vielfältigen methodischen Anforderungen von Schulleistungsmessungen und Schulleistungsvergleichen optimal bewältigen?
- Welche Faktoren beeinflussen den Lern- und Bildungserfolg Heranwachsender?
- Wie können Sprach- und Lesekompetenzen, die für schulisches Lernen erforderlich sind, erfasst und gezielt gefördert werden?
- Inwieweit und unter welchen Bedingungen gelingt es Lehrkräften, ihren Unterricht kompetenzorientiert zu gestalten?

Die Forschungsarbeiten des IQB tragen auch zur Qualifikation des wissenschaftlichen Nachwuchses und zum Aufbau von Expertise an der Schnittstelle zwischen Bildungsforschung, Bildungspolitik und Bildungspraxis bei.

Wissenschaftlicher Austausch

Das Forschungsdatenzentrum (FDZ) am IQB archiviert und dokumentiert die Datensätze der großen nationalen und internationalen Schulleistungsstudien, wie etwa den IQB-Ländervergleichsstudien und PISA, und stellt sie Wissenschaftlerinnen und Wissenschaftlern für Re- und Sekundäranalysen zur Verfügung. Darüber hinaus bietet das FDZ Veranstaltungen zur methodischen Weiterbildung des wissenschaftlichen Nachwuchses an und setzt sich für den Ausbau der Forschungsdateninfrastruktur im Bildungsbereich ein.

Ferner ist am IQB seit 2012 eine Koordinierungsstelle angesiedelt, die die Zusammenarbeit zwischen dem Zentrum für Internationale Bildungsvergleichsstudien (ZIB) und dem IQB unterstützen soll. Eine zentrale Aufgabe des 2010 gegründeten ZIB, bei dem es sich um einen Verbund der Technischen Universität München (TUM), des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) sowie des Leibniz-Instituts für die Pädagogik der Naturwissenschaften und Mathematik (IPN) handelt, ist die Durchführung von internationalen Bildungsvergleichsstudien, insbesondere von PISA. Gleichzeitig forscht das ZIB zu *Educational Measurement* und *International Large-Scale-Assessment* und soll die internationale Sichtbarkeit der deutschen Bildungsforschung in diesen Bereichen stärken. Die am IQB angesiedelte Koordinierungsstelle organisiert den Austausch zwischen IQB und ZIB, um die Aktivitäten der nationalen und internationalen *Large-Scale-Assessments* abzustimmen und gemeinsame Forschungsvorhaben durchzuführen. Weiterhin ist die Koordinierungsstelle an der inhaltlichen Ausgestaltung und Moderation von Tagungen und Workshops beteiligt, an denen Mitarbeiterinnen und Mitarbeiter des IQB und des ZIB teilnehmen.



Mitarbeiterinnen und Mitarbeiter Primarstufe und im Forschungsbereich Implementation 2013 (v.l.n.r.):

Maria Engelbert, Wissenschaftliche Mitarbeiterin Deutsch
 Dr. Heino Reimers, Koordinator Deutsch und Mathematik
 Claudia Rienäcker, Wissenschaftliche Mitarbeiterin Mathematik (bis 31.08.2013)
 Sebastian Weirich, Wissenschaftlicher Mitarbeiter Deutsch
 Nicole Haag, Wissenschaftliche Mitarbeiterin Mathematik
 Alexander Roppelt, Wissenschaftlicher Mitarbeiter Mathematik
 Dr. Katrin Böhme, Wissenschaftliche Mitarbeiterin Deutsch
 Dr. Dirk Richter, Wissenschaftlicher Mitarbeiter im Forschungsbereich Implementation
 Josefine Prengel, Wissenschaftliche Mitarbeiterin Deutsch (bis 31.08.2013)



Mitarbeiterinnen und Mitarbeiter Mathematik Sekundarstufe I 2013 (v.l.n.r.):

Dr. Claudia Pöhlmann, Wissenschaftliche Mitarbeiterin (bis 31.12.2013)
 Christiane Penk, Wissenschaftliche Mitarbeiterin
 Alexander Roppelt, Wissenschaftlicher Mitarbeiter
 Elke Pietsch, Koordinatorin (ohne Foto, 01.02.2011 bis 31.01.2013)
 Michael Katzenbach, Koordinator (ohne Foto, 01.01.2007 bis 31.07.2010, seit 01.02.2013)

Ausstattung und Personalstruktur des Instituts

In den letzten Jahren hat sich das IQB kontinuierlich weiterentwickelt und neue Aufgaben übernommen, was eine Erhöhung der Zuwendung durch die Länder erforderte. Der Wirtschaftsplan für die Jahre 2012/2013 sah Ausgaben im Umfang von insgesamt 4,24 Mio. Euro jährlich vor. Hinzu kamen zeitlich begrenzte projektbezogene Mittel für die Entwicklung von Bildungsstandards und Aufgaben für die Allgemeine Hochschulreife, gemeinsame Mittel des Bundes und der Länder für das Forschungsdatenzentrum und die Koordinierungsstelle zwischen dem ZIB und dem IQB sowie Mittel der Länder Hessen, Mecklenburg-Vorpommern, Thüringen, Sachsen und Hamburg für die Aufgabenentwicklung und Testhefterstellung im Fach Deutsch im Rahmen von VERA 6.

Der Stellenplan sah für die Haushaltsjahre 2012/2013 zwei Leitungsprofessuren (W3), dreizehn Wissenschaftlerstellen sowie sieben nicht wissenschaftliche Stellen in den Bereichen Verwaltung und EDV vor. Hinzu kommen Stellen in projektbezogenen Arbeitsbereichen wie dem Forschungsdatenzentrum, der Koordinierungsstelle zwischen IQB und ZIB und die Erarbeitung von Bildungsstandards für die Allgemeine Hochschulreife.

Die fachbezogene Koordination der Aufgabenentwicklung erfolgt in der Regel durch Lehrkräfte mit einer Lehrbefähigung im entsprechenden Fach, die von den Ländern an das IQB abgeordnet werden. Das IQB erstattet den Ländern die Gehaltskosten der Koordinatorinnen und Koordinatoren; die Versorgungsrücklagen und weitere Kosten wurden hingegen bislang in der Regel von den abordnenden Ländern getragen. Die wissenschaftlichen Mitarbeiterinnen und Mitarbeiter sind für die wissenschaftlichen Anteile der Serviceaufgaben des IQB zuständig und sichern deren Qualität. Ihre Stellen sind zeitlich befristet und dienen auch der Weiterqualifikation im Sinne der Hochschulrahmengesetzgebung.

Im Haushaltsjahr 2013 waren insgesamt 55 Mitarbeiterinnen und Mitarbeiter sowie 54 studentische Hilfskräfte an der Humboldt-Universität zu Berlin für das IQB beschäftigt.



Mitarbeiterinnen und Mitarbeiter Deutsch Sekundarstufe I 2013 (v.l.n.r.):

Stefan Schipolowski, Wissenschaftlicher Mitarbeiter
Dr. Katrin Böhme, Wissenschaftliche Mitarbeiterin
Susanne Hunger, Koordinatorin
Thomas Canz, Wissenschaftlicher Mitarbeiter



Mitarbeiterinnen Fremdsprachen Sekundarstufe I 2013 (v.l.n.r.):

Bettina Neugebauer, Koordinatorin Französisch (bis 31.10.2013)
Amira Yassine, Koordinatorin Englisch
Jessica Maluch, Wissenschaftliche Mitarbeiterin
Marlen Seppelt, Wissenschaftliche Mitarbeiterin
Karoline Sachse, Wissenschaftliche Mitarbeiterin
Daniela Holm, Koordinatorin Französisch (ohne Foto, seit 01.11.2013)

Entwicklung des Instituts im Berichtszeitraum

Als besonders aufwändiges zusätzliches Arbeitsfeld im Berichtszeitraum 2007 bis 2013 übernahm das IQB 2008 bzw. 2009 die Zuständigkeit dafür, Aufgaben für die Vergleichsarbeiten in der 3. und in der 8. Jahrgangsstufe (VERA 3 und VERA 8) zu entwickeln und empirisch zu erproben sowie begleitende didaktische Materialien für die Unterrichtspraxis zur Verfügung zu stellen. Seit 2011 koordiniert das IQB zudem die Entwicklung von Aufgaben für VERA 6 im Fach Deutsch. Die Arbeitsprozesse in den VERA-Projekten, an denen eine große Anzahl von Akteuren beteiligt ist und die in jedem Jahr durchlaufen werden, sind äußerst komplex und zeitkritisch. Die Ergebnisse dieser Prozesse kommen bundesweit in so gut wie allen 3. und 8. Klassen zum Einsatz, d. h. insgesamt circa 1,5 Millionen Schülerinnen und Schüler bearbeiten die Tests und etwa 50000 Lehrkräfte werten die Ergebnisse aus. Es ist daher äußerst wichtig, die Qualität der VERA-Aufgaben zu sichern. Entsprechend lag ein Hauptaugenmerk des Instituts in den vergangenen Jahren darauf, die Arbeitsabläufe in den VERA-Projekten kontinuierlich weiterzuentwickeln und zu optimieren.

Für die Primarstufe und die Sekundarstufe I lagen die Bildungsstandards der KMK zum Zeitpunkt der Gründung des IQB bereits vor, für die Allgemeine Hochschulreife hingegen wurde die Entwicklung der Bildungsstandards vom IQB koordiniert. Dieser Prozess begann im Dezember 2009 und wurde von einer Steuerungsgruppe begleitet, die sich aus Vertreterinnen und Vertretern der Bildungsadministration ausgewählter Länder zusammensetzte und von der Vorsitzenden des Schulausschusses der KMK geleitet wurde. Die Erarbeitung der Bildungsstandards sowie illustrierender Lern- und Abiturprüfungsaufgaben in den Fächern Deutsch, fortgeführte Fremdsprache (Englisch und Französisch) sowie Mathematik erfolgte in Fachgruppen bestehend aus Mitarbeiterinnen und Mitarbeitern der Bildungsadministrations verschiedener Länder, Lehrkräften, Professorinnen und Professoren in der jeweiligen Fachdidaktik sowie Mitarbeiterinnen und Mitarbeitern des IQB. Im Oktober 2012 verabschiedete die KMK die Bildungsstandards für die Allgemeine Hochschulreife, womit sich die Länder gleichzeitig verpflichteten, sie bis zum Schuljahr 2016/2017 verbindlich einzuführen. Wann mit der Erarbeitung von Bildungsstandards für die Allgemeine Hochschulreife in den naturwissenschaftlichen Fächern begonnen werden soll, ist noch offen.



Mitarbeiterinnen und Mitarbeiter Naturwissenschaften Sekundarstufe I 2013 (v.l.n.r.):

Malte Jansen, Wissenschaftlicher Mitarbeiter
 Anne Ziemke, Wissenschaftliche Mitarbeiterin
 Dr. Nele Kampa, Wissenschaftliche Mitarbeiterin
 Martin Hecht, Wissenschaftlicher Mitarbeiter
 Nicola Klebba, Koordinatorin (bis 31.10.2012)
 Dr. Ulrich Schroeders, Wissenschaftlicher Mitarbeiter
 Dr. Patricia Heitmann, Koordinatorin (seit 01.11.2012)
 Anna Eva Lenski, Wissenschaftliche Mitarbeiterin
 Thilo Siegle, Wissenschaftlicher Mitarbeiter

Mit einem Beschluss der KMK von 2007 hatte das IQB zunächst den Auftrag erhalten, im Anschluss an die Entwicklung von Bildungsstandards für die Allgemeine Hochschulreife auch für die Sekundarstufe II Testaufgaben und Kompetenzstufenmodelle zu entwickeln. Dieser Beschluss wurde jedoch im Sommer 2011 durch den Auftrag ersetzt, einen Pool von Abiturprüfungsaufgaben zu erstellen, der den Ländern ab dem Schuljahr 2016/2017 für den möglichen Einsatz im Abitur zur Verfügung stehen soll. Die Arbeiten daran hat das IQB im Herbst 2013 aufgenommen, sie befinden sich momentan also noch in der Anfangsphase.

Das 2007 am IQB eingerichtete Forschungsdatenzentrum (FDZ) hat sich im Berichtszeitraum als eine anerkannte Institution der deutschen Bildungsforschung und der informationellen Forschungsdateninfrastruktur etabliert. Es wurde 2009 von einer wissenschaftlichen Gutachterkommission evaluiert, die empfahl, das FDZ zu verstetigen. Nachdem das FDZ zunächst fast vollständig vom Bundesministerium für Bildung und Forschung (BMBF) finanziert worden war, ist es inzwischen an die von Bund und Ländern kofinanzierte Stelle zur Koordinierung der Aktivitäten des IQB und des Zentrums für Internationale Vergleichsstudien (ZIB) angekoppelt, die 2012 am IQB eingerichtet wurde.

Arbeitsbereiche

Die Arbeitsbereiche am IQB sind weitgehend anhand einer Schulstufen- (Primarstufe, Sekundarstufe I, Sekundarstufe II) und Fächersystematik (Deutsch, Englisch, Französisch, Mathematik, Naturwissenschaften) organisiert. Die einzelnen Arbeitsgruppen sind dabei für ihr jeweiliges Fach in der jeweiligen Schulstufe sowohl für die Arbeiten zuständig, die in Bezug auf die Bildungsstandards und die Ländervergleichsstudien anfallen, als auch für die Aufgabenentwicklung in VERA (Ausnahmen bilden die naturwissenschaftlichen Fächer und die Sekundarstufe II, für die keine Vergleichsarbeiten vorgesehen sind). Eine Herausforderung besteht bei dieser Struktur darin, die Aktivitäten fächer- und schulstufenübergreifend abzustimmen. Für die VERA-Projekte ist hierfür eine wissenschaftliche Mitarbeiterin zuständig, die auch als Ansprechpartnerin für die Länder zur Verfügung steht.

Die erforderliche methodische Abstimmung zwischen den Projekten, z. B. hinsichtlich der Skalierung von Testdaten oder der Verankerung von Testwerten auf der Metrik der Bildungsstandards, erfolgt weitgehend durch Arbeitsgruppen, die von befristet beschäftigten wissenschaftlichen Mitarbeiterinnen und Mitarbeitern organisiert werden. Aufgrund der Expertise, die sich diese Mitarbeiterinnen und Mitarbeiter



**Mitarbeiterinnen und Mitarbeiter Sekundarstufe II 2013
(von links oben im Uhrzeigersinn):**

Dr. Bernd Tesch, Koordinator Fremdsprachen (bis 31.09.2013)
Camilla Rjosk, Wissenschaftliche Mitarbeiterin Englisch
Gabriele Gippner, Koordinatorin Deutsch (bis 31.01.2013)
Lars Hoffmann, Wissenschaftlicher Mitarbeiter Deutsch
Maike Wäckerle, Wissenschaftliche Mitarbeiterin Französisch
Xenia von Hammerstein, Koordinatorin Englisch (ohne Foto, seit 18.02.2013)
Heike Wirthwein, Koordinatorin Deutsch (ohne Foto, seit 15.10.2013)



**Mitarbeiterinnen und Mitarbeiter des FDZ und der Koordinierungsstelle
am IQB 2013**

(von links oben im Uhrzeigersinn):

Dr. Poldi Kuhl, FDZ-Leitung
Monika Lacher, Assistentin der FDZ-Leitung
Dr. Simon-Tiffin Richards, Koordinierungsstelle
zwischen IQB und ZIB (bis 01.10.2012)
Anna Kroth, Koordinierungsstelle zwischen IQB und ZIB
Aleksander Kocaj, Wissenschaftlicher Mitarbeiter FDZ

angeeignet haben, ist es bislang gelungen, die erheblichen methodischen Herausforderungen zu bewältigen, die mit den am IQB entwickelten Testdesigns verbunden sind. Um die Verfahren auf der Grundlage des neuesten Forschungsstands systematisch weiterzuentwickeln sowie ihre kontinuierliche und durchgängige Umsetzung in den Projekten sicherzustellen, ist mittelfristig der Aufbau einer eigenen und exzellent geleiteten Methodenabteilung unabdingbar.

Weitere Arbeitsbereiche am IQB sind das Forschungsdatenzentrum und die Koordinierungsstelle zwischen IQB und ZIB sowie die zentralen Dienste EDV und Verwaltung.

Zur Koordination der Forschungs- und Publikationsaktivitäten in den verschiedenen Arbeitsbereichen des IQB wurde 2010 eine Arbeitsgruppe eingerichtet, in der die Postdoktorandinnen und Postdoktoranden am IQB die systematische und kontinuierliche Sichtbarkeit des Instituts in der deutschen und internationalen *Scientific Community* sichern. Weiterhin übernehmen sie Betreuungsaufgaben für die Promovierenden, entwickeln gemeinsam mit der wissenschaftlichen Leitung des IQB Forschungsstrategien und gestalten das Forschungskolloquium des Instituts. Im Berichtszeitraum haben 11 Mitarbeiterinnen und Mitarbeiter des Instituts ihre Promotionen erfolgreich abgeschlossen.

Drittmittelprojekte in der Bildungsforschung

Die Finanzierung des IQB durch die Länder ist für Forschungs- und Entwicklungsaufgaben vorgesehen, die sich auf die Präzisierung und Weiterentwicklung von Bildungsstandards sowie die standardbasierte Kompetenzerfassung beziehen. Weiterhin werden im Rahmen dieser Finanzierung vertiefende Analysen der IQB-Daten durchgeführt, die sich mit Determinanten des Bildungserfolgs von Kindern und Jugendlichen beschäftigen. Darüber hinaus haben die wissenschaftlichen Leiter des IQB eine größere Anzahl von Projekten im Bereich der empirischen Bildungsforschung eingeworben, die in enger Kooperation mit der Humboldt-Universität durchgeführt werden. Insgesamt waren 2013 in den Drittmittelprojekten 15 wissenschaftliche Mitarbeiterinnen und Mitarbeiter an der Humboldt-Universität beschäftigt, die überwiegend vom Bundesministerium für Bildung und Forschung (BMBF) oder von der Deutschen Forschungsgemeinschaft (DFG) finanziert wurden. Mit diesen Projekten arrondiert das IQB sein Forschungsprofil und stärkt die wissenschaftliche Sichtbarkeit des Instituts.



Mitarbeiterinnen und Mitarbeiter in Drittmittelprojekten und auf Stellen der Humboldt-Universität (HU) 2013:

Obere Reihe v. l. n. r.: Dr. Sebastian Kempert (Phono und HU), Annkathrin Darsow (BeFo), Aileen Edele (NEPS-SPP), Katrin Wolf (HU), Doreen Schöppe (HU), Tatjana Taraszow (NEPS), Sarah Gentrup (KuL), Stefan Hartmann (Ko-WADiS)

Untere Reihe v. l. n. r.: Jennifer Paetsch (BeFo), Kristin Schotte (NEPS), Birgit Heppt (BiSpra), Alexandra Marx (HU), Dr. Anja Felbrich (BeFo und HU), Sebastian Wurster (StABil), Anna Schliesing (StABil), Dr. Luna Beck (ohne Foto, BiSS, seit 15.08.2013 am IQB), Dr. Sofie Henschel (ohne Foto, BiSS, seit 01.10.2013 am IQB), Charlotte Kohrs (ohne Foto, BiSS, seit 18.11.2013 am IQB)



Mitarbeiterin und Mitarbeiter EDV 2013:

Daniela Holm (bis 31.10.2013)
Martin Mechtel, EDV-Leitung
Ralf Schultzik
Hannes Schurig (bis 31.03.2013)
Frank Zahnnow

Einbindung in die Humboldt-Universität

Das IQB ist seit 2004 im Sinne von § 85 des Berliner Hochschulgesetzes ein An-Institut der Humboldt-Universität (HU), das mit der Philosophischen Fakultät IV (seit April 2014 mit der neu gebildeten Kultur-, Sozial- und Bildungswissenschaftlichen Fakultät) assoziiert ist. Die beiden Direktoren sind Professoren am Institut für Erziehungswissenschaften und haben für einen Zeitraum von zunächst fünf Jahren die Dienstaufgabe übernommen, das IQB zu leiten. In der Lehre übernehmen sie regelmäßig Veranstaltungen in den Bereichen Diagnostik, Rückmeldung und Evaluation sowie Testtheorie und Kausale Modellierung in der Schulleistungsforschung. Die Leitung des IQB ist zudem Mitglied im Interdisziplinären Zentrum für Bildungsforschung (IZBF) der Humboldt-Universität. Weiterhin kooperiert das Institut mit der Humboldt-Universität im Rahmen von Drittmittelprojekten, die von der Leitung des IQB eingeworben wurden und formal an der HU angesiedelt sind.

Kooperationspartner

Das IQB kooperiert mit einer Vielzahl von renommierten Universitäten und Forschungseinrichtungen im In- und Ausland. Die Kooperationen beziehen sich auf die Serviceaufgaben des IQB, insbesondere im Bereich der Aufgabenentwicklung, auf die vielfältigen Forschungs- und Entwicklungsaktivitäten des Instituts und auf die Qualifizierung des wissenschaftlichen Nachwuchses. Das Forschungsdatenzentrum am IQB verfügt über weitere Kooperationsbeziehungen, die sich auf die spezifischen Aufgaben dieses Arbeitsbereichs beziehen. Zudem sind an den Drittmittelprojekten der IQB-Leitung, die in Kooperation mit der Humboldt-Universität durchgeführt werden, in der Regel auch Kolleginnen und Kollegen anderer Universitäten bzw. Forschungseinrichtungen beteiligt.

Kooperationen im Bereich der Aufgabenentwicklung

Die fachbezogene Koordination der Aufgabenentwicklung in den Projekten erfolgt am IQB durch abgeordnete Lehrkräfte und wissenschaftliche Mitarbeiterinnen und Mitarbeiter. In der Gründungsphase des Instituts wurden Kooperationen mit zahlreichen fachdidaktischen Lehrstühlen in den jeweiligen Fächern angebahnt, die im Berichtszeitraum weiter aus-



Mitarbeiterinnen Verwaltung 2013 (v.l.n.r.):

Monika Lacher (Assistenz FDZ-Leitung)
 Martina Rosin (Reisekosten / Assistenz Geschäftsführung, bis 31.01.2013)
 Mandy Ladenthin (Vertragswesen/Reisekosten)
 Anne Dressler (Verwaltungsleitung, bis 30.09.2013)
 Mandy Sulejmani (stellv. Verwaltungsleitung / Buchhaltung)
 Ricarda Klein (Assistenz Institutsleitung)
 Kerstin Dolch (Leitung Veranstaltungsmanagement)
 ohne Foto:
 Kerstin Dröbig (Verwaltungsleitung/Personalwesen, seit 01.09.2013)
 Fränze Schröder (Haushalts- und Personalwesen, bis 31.01.2013)
 Anica Thiele (Reisekosten / Assistenz Geschäftsführung)



**Mitarbeiterinnen VERA 2013
 (v.l.n.r.):**

Bettina Neugebauer,
 Koordinatorin VERA 3 und 8
 (seit 01.11.2013)
 Xenia Steigerwald,
 Koordinatorin VERA 6
 Dr. Anke Greve,
 Koordinatorin VERA 3 und 8
 (ohne Foto, bis 31.10.2013)



Geschäftsführer 2013:
 Kirk Fänderich

gebaut werden konnten. Auf den folgenden Seiten werden die fachdidaktischen Kooperationspartnerinnen und -partner des IQB sowie die Aufgabenbewerterinnen und -bewerter aufgeführt. Weiterhin arbeitet das Institut mit zahlreichen Lehrkräften aus den Ländern zusammen, die dafür zuständig sind, die Aufgaben zu entwickeln. Ihr Engagement bildet das Fundament der Testentwicklung am IQB.

Kooperationen im Bereich der Forschung und Entwicklung

Als wissenschaftliche Einrichtung ist das IQB mit seinen Arbeiten sowohl national als auch international sichtbar und anschlussfähig. Dies gelingt durch intensiven Austausch mit nationalen und internationalen Kooperationspartnern in formellen und informellen Netzwerken. Die vergangenen Jahre wurden genutzt, um das vorhandene Netzwerk weiter auszubauen. Neben den bereits genannten Kooperationen arbeitet das IQB mit den auf den folgenden Seiten aufgeführten Institutionen und Personen eng zusammen.

Kooperationen im Bereich der Qualifikation von wissenschaftlichem Nachwuchs

Die Anforderungen, die an wissenschaftlichen Nachwuchs etwa in Bezug auf Methodenkenntnisse, Publikationsleistungen und Drittmittelinwerbung gestellt werden, sind deutlich gestiegen. Entsprechend hat die Anzahl der Angebote zugenommen, die darauf abzielen, Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftler in diesen Bereichen zu qualifizieren. Auch in der empirischen Bildungsforschung stehen solche Angebote zur Verfügung, die teilweise in Berlin angesiedelt sind. An den Programmen, die auf den folgenden Seiten aufgeführt sind, ist das IQB aktiv beteiligt.

Primarstufe

Mathematik

Fachdidaktische Kooperationspartnerinnen und -partner:

Gabriele Moll
Technische Universität München
Prof. Dr. Kristina Reiss
Technische Universität München

Fachdidaktische Bewerberinnen und Bewerber:

Prof. Dr. Hedwig Gasteiger
Ludwig-Maximilians-Universität München
Prof. Dr. Marja van den Heuvel-Panhuizen
Universität Utrecht
Prof. Dr. Ingmar Hosenfeld
Universität Koblenz-Landau
Prof. Dr. Mareike Kunter
Goethe-Universität Frankfurt
Prof. Dr. Eva-Maria Lankes
Technische Universität München, School of Education
Prof. Dr. Jens Holger Lorenz
Goethe-Universität Frankfurt
Dr. Johanna Neubrand
Christian-Albrechts-Universität zu Kiel
Dr. Paul Olbrich
Staatliches Schulamt Dillingen
Prof. Dr. Andrea Peter-Koop
Universität Oldenburg
Prof. Dr. Bernd Wollring
Universität Kassel
Prof. Dr. Gerd Walther
Christian-Albrechts-Universität zu Kiel

Deutsch

Fachdidaktische Kooperationspartnerinnen und -partner:

Dr. Ulrike Behrens
Universität Duisburg-Essen
Prof. Dr. Albert Bremerich-Vos
Universität Duisburg-Essen
Michael Krelle
Universität Duisburg-Essen
Dr. Christopher Sappok
Leuphana Universität Lüneburg

Fachdidaktische Bewerberinnen und Bewerber:

Prof. Dr. Wolfgang Eichler
Universität Oldenburg
Prof. Dr. Peter Conrady
Universität Dortmund
Prof. Dr. Ingmar Hosenfeld
Universität Koblenz-Landau
Prof. Dr. Eva-Maria Lankes
Technische Universität München, School of Education
Prof. Dr. Thomas Lindauer
Pädagogische Hochschule Aargau Nordwestschweiz
Prof. Dr. Ingelore Oomen-Welke
Pädagogische Hochschule Freiburg
Prof. Dr. Rainer Peek (†)
Universität zu Köln
Prof. Dr. Angelika Speck-Hamdan
Ludwig-Maximilians-Universität München
Prof. Dr. Annegret v. Wedel-Wolff
Pädagogische Hochschule Schwäbisch Gmünd

Sekundarstufe I

Mathematik

Fachdidaktische Kooperationspartnerinnen und -partner an der Universität Kassel:

Kay Achmetli
Dr. Michael Besser
Prof. Dr. Werner Blum
Christina Drüke-Noe
Dr. Alexander Jordan (†)
Katrin Keller
Janina Krawitz
André Krug
Universität Münster
Prof. Dr. Dominik Leiß
Leuphana Universität Lüneburg
Marcel Müller
Sebastian Vogel
Dr. Bernd Wiegand

Fachdidaktische Bewerberinnen und Bewerber:

Prof. Dr. Angelika Bikner-Ahsbahr
Universität Bremen
Prof. Dr. Werner Blum
Universität Kassel
Prof. Dr. Andreas Büchter
Universität zu Köln
Prof. Dr. Gilbert Greefrath
Universität Münster
Prof. Dr. Aiso Heinze
Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik
Prof. Dr. Wilfried Herget
Universität Halle-Wittenberg
Nikola Leufer
Landesinstitut für Schule, Bremen
Prof. Dr. Michael Neubrand
Universität Oldenburg
Prof. Dr. Alexander Wynands
Universität Bonn

Deutsch

Fachdidaktische Kooperationspartnerinnen und -partner:

Prof. Dr. Michael Becker-Mrotzek
Universität zu Köln
Dr. Ulrike Behrens
Universität Duisburg-Essen
Prof. Dr. Albert Bremerich-Vos
Universität Duisburg-Essen
Necle Bulut
Universität zu Köln
Dr. Jörg Jost
Universität Paderborn
Michael Krelle
Universität Duisburg-Essen
Michaela Mörs
Universität zu Köln
Dr. Uwe Neugebauer
Universität zu Köln
Miriam Possmayer
Universität zu Köln
Dr. Christopher Sappok
Leuphana Universität Lüneburg

Fachdidaktische Bewerberinnen und Bewerber:

Prof. Dr. Anne Berkemeier
Pädagogische Hochschule Heidelberg
Prof. Dr. Albert Bremerich-Vos
Universität Duisburg-Essen
Prof. Dr. Wolfgang Eichler
Universität Oldenburg
Prof. Dr. Volker Frederking
Universität Erlangen-Nürnberg
Prof. Dr. Hartmut Günther
Universität zu Köln
Prof. Dr. Rüdiger Grotjahn
Ruhr-Universität Bochum
Prof. Dr. Johannes Hartig
Deutsches Institut für Internationale Pädagogische Forschung
Prof. Dr. Michael Kämper-van den Boogaart
Humboldt-Universität zu Berlin
Prof. Dr. Carl Ludwig Naumann
Leibniz Universität Hannover
Prof. Dr. Astrid Neumann
Leuphana Universität Lüneburg
Prof. Dr. Rainer Peek (t)
Universität zu Köln
Prof. Dr. Cornelia Rosebrock
Universität Frankfurt am Main
Prof. Dr. Petra Stanat
Humboldt-Universität zu Berlin
Prof. Dr. Swantje Weinhold
Leuphana Universität Lüneburg
Prof. Dr. Oomen-Welke
Pädagogische Hochschule Freiburg

Französisch

Fachdidaktische Kooperationspartnerinnen und -partner:

Prof. Dr. Daniela Caspari
Freie Universität Berlin
Prof. Dr. Adelheid Hu
Universität Luxemburg
Almut Keller
Freie Universität Berlin
Prof. Dr. Eynar Leupold
Pädagogische Hochschule Freiburg
Prof. Dr. Franz-Joseph Meißner
Universität Gießen
Dr. Elke Philipp
abgeordnete Lehrkraft Land Berlin
Kerstin Rauch
abgeordnete Lehrkraft Land Berlin

Fachdidaktische Bewerberinnen und Bewerber:

Prof. Dr. Rüdiger Grotjahn
Ruhr-Universität Bochum
Anette Kamitz
Ruhr-Universität Bochum
Prof. Dr. Karin Kleppin
Ruhr-Universität Bochum
Dr. Astrid Reich
Ruhr-Universität Bochum
Heidi Schmitt-Ford
Lehrkraft Rheinland-Pfalz
Dr. Hanno Werry
Landesinstitut für Pädagogik und Medien, Saarland

Testmethodische Partnerinnen und Partner am Centre international d'études pédagogiques (CIEP), Sèvres:

Sylvie Lepage
Roselyne Marty
Patrick Riba

Englisch

Fachdidaktische Kooperationspartnerinnen und -partner:

Prof. Dr. Petra Burmeister
Pädagogische Hochschule Weingarten
Dr. Rita Green
Test Development Training and Analysis Ltd., Old Stratford
Prof. Dr. Günther Nold
Universität Dortmund
Prof. Dr. Konrad Schröder
Universität Augsburg
Prof. Dr. Wolfgang Zydati
Freie Universität Berlin

Fachdidaktische Bewerberinnen und Bewerber:

Prof. Dr. Petra Burmeister
Pädagogische Hochschule Weingarten
Dr. Rita Green
Test Development Training and Analysis Ltd., Old Stratford

Weitere Beraterinnen und Berater der Aufgabenentwicklergruppe:

Graham Hyatt
Ministerium für Bildung, Wissenschaft und Kultur
Mecklenburg-Vorpommern
Hella Tinis-Faur
Staatsinstitut für Schulqualität und Bildungsforschung
München

Biologie, Chemie, Physik

Fachdidaktische Partnerinnen und Partner im Fach Biologie:

Julia Arnold
Universität Kassel
Dr. Stefan Hartmann
Humboldt-Universität zu Berlin
Kathrin Klöpfel
Universität Kassel
Dr. Kerstin Kremer
Universität Kassel
Prof. Dr. Jürgen Mayer
Universität Kassel
Prof. Dr. Julia Schwanewedel
Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik
Dr. Nicole Wellnitz
Universität Kassel
Melanie Werner
Universität Kassel

Fachdidaktische Partnerinnen und Partner im Fach Chemie:

Dr. Julia Hostenbach
Universität Duisburg-Essen
Iwen Kobow
Universität Duisburg-Essen
Prof. Dr. Mathias Ropohl
Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik
Prof. Dr. Elke Sumfleth
Universität Duisburg-Essen
Prof. Dr. Maik Walpuski
Universität Duisburg-Essen

Fachdidaktische Partnerinnen und Partner im Fach Physik:

Sara Dehn
Universität Duisburg-Essen
Prof. Dr. Hans E. Fischer
Universität Duisburg-Essen
Prof. Dr. Hendrik Härtig
Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik
Prof. Dr. Alexander Kauertz
Universität Koblenz-Landau
Dr. Irene Neumann
Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik
Raffaella Römer
Universität Duisburg-Essen
Melanie Wächter
Pädagogische Hochschule Weingarten

Sekundarstufe II

Mathematik

Beteiligte an der Standardentwicklung:

Thomas Jurke
Kultusministerium Stuttgart
Manfred Herbst
Gymnasium Herzogenaurach
Christian Bänsch
Senatsverwaltung Berlin
Peter Lorenz
Mecklenburg-Vorpommern
Hans-Dieter Stenten-Langenbach
Niedersachsen
Dr. Rainer Heinrich
Kultusministerium Sachsen
Ardito Messner
Sachsen-Anhalt
Dr. Dieter Bartneck
Kultusministerium Sachsen
Prof. Dr. Werner Blum
Universität Kassel
Prof. Dr. Regina Bruder
Technische Universität Darmstadt
Prof. Dr. Michael Kleine
Universität Bielefeld

Wissenschaftliche Beratung der Aufgabenentwicklergruppe:

Prof Dr. Werner Blum
Universität Kassel
Prof. Dr. Regina Bruder
Technische Universität Darmstadt
Prof Dr. Michael Kleine
Universität Bielefeld
Prof. Dr. Michael Neubrand
Universität Oldenburg
Prof. Dr. Alexander Wynands
Universität Bonn
Prof Dr. Gabriele Kaiser
Universität Hamburg

Deutsch

Beteiligte an der Standardentwicklung:

Prof. Dr. Michael Becker-Mrotzek
Universität zu Köln
Dr. Désirée Burba
MBK Schleswig-Holstein
Antonia Dicken-Begrich
Bezirksregierung Düsseldorf
Robert Grüschow
Senatorin für Bildung und Wissenschaft, Bremen
Dr. Jörg Heinig
Ministerium für Bildung, Wissenschaft und Kultur, Mecklenburg-Vorpommern
Prof. Dr. Michael Kämper-van den Boogaart
Humboldt-Universität zu Berlin
Birgit Kölle
Senatsverwaltung Berlin
Prof. Dr. Juliane Köster
Friedrich-Schiller-Universität Jena
Dr. Elke Richlick
Niedersächsische Landesschulbehörde
Jürgen Rotschedl
Staatsinstitut für Schulqualität und Bildungsforschung, Bayern
Dr. Oliver Schoell
Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen

Wissenschaftliche Beratung der Aufgabenentwicklergruppe:

Prof. Dr. Michael Becker-Mrotzek
Universität zu Köln
Prof. Dr. Helmuth Feilke
Universität Gießen
Prof. Dr. Volker Frederking
Universität Erlangen-Nürnberg
Dr. Jörg Jost
Universität Paderborn
Prof. Dr. Juliane Köster
Universität Jena
Michaela Mörs
Universität zu Köln
Prof. Dr. Astrid Neumann
Leuphana Universität Lüneburg
Prof. Dr. Torsten Steinhoff
Technische Universität Dortmund
Dr. Dorothee Wieser
Humboldt-Universität zu Berlin
Prof. Dr. Iris Winkler
Universität Jena

Englisch

Beteiligte an der Standardentwicklung:

Dr. Christof Arnold
*Ministerium für Bildung, Wissenschaft und Kultur
 Mecklenburg-Vorpommern*
 Uwe Christiansen
Ministerium für Bildung und Kultur Schleswig-Holstein
 Dr. Rupert Deppe
Thüringisches Ministerium für Bildung, Wissenschaft und Kultur
 Prof. Dr. Günter Nold
Technische Universität Dortmund
 Dr. Manfred Pulm
Bezirksregierung Köln
 Prof. Dr. Konrad Schröder
Universität Augsburg
 Ulrike Schuh-Fricke
Neue Oberschule Braunschweig
 Günther Sommerschuh
Institut für Qualitätsentwicklung an Schulen Schleswig Holstein
 Dr. Rolf Theis
Goethe-Gymnasium, Frankfurt am Main
 Manfred Thönicke
Hamburger Institut für Berufliche Bildung
 Prof. Dr. Helmut Johannes Vollmer
Universität Osnabrück

Wissenschaftliche Beratung der Aufgabenentwicklergruppe:

Prof. Dr. Eva Burwitz-Melzer
Universität Gießen
 Prof. Dr. Günter Nold
Technische Universität Dortmund
 Prof. Dr. Konrad Schröder
Universität Augsburg
 Prof. Dr. Helmut Johannes Vollmer
Universität Osnabrück

Französisch

Beteiligte an der Standardentwicklung:

Dorothee Bauni
*Ministerium für Bildung, Wissenschaft, Jugend und Kultur
 des Landes Rheinland-Pfalz*
 Prof. Dr. Daniela Caspari
Freie Universität Berlin
 Annette de la Motte
Ministerium für Bildung und Kultur Schleswig-Holstein
 Prof. Dr. Eynar Leupold
Pädagogische Hochschule Freiburg
 Prof. Dr. Franz-Joseph Meißner
Universität Gießen
 Joachim Mohr
Ministerium für Bildung, Saarbrücken
 Henny Rönneper
Ministerium für Schule, Jugend und Kultur NRW
 Véronique Zettl
Kaufmännische Schule Schwäbisch Gmünd
 Anneliese Zorn
Johannes-Kepler-Gymnasium, Weil der Stadt

Wissenschaftliche Beratung der Aufgabenentwickler:

Prof. Dr. Daniela Caspari
Freie Universität Berlin
 Prof. Dr. Karin Kleppin
Ruhr-Universität Bochum
 Prof. Dr. Hélène Martinez
Universität Gießen
 Prof. Dr. Franz-Joseph Meißner
Universität Gießen

Kooperationen im Bereich der Forschung und Entwicklung

Association for Educational Assessment Europe (AEA-E)

Institutionelle Mitgliedschaft des IQB, Kooperation bei der Ausrichtung der AEA-E Jahrestagung 2012 in Berlin, www.aea-europe.net

BeLDACH-Verbund bestehend aus

- Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens (bifie), www.bifie.at
- Ministerium der Deutschsprachigen Gemeinschaft in Belgien, www.dglive.be
- Ministère de l'Éducation nationale et de la Formation professionnelle Luxembourg (MEN, www.men.public.lu)
- Universität Luxemburg, Forschungseinheit Educational Measurement and Applied Cognitive Science (EMACS), www.uni.lu/recherche/flshase/educational_measurement_and_applied_cognitive_science_emacs
- Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK), www.edk.ch

Regelmäßige informelle Kooperation und Konsultation der deutschsprachigen Länder im Bereich der Qualitätsentwicklung im Bildungswesen.

Cito Niederlande und Cito Deutschland

Kooperation bei der Bearbeitung statistischer und testtheoretischer Fragestellungen der Kompetenzmessung, gemeinsame Entwicklung von Übungsmaterialien in den Fremdsprachen, www.de.cito.com

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

Kooperation im Rahmen der Entwicklung und Erprobung technologiebasierter Assessments; Kooperation für die Erstellung eines Forschungsberichts mit Validitätsanalysen für einen Prozessindikator zur Qualität von Schule und Unterricht; Entwicklung eines Musterindikators, www.dipf.de

Deutsches Zentrum für Lehrerbildung Mathematik (DZLM)

Institutionelle Mitgliedschaft des IQB, Kooperation im Bereich der Evaluation von Lehrerbildung, www.dzlm.de

Educational Testing Service (ETS), Princeton, NJ, USA

Kooperation im Bereich der Entwicklung und Anwendung aktueller Ansätze zur Individualdiagnostik (kognitive Diagnosemodelle), www.ets.org

IGLU/TIMSS-Konsortium

(Leitung: Prof. Dr. Wilfried Bos)

Kooperation im Rahmen des nationalen und internationalen Bildungsmonitorings in der Primarstufe, www.ifs-dortmund.de

Institut für Schulqualität der Länder Berlin und Brandenburg e. V. (ISQ)

Kooperation im Rahmen der Überprüfung von Prüfungsaufgaben für den Mittleren Schulabschluss (MSA) in Berlin und in verschiedenen Forschungs- sowie Evaluationsvorhaben, www.isq-bb.de

Landesinstitut für Schule und Medien Berlin-Brandenburg (LISUM)

Erarbeitung und Durchführung einer Lehrerfortbildung zur Bewertung und Weiterentwicklung von Lernaufgaben (Projekt LAGOS), www.lisum.berlin-brandenburg.de

Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik an der Universität Kiel (IPN)

(Leitung: Prof. Dr. Olaf Köller)

Kooperation in verschiedenen Forschungsprojekten, z. B. einer Untersuchung zur Vergleichbarkeit der in NEPS, PISA und im IQB-Ländervergleich 2012 erfassten Mathematik- und Naturwissenschaftskompetenzen; Studie zu Effekten des Programms SINUS an Grundschulen und zur Entwicklung der mathematisch-naturwissenschaftlichen Kompetenzen in der beruflichen Erstausbildung; Studie zum Vergleich der Rahmenkonzepte und Validierung der CITO-Testinstrumente in Deutsch, www.ipn.uni-kiel.de

Leibniz-Forschungsverbund ‚Bildungspotenziale‘ (LERN),

außer dem IQB bestehend aus

- Deutsches Institut für Erwachsenenbildung – Leibniz-Zentrum für Lebenslanges Lernen e. V. (DIE)
- Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
- Deutsches Institut für Wirtschaftsforschung (DIW Berlin), Sozio-oekonomisches Panel (SOEP),
- Georg-Eckert-Institut – Leibniz-Institut für internationale Schulbuchforschung (GEI)
- GESIS – Leibniz-Institut für Sozialwissenschaften
- ifo-Institut – Leibniz-Institut für Wirtschaftsforschung an der Universität München
- Leibniz-Institut für Bildungsverläufe e.V. (LifBi)
- Leibniz-Institut für Agrarentwicklung in Mittel- und Osteuropa (IAMO), (assoziiertes Mitglied)
- Leibniz-Institut für Wissensmedien (IWM)
- Leibniz-Institut für Neurobiologie (LIN)
- Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN)
- Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID)
- Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
- Wissenschaftszentrum Berlin für Sozialforschung (WZB)
- Zentrum für Europäische Wirtschaftsforschung (ZEW)

Das IQB ist als assoziiertes Mitglied des Forschungsverbundes an der interdisziplinären Erforschung von Bedingungen zur Optimierung der gesellschaftlichen Entwicklung und Nutzung von Bildungspotenzialen beteiligt, www.leibniz-gemeinschaft.de/forschung/leibniz-forschungsverbuende/bildungspotenziale/

Leuphana-Universität Lüneburg, Institut für Bildungswissenschaft

(Prof. Dr. Timo Ehmke)

Kooperation im Rahmen einer Untersuchung zur Vergleichbarkeit der in NEPS, PISA und im IQB-Ländervergleich 2012 erfassten Mathematik- und Naturwissenschaftskompetenzen, www.leuphana.de/timo-ehmke.html

Network of Large Scale Studies on Students with Special Educational Needs (NELSEN),

außer IQB bestehend aus DIPF (siehe oben), WZB, NEPS, ZIB (siehe unten), Zusammenarbeit bei Fragen der Identifizierung und Kategorisierung von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf in Vergleichsstudien

Qualitätsagentur am Staatsinstitut für Schulqualität und Bildungsforschung München (ISB)

(Prof. Dr. Eva-Maria Lankes)

Kooperation zur Untersuchung der Skalierbarkeit von Kompetenztestdaten in der Domäne „Schreiben“ der VERA-3-Pilotierungserhebung, <http://www.isb.bayern.de/ueber-das-isb/organisation/qualitaetsagentur/>

Technische Universität München

(Prof. Dr. Kristina Reiss), in Zusammenarbeit mit der

National Taichung University of Education

Kooperation mit dem Ziel der Optimierung der Testkonstruktion der Vergleichsarbeiten (VERA) und der systematischen Diagnostik fehlerhafter Lösungsstrategien von Schülerinnen und Schülern beim Bearbeiten von Testaufgaben, <http://www.ma.edu.tum.de/index.php?id=16>

Universität Potsdam

(Prof. Dr. Miriam Vock)

Kooperation im Rahmen des DFG-Projekts „Das Überspringen einer Klassenstufe als schulische Akzelerationsmaßnahme: Bestandsaufnahme, Bedingungen und Auswirkungen“, <http://www.uni-potsdam.de/unterrichtsinterventionsforsch/profdmirmiamvock.html>

University of California, Berkeley, USA, Berkeley Evaluation and Assessment Research (BEAR) Center

(Leitung: Prof. Dr. Mark Wilson)

Kooperation in Fragen der Entwicklung und Durchführung von Standard-Setting-Verfahren, <http://bearcenter.berkeley.edu>

Wissenschaftszentrum Berlin für Sozialforschung (WZB)

(Jutta Höhne)

Kooperation zur Erstellung des Zweiten Integrationsindikatorenberichts für die Bundesbeauftragte für Migration, Flüchtlinge und Integration, <http://www.wzb.eu/de/personen/jutta-hoehne>

Zentrum für Internationale Vergleichsstudien (ZIB)

(Prof. Dr. Manfred Prenzel, Prof. Dr. Eckhard Klieme, Prof. Dr. Olaf Köller)

Kooperation im Rahmen des nationalen und internationalen Bildungsmonitorings in der Sekundarstufe I, Zusammenarbeit im Rahmen der Studie PISA-Follow-Up 2013 („PISA Plus“), Gemeinsame Ausrichtung der Jahrestagung der Association for Educational Assessment Europe (AEA-Europe) 2012, www.zib-cisa.de

Kooperationen im Bereich der Qualifikation von wissenschaftlichem Nachwuchs

Berlin Interdisciplinary Education Research Network (BIEN)

Monatliche Seminare, Jahrestagungen und Methodenworkshops für Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftler aus verschiedenen Disziplinen im Bereich der Bildungsforschung, <http://bien-edu.net/>

College for Interdisciplinary Educational Research – gemeinsame Initiative des BMBF, der Jacobs Foundation und der Leibniz-Gemeinschaft

Programm für Postdoktorandinnen und Postdoktoranden kurz nach der Promotion mit dem Ziel der Förderung ihrer frühen wissenschaftlichen Eigenständigkeit im Bereich interdisziplinärer Forschung zur Bildung im Lebenslauf, www.ciderweb.org/

International Max Planck Research School on the Life Course (LIFE)

Gemeinsames Graduiertenprogramm des Max-Planck-Instituts für Bildungsforschung, der Freien Universität Berlin, der Humboldt-Universität zu Berlin, der University of Michigan, der University of Virginia und der Universität Zürich (UZH) mit Seminaren, Akademien und Möglichkeiten für Auslandsaufenthalte sowie internationale und interdisziplinäre Kooperationen, www.imprs-life.mpg.de/en

Kooperationspartner des Forschungsdatenzentrums (FDZ)

DIPF – Informationszentrum Bildung (IZB)

Austausch im Bereich Datenmanagement und Ausbau der Forschungsdateninfrastruktur, Zusammenarbeit bei der Bereitstellung und Verlinkung von Instrumenten der am FDZ vorgehaltenen Daten, www.dipf.de/de/forschung/abteilungen/informationszentrum-bildung

GESIS – Zentralarchiv für empirische Sozialforschung (ZA) Köln

Austausch im Bereich Datenmanagement und Ausbau der Forschungsdateninfrastruktur, www.gesis.org/das-institut/

IEA-ETS Research Institute

Zusammenarbeit bei der Organisation internationaler Fortbildungsveranstaltungen zu *Large-Scale-Assessments*, www.ierinstitute.org/

Institut zur Zukunft der Arbeit (IZA)

Bereitstellung des Remote-Computing-Systems JoSuA durch das IZA und Zusammenarbeit bei der Entwicklung von Metadatenstandards, www.iza.org/de

RatSWD – Ausschuss Forschungsdateninfrastruktur (fdi)

Austausch zu aktuellen Fragen des Datenangebots und -zugangs mit circa 25 Forschungsdaten- und Datenservicezentren in ganz Deutschland, <http://ratswd.de/forschungsdaten/fdi>

Kooperationsbeziehungen im Rahmen von Drittmittelprojekten der IQB-Leitung an der Humboldt-Universität

Bund-Länder-Initiative „Bildung durch Sprache und Schrift“ (BISS)

Mitglied im Trägerkonsortium (Prof. Dr. Michael Becker-Mrotzek, Prof. Dr. Marcus Hasselhorn, Prof. Dr. Hans-Joachim Roth, Prof. Dr. Petra Stanat) zur Umsetzung des fünfjährigen Forschungs- und Entwicklungsprogramms zur Sprach- und Leseförderung, <http://www.mercator-institut-sprachfoerderung.de/projekte/bildung-durch-sprache-und-schrift-biss.html>

BMBF-Projekt „Schulen als Steuerungsakteure im Bildungssystem“ (StABil)

Verbundprojekt mit der Freien Universität Berlin (Prof. Dr. Felicitas Thiel) im Rahmen des Forschungsschwerpunkts „Steuerung im Bildungssystem“ (SteBis), www.stebis.de/forschungsprojekte/thielpant

DFG-Projekt „Literaturästhetische Urteilskompetenz“ (LUK)

Verbundprojekt mit der Humboldt-Universität zu Berlin (Prof. Dr. Thorsten Roick) und der Universität Erlangen-Nürnberg (Prof. Dr. Volker Frederking) im Rahmen des DFG-Schwerpunktprogramms „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“, <http://kompetenzmodelle.dipf.de/de/projekte/projekt-literaeraesthetische-urteilskompetenz>

BMBF-Projekt „Effekte kombinierter musikalischer und phonologischer Frühförderung auf die Entwicklung phonologischer Bewusstheit bei Kindergartenkindern deutscher und nicht-deutscher Herkunftssprache“ (Phono II)

Verbundprojekt im Rahmen der Forschungsinitiative Sprachdiagnostik Sprachförderung (FISS) mit der Julius-Maximilians-Universität Würzburg (Prof. Dr. Wolfgang Schneider) und der Otto-Friedrich-Universität Bamberg (Prof. Dr. Cordula Artelt), www.fiss-bmbf.uni-hamburg.de/projekt-stanat%20phonologische%20bewusstheit.html

Nationales Bildungspanel (National Educational Panel Study, NEPS)

Zusammenarbeit im Rahmen der Säule 4 „Bildungserwerb von Personen mit Migrationshintergrund“ des NEPS (Prof. Dr. Frank Kalter, Prof. Dr. Cornelia Kristen), www.neps-data.de/

BMBF-Projekt Kompetenzerwerb und Lernvoraussetzungen (KuL)

Verbundprojekt mit der Otto-Friedrich-Universität Bamberg (Prof. Dr. Cornelia Kristen) und der Universität Mannheim (Prof. Dr. Irena Kogan), www.forschung-kul.de

DFG-Projekt „Immigrants’ first and second language proficiency and social integration“

Verbundprojekt im Rahmen des DFG-Schwerpunktprogramms „Education as a Lifelong Process“ mit der Otto-Friedrich-Universität Bamberg (Prof. Dr. Cornelia Kristen), <https://spp1646.neps-data.de/Projects/ProjectKristenStanat.aspx>

BMBF-Projekt „Bildungssprachliche Kompetenzen: Anforderungen, Sprachverarbeitung und Diagnostik“ (BiSpra II)

Verbundprojekt im Rahmen der Forschungsinitiative Sprachdiagnostik Sprachförderung (FISS) mit der Otto-Friedrich-Universität Bamberg (Prof. Dr. Sabine Weinert), www.fiss-bmbf.uni-hamburg.de/projekt-stanat%20bispra.html

BMBF-Projekt Kompetenzmodellierung und -erfassung zum Wissenschaftsverständnis über naturwissenschaftliche Arbeits- und Denkweisen bei Studierenden (Lehramt) in den drei naturwissenschaftlichen Fächern Biologie, Chemie und Physik (Ko-WADiS)

Verbundprojekt im Rahmen des BMBF-Forschungsprogramms „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor“ (KoKoHs) mit der Universität Duisburg-Essen (Prof. Dr. Philipp Schmiemann), der Universität Wien (Prof. Dr. Günther Pass, Prof. Dr. Martin Hopf, Dr. Franz Radits), der Universität Innsbruck (Dr. Suzanne Kapilari), der Freien Universität Berlin (Prof. Dr. Dirk Krüger, Prof. Dr. Volkhard Nordmeier, Prof. Dr. Felicitas Thiel) und der Humboldt-Universität zu Berlin (Prof. Dr. Annette Upmeyer zu Belzen, Prof. Dr. Rüdiger Tiemann), www.hu-berlin.de/biologie/biodidaktik/forschung/kowadis.html

Universität Ulm, Institut für Psychologie und Pädagogik (Prof. Dr. Oliver Wilhelm)

Zusammenarbeit bei der längsschnittlichen Untersuchung individueller schulischer Entwicklungsprozesse sowie im Bereich der technologiebasierten Kompetenzdiagnostik, Finanzierung durch den Stifternverband für die Deutsche Wissenschaft e. V., www.uni-ulm.de/in/psy-paed/dia/team/oliver-wilhelm.html



02 // Das Kerngeschäft des IQB

Bildungsstandards und ihre Operationalisierung mit Aufgaben

Die Bildungsstandards der Kultusministerkonferenz (KMK) bilden eine wichtige Grundlage der Qualitätsentwicklung und Qualitätssicherung im deutschen Bildungswesen und damit auch für die Arbeit des IQB. Zu den Aufgaben des IQB gehören die Operationalisierung, Normierung und Weiterentwicklung der Bildungsstandards sowie die Überprüfung, inwieweit die länderübergreifend vereinbarten Kompetenzziele erreicht werden. In diesem Kapitel werden die ersten drei dieser Aufgaben genauer beschrieben. Auf die Überprüfung des Erreichens der Bildungsstandards wird in Kapitel 3 ausführlich eingegangen.

Geschichte der Bildungsstandards in Deutschland

Mit Beschlüssen von 2003, 2004 und 2012 hat die KMK für ausgewählte Fächer Bildungsstandards für den Primarbereich, den Hauptschulabschluss (HSA) und den Mittleren Schulabschluss (MSA) sowie für die Allgemeine Hochschulreife verabschiedet. Diese Vorgaben sind für alle 16 Länder in der Bundesrepublik verbindlich.

Klieme und Kollegen (2007, S. 24 ff.) sowie Köller (2009, S. 530 f.) benennen zentrale Merkmale von Bildungsstandards, die als Grundlage für die Weiterentwicklung von Bildungsqualität geeignet sind. Demnach zeichnen sich „gute“ Bildungsstandards neben ihrer Fachspezifität und Fokussierung auf die Kernbereiche des jeweiligen Faches auch dadurch aus, dass sie realistische Erwartungen spezifizieren und kumulative Lernprozesse begünstigen. Ferner sollten sie die Möglichkeit zur Differenzierung verschiedener Kompetenzniveaus eröffnen und so formuliert sein, dass sich Messinstrumente entwickeln lassen, mit denen die beschriebenen Kompetenzen erfasst werden können. Bei den Bildungsstandards der KMK handelt es sich um Regelstandards, die definieren, welche Kompetenzerwartungen *im Durchschnitt* bis zum jeweiligen Bildungsabschnitt von den Schülerinnen und Schülern erfüllt werden sollen.

Bildungsstandards sind normative Kompetenzerwartungen. Sie sind das Resultat eines Aushandlungsprozesses zwischen unterschiedlichen Entscheidungsträgern im Bildungssektor – zwischen Bildungspolitik und Bildungsadministration, Fachdidaktiken, Erziehungs- und Bildungswissenschaften sowie Vertretern der Schulpraxis. Mit der Einführung der Bildungsstandards wurde in Deutschland eine

erhebliche Veränderung in der Steuerung des allgemein bildenden Schulsystems ausgelöst, die häufig auch als Paradigmenwechsel von der Input- zur Outputsteuerung beschrieben wird (vgl. Klieme et al., 2007). Dabei erhalten in der Steuerung von Bildungsqualität die tatsächlich erreichten Bildungserträge auf allen Ebenen des Schulsystems stärkere Bedeutung als es bei der bisher dominierenden Inputsteuerung über neue didaktische Modelle, modifizierte Lehrpläne oder veränderte Lehramtsausbildungen der Fall war. Mit den Bildungsstandards ist das Ziel verbunden, die Kompetenzorientierung im Bildungssystem zu stärken. Diese Fokussierung soll dazu beitragen, dass im Unterricht anstelle von trägem Wissen, das Schülerinnen und Schüler nur zur Beantwortung von eng begrenzten und bekannten Aufgabenstellungen abrufen können, vernetztes Wissen entwickelt wird, das zur flexiblen Bewältigung von Problemen angewendet werden kann (vgl. z. B. Blum, Drücke-Noe, Leiß, Wiegand & Jordan, 2005; Bremerich-Vos, 2009). Dabei handelt es sich um einen sehr ambitionierten fachdidaktischen und pädagogischen Anspruch, bei dessen Einlösung Tests und Leistungsrückmeldungen lediglich eine unterstützende Funktion haben können.

Die verbindliche Einführung der Bildungsstandards erfolgte für den Primarbereich, den HSA und den MSA in allen Ländern zum Schuljahresbeginn 2004/2005 bzw. 2005/2006 (siehe Tabelle 2.1). Im Oktober 2007 beschloss die KMK, auch Bildungsstandards für die Allgemeine Hochschulreife zu entwickeln. Diese wurden für die Fächer Deutsch, die fortgeführte Fremdsprache (Englisch, Französisch) und Mathematik im Oktober 2012 verabschiedet. Ob bzw. wann Bildungsstandards für die Allgemeine Hochschulreife auch in den Fächern Biologie, Chemie und Physik entwickelt werden, wird die KMK nach Auswertung des IQB-Ländervergleichs

Tabelle 2.1: Von der KMK verabschiedete Bildungsstandards (Stand: Dezember 2013)

	Primarbereich Ende der 4. Jahrgangsstufe	Ende der Sekundarstufe I		Ende der Sekundarstufe II Allgemeine Hochschulreife
		Hauptschulabschluss	Mittlerer Abschluss	
Deutsch	x	x	x	x
Mathematik	x	x	x	x
Englisch		x	x	x
Französisch		x	x	x
Biologie			x	*
Chemie			x	*
Physik			x	*

Anmerkung: * Über die Entwicklung dieser Bildungsstandards wird voraussichtlich im Jahr 2015 entschieden.

2012 (siehe Kapitel 3) entscheiden, in dem erstmals das Erreichen der Bildungsstandards für den MSA in den naturwissenschaftlichen Fächern überprüft wurde. Alle aktuell in Deutschland vorliegenden Bildungsstandards sind auf den Internetseiten der KMK zu finden (<http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsstandards/dokumente.html>).

Bildungsstandards beschreiben die zu entwickelnden Kompetenzen in jedem Fach möglichst so konkret, dass das Erreichen der Erwartungen mit Testverfahren überprüft werden kann. Eine Hauptaufgabe des IQB besteht darin, auf der Grundlage der Bildungsstandards für den Primarbereich und die Sekundarstufe I Testaufgaben zu entwickeln und im Ländervergleich zu überprüfen, inwieweit die angestrebten Kompetenzniveaus erreicht wurden. Die regelmäßige Überprüfung des Erreichens der Bildungsstandards („Bildungsmonitoring“) soll Stärken und Schwächen in den Bildungssystemen der Länder identifizieren und Hinweise auf möglichen Handlungsbedarf liefern.

Nach einem im Frühjahr 2011 getroffenen Beschluss der KMK sollen zu den Bildungsstandards für die Allgemeine Hochschulreife keine Testaufgaben und Kompetenzstufenmodelle entwickelt werden, so dass sich die Arbeiten des IQB im Bereich der Sekundarstufe II von denen in der Primarstufe und in der Sekundarstufe I unterscheiden. Sie werden daher am Ende dieses Kapitels separat beschrieben.

Der Prozess der Aufgabenentwicklung für die Ländervergleichsstudien des IQB in der Primarstufe und in der Sekundarstufe I

Ob und inwieweit es gelingt, die mit den Bildungsstandards definierten Kompetenzziele zu erreichen, wird regelmäßig in mehrjährigen Abständen vom IQB überprüft und berichtet. Die Ergebnisse werden dabei für die einzelnen Länder in der Bundesrepublik Deutschland dargestellt (siehe Kapitel 3). Zur Durchführung dieser Ländervergleichsstudien wird eine große Zahl von Testaufgaben¹ benötigt, die in ihrer Gesamtheit die Vorgaben der Bildungsstandards angemessen abbilden und möglichst zuverlässige und umfassende Informationen über die Kompetenzen der Schülerinnen und Schüler in den getesteten Bereichen liefern. Die Entwicklung von Aufgaben, die diesen Anforderungen genügen, ist aufwändig, anspruchsvoll und komplex. In der ersten Arbeitsphase des IQB in den Jahren 2005 und 2006 wurden die Grundsteine für

die Entwicklung von Testaufgaben gelegt (vgl. Köller, 2008). So initiierte das IQB in den Anfangsjahren die Testkonstruktion für die Fächer Deutsch und Mathematik im Primarbereich und für die Fächer Mathematik, Deutsch, Englisch und Französisch in der Sekundarstufe I. Dabei wurde aus fachdidaktischer und lerntheoretischer Sicht präzisiert, was unter den jeweiligen Kompetenzen in den einzelnen Fächern zu verstehen ist. Weiterhin wurden Schulungen konzipiert und umgesetzt, in denen die Aufgabenentwicklerinnen und Aufgabenentwickler in die Lage versetzt werden sollen, fachdidaktisch und messtechnisch geeignete Testaufgaben zu konstruieren.

Allgemein umfasst der Prozess der Entwicklung eines normierten Testaufgabenpools am IQB die in Abbildung 2.1 dargestellten fünf Teilschritte, von denen zu Beginn des Berichtszeitraums bereits folgende Schritte begonnen worden waren:

- im Primarbereich in den Fächern Deutsch und Mathematik die Schritte 1 bis 3
- in der Sekundarstufe I in den Fächern Deutsch, Englisch und Französisch die Schritte 1 und 2
- in der Sekundarstufe I im Fach Mathematik die Schritte 1 bis 3 sowie Teile von Schritt 4 (entscheidende Arbeitsanteile waren hier vom deutschen PISA-Konsortium [PISA-2006] geleistet worden; Federführung für PISA insgesamt: Prof. Dr. Manfred Prenzel; Federführung Mathematik: Prof. Dr. Werner Blum).

Im Berichtszeitraum 2007 bis 2013 wurde der Prozess der Aufgabenentwicklung in allen genannten Fächern und Schulstufen weitergeführt bzw. in den naturwissenschaftlichen Fächern Biologie, Chemie und Physik der Sekundarstufe I begonnen und mit der Entwicklung von Kompetenzstufenmodellen abgeschlossen.

1. Teilschritt: Aufgabenkonstruktion

Um empirisch und fachdidaktisch fundierte Kompetenzstufenmodelle definieren zu können, wird am IQB in jedem Fach eine große Anzahl von Items entwickelt, mit denen ein breites Fähigkeitsspektrum auf Seiten der Schülerinnen und Schüler erfasst werden kann.² Der Aufbau eines solchen Itempools erfordert interdisziplinäre Arbeitsstrukturen, in denen erfahrene und intensiv geschulte Lehrkräfte, ausgewiesene Vertreterinnen bzw. Vertreter der Fachdidaktiken sowie Expertinnen und Experten aus der empirischen Bildungsforschung eng kooperieren.

¹ Eine Testaufgabe besteht meistens aus mehreren Teilaufgaben, den sogenannten Items.

² Im Fach Französisch konnten zusätzlich zu Eigenentwicklungen des IQB 360 Items in den Aufgabenpool aufgenommen werden, die vom *Centre International d'Etudes Pédagogiques* (CIEP) in Sèvres konstruiert worden waren.

Die Entwicklung von Testaufgaben verläuft in den verschiedenen Fächern und Schulstufen sehr ähnlich. Zunächst legen die Lehrkräfte, die am Aufgabenentwicklungsprozess beteiligt sind, erste Ideen und Entwürfe für Testaufgaben vor. Diese werden in den Aufgabenentwicklergruppen, in denen in der Regel Lehrkräfte aus allen 16 Ländern zusammenarbeiten und die sich mehrmals im Jahr treffen, gemeinsam mit Fachdidaktikerinnen und Fachdidaktikern diskutiert und ggf. verändert. Um authentische Schülerlösungen für die Weiterentwicklung der Aufgaben zu erhalten, erproben die Gruppen diese in einzelnen Schulklassen (Pilotierungen).

2. Teilschritt: Optimierung der Aufgaben

Expertinnen und Experten aus der empirischen Bildungsforschung und der Fachdidaktik beurteilen und kommentieren in einem zweiten Schritt die Items. Die in der Gründungsphase des IQB angebahnten Kooperationen mit Expertinnen und Experten in den jeweiligen Fächern wurden im Berichtszeitraum weiter ausgebaut (siehe Kapitel 1). Alle Items, die in diesem Begutachtungsprozess nach meist mehreren Korrekturschleifen zwischen den Aufgabenentwicklerteams und den Aufgabenbewerterinnen und -bewertern als prinzipiell geeignet eingeschätzt wurden, werden in einer Itemdatenbank verwaltet. Aus dieser Datenbank können sie dann für den weiteren Einsatz in Pilotierungs- und Normierungsstudien ausgewählt werden.

3. Teilschritt: Pilotierungsstudien

Im Rahmen der Pilotierungsstudien werden einer Stichprobe von Schülerinnen und Schülern Testhefte mit Items des jeweiligen Fachs zur Bearbeitung vorgelegt (siehe Tabelle 2.2). Diese Stichprobe wird dabei so ausgewählt, dass sie die

Heterogenität der Schülerschaft in Deutschland abbildet, ohne im strengen Sinne repräsentativ zu sein. In diesem Schritt der Aufgabenentwicklung geht es nicht darum, die Kompetenzausprägungen der einzelnen Schülerinnen und Schüler zu messen, sondern es soll überprüft werden, ob die Items aus psychometrischer Sicht zur Erfassung von Kompetenzen geeignet sind. Da solche Messungen objektiv, reliabel, valide und fair sein sollen, müssen die Items verschiedene Kriterien erfüllen, die das IQB anhand der Pilotierungsdaten überprüft. Dies betrifft unter anderem die folgenden Aspekte:

- **Schwierigkeit:** Die Items dürfen weder zu leicht noch zu schwer sein, sollen aber zugleich ein relativ breites Kompetenzspektrum abdecken.
- **Modellpassung:** Die Items müssen mit Modellen der probabilistischen Testtheorie (auch: *Item-Response-Theorie*, IRT) übereinstimmen, die in Schulleistungsstudien häufig angewendet werden. Inwieweit dies der Fall ist, wird anhand eines itemspezifischen Fit-Index erfasst. Ein Item mit „schlechtem“ Fit wäre beispielsweise eines, dessen Lösungswahrscheinlichkeit nicht oder nur wenig mit der Kompetenzausprägung der Schülerin oder des Schülers in Zusammenhang steht.
- **Fairness:** Ob ein Item richtig gelöst wird, soll nur von der Kompetenzausprägung der Schülerin oder des Schülers abhängen und nicht von anderen Merkmalen der Person (z. B. dem Geschlecht) oder des Items (z. B. ob das Item an vorderer oder hinterer Stelle im Testheft steht).

Items, die eines oder mehrere der Kriterien nicht erfüllen, werden genauer analysiert und gegebenenfalls überarbeitet, bevor sie in weiteren Studien eingesetzt werden, oder sie werden ausgeschlossen. Das Endergebnis dieses Arbeitsschrittes ist ein Pool mit optimierten Testaufgaben, die in Normierungsstudien eingesetzt werden können.

■ **Abbildung 2.1: Die fünf Teilschritte zur Entwicklung eines Pools von normierten Testaufgaben am IQB**

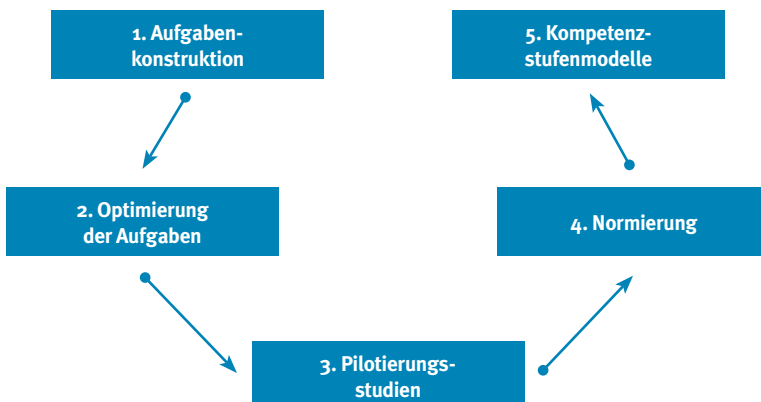


Tabelle 2.2: Überblick über die Pilotierungsstudien des IQB

	Zeitpunkt der Datenerhebung	Anzahl getesteter Items/Aufgaben¹	Stichprobenumfang^{1,2}
Mathematik Sekundarstufe I	2005	650 Items	6 000
Mathematik Primarstufe	2006	650 Items	10 300
Deutsch Primarstufe	2006	1060 Items	10 600
Deutsch Sekundarstufe I	2007	400 Items im Bereich <i>Zuhören</i> 600 Items im Bereich <i>Lesen</i> 500 Items in den Bereichen <i>Rechtschreibung</i> und <i>Sprachgebrauch</i> 15 Aufgaben im Bereich <i>Schreiben</i>	3 900
Englisch Sekundarstufe I	2007	350 Items im Bereich <i>Hörverstehen</i> 390 Items im Bereich <i>Lesen</i> 19 Aufgaben im Bereich <i>Schreiben</i>	3 000
Französisch Sekundarstufe I	2007	360 Items im Bereich <i>Hörverstehen</i> und <i>Lesen</i> , 17 Aufgaben im Bereich <i>Schreiben</i>	2 700
Naturwissenschaften Sekundarstufe I	2009	Bereich <i>Erkenntnisgewinnung</i> : 190 Items für Biologie 120 Items für Chemie 240 Items für Physik Bereich <i>Fachwissen</i> : 160 Items für Biologie 180 Items für Chemie 160 Items für Physik	6 900
Mathematik Sekundarstufe I	2009	210 Items	760
Deutsch Sekundarstufe I	2010	13 Aufgaben im Bereich <i>Schreiben</i>	2 000
Naturwissenschaften Sekundarstufe I	2011	Bereich <i>Bewertung</i> : 170 Items für Biologie 120 Items für Chemie 100 Items für Physik	3 100
Mathematik Sekundarstufe I	2011	410 Items	3 500

Anmerkungen:

¹ Gerundete Angaben.

² Zur Schätzung der Güte eines Items muss dieses von einer ausreichend großen Schülerstichprobe bearbeitet worden sein. Unterschiedliche Bearbeitungszeiten je nach Fach und Kompetenzbereich, verschieden lange Testzeiten in den Klassenstufen und unterschiedliche Itemzahlen führen dazu, dass die Gesamtstichprobenumfänge der Pilotierungsstudien variieren.

Die in Pilotierungs- und Normierungsstudien eingesetzten Itemmengen sind so groß, dass die einzelnen Schülerinnen und Schüler sie nicht alle bearbeiten könnten. Jede Schülerin und jeder Schüler erhält daher ein Testheft, das nur eine Teilmenge der Aufgaben enthält. Damit die Items dennoch bei der Auswertung in ihrer Schwierigkeit auf einer gemeinsamen Skala abgebildet werden können, sind die Testhefte so konzipiert, dass bestimmte Aufgaben in mehreren Heften vorkommen. Die Zusammenstellung der Testhefte erfolgt dabei so, dass alle Aufgaben indirekt miteinander verlinkt sind. Ein solches komplexes Testdesign wird auch als *Multi-Matrix-Design* bezeichnet. Die Erarbeitung des Designs und die Auswertung der Daten aus den Pilotierungsstudien liegt in der Verantwortung des IQB; mit der Planung und Durchführung der Datenerhebungen (Stichprobenziehung, Auswahl und Schulung der Testleiterinnen und Testleiter, Versendung des Erhebungsmaterials, Koordinierung der Testdurchführung etc.) sowie mit der Dateneingabe wurde das *Data Processing Center* (DPC) in Hamburg als externer Dienstleister beauftragt.

4. Teilschritt: Normierung

Die Konzeption und Durchführung der Normierungsstudien folgt im Wesentlichen den gleichen Prinzipien wie bei den Pilotierungsstudien. Die Testaufgaben, die sich in der Pilotierung bewährt haben, werden wiederum in einer größeren Anzahl von untereinander verlinkten Testheften zusammengestellt und von einer großen Stichprobe von Schülerinnen und Schülern aus allen Ländern der Bundesrepublik Deutschland bearbeitet (siehe Tabelle 2.3), wobei die Stichprobe einer Normierungsstudie im statistischen Sinne repräsentativ sein muss. Auf der Basis der resultierenden Daten werden wiederum – wie auch in den Pilotierungsstudien – die statistischen Kennwerte der einzelnen Aufgaben bestimmt. Anhand dieser Kennwerte werden zunächst Items aus dem Pool entfernt, die sich trotz Überarbeitung im Anschluss an die Pilotierung als problematisch erweisen. Anschließend werden die Daten für die verbleibenden Items auf eine Skala transformiert, die den länderübergreifenden Maßstab für die erreichten Kompetenzen der Schülerinnen und Schüler bilden. Diese Metrik (in der Regel Mittelwert $M = 500$, Standardabweichung $SD = 100$) wird sowohl in den Ländervergleichsstudien des IQB (siehe Kapitel 3) als auch in den Vergleichsarbeiten (siehe Kapitel 4) verwendet. Anhand der Normierungsdaten werden zudem Kompetenzstufenmodelle für jedes Fach entwickelt, die ebenfalls in den Ländervergleichsstudien und in den Vergleichsarbeiten zum Einsatz kommen.

5. Teilschritt: Kompetenzstufenmodelle

Um die Ergebnisse von Kompetenztestungen anschaulich darstellen zu können, werden diese oft nicht nur in Form von Punktwerten berichtet, sondern zusätzlich auch in Bezug auf Kompetenzstufenmodelle. Ein solches Kompetenzstufenmodell ist exemplarisch in Abbildung 2.2 auf Seite 35 für das Fach Mathematik in der Primarstufe dargestellt. Kompetenzstufenmodelle ermöglichen es, jede getestete Person und gleichzeitig jedes eingesetzte Item einem konkreten, inhaltlich beschreibbaren Kompetenzniveau zuzuordnen. So sind Aussagen darüber möglich, welche Anforderungen Schülerinnen und Schüler, die ein bestimmtes Kompetenzniveau erreicht haben, mit hoher Wahrscheinlichkeit bewältigen können.

In einem sogenannten *Standard-Setting-Verfahren* legen im Idealfall Expertinnen und Experten aus den Bereichen Fachdidaktik, empirische Bildungsforschung, Schulverwaltung und Schulpraxis in gemeinsamen Sitzungen sogenannte Schwellenwerte (*Cut-Scores*) fest, die Grenzen zwischen qualitativ unterscheidbaren Kompetenzniveaus markieren (vgl. Pant, Tiffin-Richards & Köller, 2010; Tiffin-Richards & Köller, 2010). Dabei wurde für die Primarstufe und den MSA so vorgegangen, dass die Kompetenzstufe III die in den Bildungsstandards definierten Regelanforderungen der KMK repräsentiert, d. h. das in der Regel am Ende der Bildungsetappe erwartete Kompetenzniveau. Die Kompetenzstufe II definiert den Mindeststandard, den *alle* Schülerinnen und Schüler der Zielpopulation erfüllen sollen. Die Kompetenzstufe IV wird als „Regelstandard Plus“ und die Kompetenzstufe V als „Optimalstandard“ bezeichnet.

Die Stufenbeschreibungen der Modelle spezifizieren, welche kognitiven Anforderungen Schülerinnen und Schüler bewältigen können, wenn sie die jeweilige Kompetenzstufe erreicht haben. Auf diese Weise wird es möglich, die erzielten Schülerleistungen qualitativ zu beschreiben und zu bestimmen, welcher Anteil von Schülerinnen und Schülern die Anforderungen einer Kompetenzstufe bereits mit hoher Sicherheit bewältigt. Für die Darstellung und Interpretation der Ergebnisse der IQB-Ländervergleichsstudien spielen diese Kompetenzstufenmodelle eine zentrale Rolle (siehe Kapitel 3).

Tabelle 2.4 gibt einen Überblick darüber, wann die vom IQB vorgelegten Kompetenzstufenmodelle von der Kultusministerkonferenz verabschiedet wurden (alle verabschiedeten Kompetenzstufenmodelle sind auf den Internetseiten des IQB zu finden: <http://www.iqb.hu-berlin.de/bista/ksm>). Zur Illustration wird im Folgenden die Entwicklung der auf den MSA bezogenen Kompetenzstufenmodelle für den Bereich *Schreiben* im Fach Deutsch und für die naturwissenschaftlichen Fächer ausführlicher dargestellt.

Tabelle 2.3: Überblick über die Normierungsstudien des IQB

	Zeitpunkt der Datenerhebung	Anzahl getesteter Items/Aufgaben	Stichprobenumfang
Mathematik Sekundarstufe I	2006 und 2007	520 Items	14 000
Mathematik Primarstufe	2007	570 Items	7 300
Deutsch Primarstufe	2007	90 Items im Bereich <i>Zuhören</i> 120 Items im Bereich <i>Lesen</i> 150 Items im Bereich <i>Rechtschreibung</i> 110 Items im Bereich <i>Sprachgebrauch</i> 8 Aufgaben im Bereich <i>Schreiben</i>	11 400
Deutsch Sekundarstufe I	2008	290 Items im Bereich <i>Zuhören</i> 300 Items im Bereich <i>Lesen</i> 230 Items im Bereich <i>Rechtschreibung</i> 250 Items im Bereich <i>Sprachgebrauch</i> 11 Aufgaben im Bereich <i>Schreiben</i>	7 900
Englisch Sekundarstufe I	2008	310 Items im Bereich <i>Hörverstehen</i> 350 Items im Bereich <i>Lesen</i> 14 Aufgaben im Bereich <i>Schreiben</i>	5 900
Französisch Sekundarstufe I	2008	370 Items in den Bereichen <i>Hörverstehen</i> und <i>Lesen</i> 12 Aufgaben im Bereich <i>Schreiben</i>	6 000
Deutsch Sekundarstufe I	2011	12 Aufgaben im Bereich <i>Schreiben</i>	3 000
Naturwissenschaften Sekundarstufe I	2011	Bereich <i>Erkenntnisgewinnung</i> : 180 Items für Biologie 140 Items für Chemie 190 Items für Physik Bereich <i>Fachwissen</i> : 180 Items für Biologie 180 Items für Chemie 170 Items für Physik	9 050

Kompetenzstufenmodell für den Bereich Schreiben im Fach Deutsch für den MSA

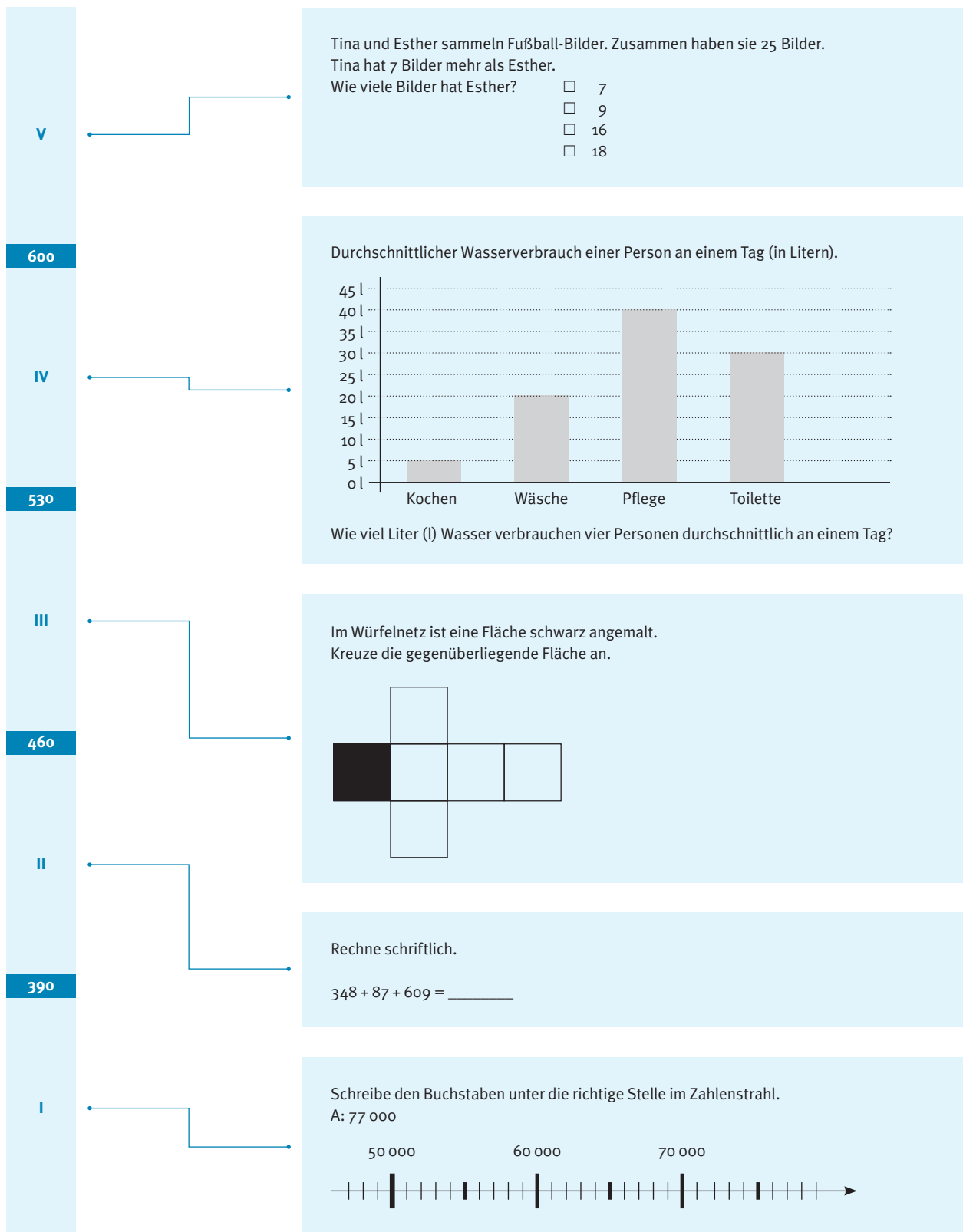
Im Fach Deutsch in der Sekundarstufe I wurden im Zeitraum von Januar 2007 bis Dezember 2013 Kompetenzstufenmodelle für den MSA in den Bereichen *Zuhören*, *Lesen*, *Rechtschreibung*, *Sprache und Sprachgebrauch untersuchen* sowie *Schreiben* entwickelt (siehe Abbildung 2.3). Für die Bereiche *Zuhören*, *Lesen* und *Rechtschreibung* wurden bereits auf der Grundlage der 2008 durchgeführten Normierungsstudie Kompetenzstufenmodelle entwickelt und von der KMK im Jahr 2009 verabschiedet. Für die Bereiche *Sprache und Sprachgebrauch untersuchen* sowie *Schreiben* wurden die Kompetenzstufenmodelle im Frühjahr 2014 von der KMK verabschiedet.

Für den Bereich *Schreiben* wurden im Rahmen der Aufgabenentwicklung jeweils mehrere Aufgaben zu den Textmustern *Argumentation*, *Information* und *Narration* entwickelt. Alle Aufgaben bestanden aus einem Stimulus mit Bild- und/oder Textmaterial sowie einer Instruktion. Die Schülerinnen und Schüler wurden aufgefordert, einen zusammenhängenden vollständigen Text zu einer Frage, einem Thema oder einer Problemstellung zu schreiben. Die *argumentierenden Aufgaben* umfassten sowohl einseitige Argumentationen, etwa das Verfassen eines Bewerbungsschreibens oder die Widerlegung einer vorgegebenen Position, als auch zweiseitige Argumentationen, bei denen sich die Schülerinnen und Schüler mit dem Für und Wider einer Problematik auseinandersetzen sollten. Unter den *informierenden Aufgaben* befanden sich beschreibend-instruierende Aufgaben, wie etwa das Schreiben einer Bauanleitung, sowie berichtende

Tabelle 2.4: Von der KMK verabschiedete Kompetenzstufenmodelle (Stand: Dezember 2013)

Fach und Abschlussart	Kompetenzbereiche	Zeitpunkt (Abschluss)
Mathematik Sekundarstufe I	Globalmodell (umfasst fünf inhaltsbezogene mathematische Kompetenzen)	2008 (MSA), 2009 (HSA), 2011 (integriertes Modell für HSA und MSA)
Mathematik Primarstufe	Globalmodell (umfasst fünf inhaltsbezogene mathematische Kompetenzen)	2008
Englisch Sekundarstufe I	<i>Lesen und Hörverstehen</i>	2009 (getrennt für HSA und MSA)
Französisch Sekundarstufe I	<i>Lesen und Hörverstehen</i>	2009 (MSA)
Deutsch Sekundarstufe I	<i>Zuhören</i> <i>Lesen</i> <i>Rechtschreibung</i>	2009 (MSA)
Deutsch Primarstufe	<i>Zuhören</i> <i>Lesen</i> <i>Rechtschreibung</i> <i>Sprache und Sprachgebrauch untersuchen</i>	2011
Naturwissenschaften (Biologie, Chemie, Physik) Sekundarstufe I	<i>Fachwissen</i> <i>Erkenntnisgewinnung</i>	2011 (MSA)
Deutsch Sekundarstufe I	<i>Schreiben</i> <i>Sprache und Sprachgebrauch untersuchen</i>	2014 (MSA)

Abbildung 2.2: Illustrierende Aufgaben zum globalen Kompetenzstufenmodell im Fach Mathematik für die Primarstufe



Aus Platzgründen sind die Aufgaben in einem modifizierten Layout dargestellt.

Aufgaben, wie das Verfassen einer Zeitungsnachricht (vgl. Abbildung 2.5). Bei den *narrativen Aufgaben* sollten die Jugendlichen einen Text verfassen, in dem eine Ereignisfolge vollständig oder in wesentlichen Teilen erzählend darzustellen war.

Nach Abschluss der Pilotierung dieser Aufgaben im Jahr 2010 konnten ab Frühjahr 2011 die Schülerantworten zu den argumentierenden, informierenden und narrativen Aufgabentypen analysiert werden. Die Kodierung von Schülertexten ist in hohem Maße komplex und erforderte die Entwicklung zweier Auswertungsschemata. Für ein *analytisches Auswertungsschema* definierten die fachdidaktischen Partnerinnen und Partner des IQB im Fach Deutsch 18 Kriterien zur Erfassung sprachlicher, inhaltlicher und struktureller Merkmale der Schülertexte. Diese umfassen unter anderem den Umfang des verwendeten Wortschatzes, Zeichensetzungsfehler und das Vorhandensein einer angemessenen Überschrift. Um neben diesen analytischen Kriterien eine ganzheitliche Beurteilung der Schülertexte vornehmen zu können, wurde am IQB zusätzlich ein *holistisches Auswertungsschema* entwickelt. Dies beinhaltet je eine textmuster-spezifische Globalskala sowie drei semiholistische Subskalen zu Inhalt, Stil und sprachlicher Richtigkeit. In Abbildung 2.4 ist exemplarisch die Subskala *Stil* des holistischen Schemas dargestellt.³

Insgesamt erwiesen sich zwölf Aufgaben für den weiteren Einsatz in der Normierungsstudie für den Kompetenzbereich *Schreiben* als geeignet. Die Aufgabenauswahl umfasste je vier Aufgaben zu argumentativen, informierenden und narrativen Texten. Ein Aufgabenbeispiel für einen informierenden Text ist in Abbildung 2.5 zu finden.

Die Schülertexte aus der 2011 durchgeführten Normierungsstudie wurden anhand der analytischen und holistischen Auswertungsschemata ausgewertet, um auf dieser Grundlage jeweils die textsortenspezifischen Kompetenzstufenmodelle zu argumentativen informierenden und narrativen Texten zu entwickeln. Dabei musste berücksichtigt werden, dass es sich bei den Schreibaufgaben nicht wie in anderen Kompetenzbereichen um eine Vielzahl von Teilaufgaben unterschiedlicher Schwierigkeit, sondern um jeweils vier relativ ähnlich schwere, sehr umfangreiche Einzelaufgaben handelt. Zur Festlegung der Stufengrenzen beim *Standard-Setting* musste daher eine Methode adaptiert und weiterentwickelt werden, die für diese Art von Aufgabenmaterial geeignet ist (vgl. die „Body-of-Work“-Methode bei Kingston, Kahl, Sweeney & Bay, 2001).

■ **Abbildung 2.3: Entwicklung von Kompetenzstufenmodellen im Fach Deutsch in der Sekundarstufe I**



³ Vorbehaltlich redaktioneller Änderungen.

■ Abbildung 2.4: Subskala „Stil“ des holistischen Auswertungsschemas für narrative Texte

Zur Bewertung des Stils ist auf folgende Kriterien zu achten:

- zutreffende Textsorte (...)
- einheitliche Perspektive und ggf. Adressierung (...)
- Kohärenzstiftung / Einsatz von Verknüpfungsmitteln
- Wortwahl
- Vielfalt und Komplexität der Satzkonstruktion
- Tempusgebrauch (...)

Aufgabenspezifische Textelemente:

- Verwendung rhetorischer Stilmittel (Metaphern, Vergleiche, Ironie etc.)

Stufe 4

Texte der Stufe 4 halten die zutreffende Textsorte ein. Die Erzählperspektive ist einheitlich und zutreffend. Aufgrund funktional eingesetzter Verknüpfungsmittel wirken die Texte durchgängig kohärent. Die Wortwahl ist stets angemessen. Der Satzbau ist abwechslungsreich und nicht übermäßig komplex. Der Tempusgebrauch ist konsistent und angemessen. Aufgabenspezifische Textelemente sind durchgängig vorhanden.

Stufe 3

Texte der Stufe 3 halten die zutreffende Textsorte ein. Die Erzählperspektive ist einheitlich und zutreffend. Verknüpfungsmittel sind größtenteils funktional. Inkohärenzen treten nur vereinzelt auf. Die Wortwahl ist zumeist angemessen. Der Satzbau ist größtenteils abwechslungsreich und nicht übermäßig komplex. Der Tempusgebrauch ist überwiegend konsistent und angemessen. Aufgabenspezifische Textelemente sind mehrheitlich vorhanden.

Stufe 2

Texte der Stufe 2 halten die zutreffende Textsorte größtenteils ein. Die Erzählperspektive ist vorwiegend einheitlich und zutreffend. Verknüpfungsmittel finden sich kaum bzw. sind selten funktional. Oftmals sind Teile des Textes inkohärent. Die Wortwahl ist mitunter unangemessen. Die Sätze sind teilweise monoton oder übermäßig komplex konstruiert. Der Tempusgebrauch ist gelegentlich inkonsistent oder unangemessen. Aufgabenspezifische Textelemente sind teilweise vorhanden.

Stufe 1

Texte der Stufe 1 verfehlen meistens die zutreffende Textsorte. Die Erzählperspektive ist bisweilen unklar oder unzutreffend. Verknüpfungsmittel werden nicht oder nicht funktional eingesetzt. Textteile wirken unzusammenhängend. Inkohärenzen treten gehäuft auf. Die Wortwahl ist kaum angemessen. Die Sätze sind überwiegend monoton konstruiert oder unverständlich. Der Tempusgebrauch ist oftmals inkonsistent oder unangemessen. Aufgabenspezifische Textelemente fehlen.

Stufe 0

Texte der Stufe 0 sind zu kurz und bieten keine hinreichende Substanz für eine zuverlässige Bewertung.

Anmerkungen: Einige der aufgeführten Kriterien wurden aufgabenspezifisch konkretisiert. Für nicht auswertbare Schülertexte (beispielsweise bei Verweigerung der Aufgabenbearbeitung) wurden spezielle Bewertungskategorien definiert, die hier nicht dargestellt sind.

Für jedes Textmuster erfolgte das *Standard-Setting* in folgenden Schritten. Zunächst beschrieben die fachdidaktischen Partnerinnen und Partner des IQB die zu erreichenden Kompetenzen für jede der fünf Stufen der Kompetenzstufenmodelle. Expertinnen und Experten aus Schulpraxis, Fachdidaktik, empirischer Bildungsforschung und Bildungsadministration ordneten dann eine Auswahl von Schülerlösungen aus der Normierungsstudie jeweils einer der fünf Stufenbeschreibungen zu. Anschließend erhielten sie vom IQB eine Rückmeldung darüber, bei welchen Einstufungen Uneinigkeit innerhalb der Expertengruppe bestand. Daraufhin wurden die Gründe für die jeweiligen Einstufungen in den Arbeitsgruppen diskutiert und es folgte eine erneute Zuordnung derselben Schülerlösungen zu den Kompetenzstufen. In einem abschließenden Arbeitsschritt wurde eine sogenannte Feinjustierung auf Basis der im bisherigen Ablauf ermittelten vorläufigen Stufengrenzen vorgenommen. Die Expertinnen und Experten erhielten dazu Texte von Schülerinnen und Schülern, deren Kompetenzwerte sehr dicht an den vorläufigen Stufengrenzen und somit faktisch zwischen zwei Stufen lagen, und ordneten diese wiederum jeweils einer Stufe zu. Die Ergebnisse der Feinjustierung wurden schließlich herangezogen, um rechnerisch die genauen Stufengrenzen der Kompetenzstufenmodelle zu ermitteln.

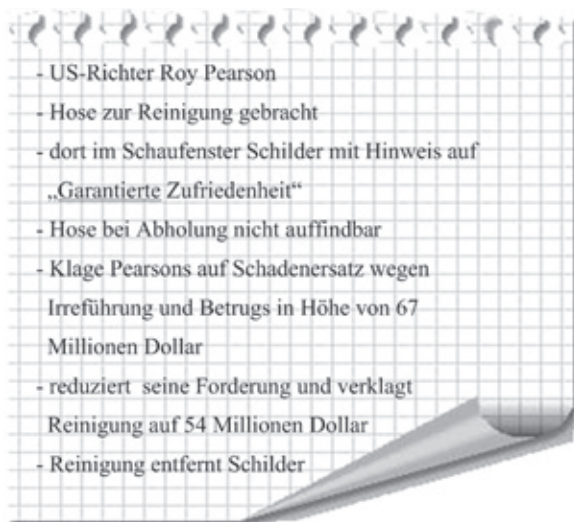
Kompetenzstufenmodelle für die Bereiche Fachwissen und Erkenntnisgewinnung in den naturwissenschaftlichen Fächern

In den naturwissenschaftlichen Fächern Biologie, Chemie und Physik fielen sämtliche Arbeitsschritte – von der Aufgabenkonstruktion bis zur Entwicklung von Kompetenzstufenmodellen – in den aktuellen Berichtszeitraum (siehe Abbildung 2.6). Die Aufgabenentwicklung erfolgte in jedem der drei Fächer für die vier Kompetenzbereiche *Fachwissen*, *Erkenntnisgewinnung*, *Bewertung* und *Kommunikation*.

Das fächerübergreifende Projekt „Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe I“ (ESNaS) nahm 2007 seine Arbeit auf. Zunächst bestand die Aufgabe der Projektleitung am IQB darin, für die drei beteiligten Fächer gemeinsame Konzeptionen und Arbeitsstrukturen als Grundlage für die Aufgabenentwicklung festzulegen. So wurden zum Beispiel fächerübergreifende Richtlinien zur Konstruktion von Testaufgaben erarbeitet, die Anleitungen zur Aufgabenentwicklung sowie fachspezifische Aufgabenbeispiele enthielten.

■ Abbildung 2.5: Aufgabenbeispiel für den Kompetenzbereich Schreiben im Fach Deutsch in der Sekundarstufe I

Der Reporter einer Tageszeitung soll für die morgige Ausgabe in der Rubrik „Aus aller Welt“ eine Nachricht schreiben. Dazu hat er sich folgende Stichpunkte notiert:



Schreibe die Nachricht für den Reporter und beachte dabei, dass in einem Artikel am Anfang die wichtigsten Informationen kommen und gegen Ende die unwichtigeren. Denk auch an die Überschrift!

Tipp:
Nummeriere zuerst die Informationen auf dem Notizzettel nach ihrer Wichtigkeit und schreibe dann.

Bevor mit der Entwicklung von Testaufgaben zur Evaluation des Erreichens der Bildungsstandards in den Naturwissenschaften begonnen wurde, entwickelte die fachdidaktische Leitung ein fächerübergreifendes Kompetenzstrukturmodell für die drei Fächer (Walpuski, Kampa, Kauertz & Wellnitz, 2008), das eine dreidimensionale Struktur mit den Dimensionen *Kompetenzbereiche*, *Komplexität des Inhalts* und *Kognitive Prozesse* aufweist (vgl. Abbildung 2.7). Im Rahmen der Aufgabenentwicklung wurden Anforderungen, von denen theoretisch anzunehmen ist, dass sie die Schwierigkeit von Aufgaben bestimmen, für die Dimensionen *Komplexität des Inhalts* und *Kognitive Prozesse* ausdifferenziert und gezielt variiert. Bei der Dimension *Komplexität des Inhalts* handelt es sich um ein gestuftes Aufgabenmerkmal, das den Umfang und den Vernetzungsgrad der zu bearbeitenden Inhalte beschreibt. Es werden die Ausprägungen „ein Fakt“, „zwei Fakten“, „ein Zusammenhang“, „zwei Zusammenhänge“ und „übergeordnetes Konzept“ unterschieden. Der in den Anforderungsbereichen der Bildungsstandards enthaltene Aspekt der *Komplexität* kann auf diese Weise genauer beschrieben und operationalisiert werden. Die Dimension *Kognitive Prozesse* spezifiziert hingegen die Qualität der kognitiven Fähigkeiten, die Schülerinnen und Schüler zur Bearbeitung einer Aufgabe benötigen. Dabei handelt es sich ebenfalls um ein gestuftes Aufgabenmerkmal, d. h. die Schülerinnen und Schüler müssen – je nach Niveau der Anforderung – zur Lösung einer Aufgabe Informationen reproduzieren, selektieren, organisieren oder integrieren.

Auf der Basis dieser inhaltlichen Vorüberlegungen begann im Jahr 2008 der Prozess der Testaufgabenentwicklung für die Kompetenzbereiche *Fachwissen* und *Erkenntnisgewinnung*. Die Aufgabenentwicklung für die Kompetenzbereiche *Bewertung* und *Kommunikation* startete zeitlich versetzt im Jahr 2009 bzw. 2011.

Der Kompetenzbereich *Fachwissen* erwies sich im Verlauf der Aufgabenentwicklung als besondere Herausforderung. Im Fokus dieses Kompetenzbereichs steht nicht der Wissensabruf, sondern – entsprechend des Kompetenzbegriffs von Weinert (2001) – der aktive Umgang mit fachwissenschaftlichen Inhalten zum Lösen spezifischer naturwissenschaftlicher Probleme. Die in den Bildungsstandards formulierten Kompetenzerwartungen im Bereich *Fachwissen* beschreiben die Fachinhalte nur allgemein und gleichzeitig ist die inhaltliche Schnittmenge der curricularen Vorgaben der 16 Länder gering. Aufgrund dieser Rahmenbedingungen war es erforderlich, die zur Bearbeitung der jeweiligen Testaufgaben notwendigen Fachinformationen im Aufgabenstamm anzugeben. Der Kompetenzbereich *Fachwissen* wird folglich primär als „Umgang mit Fachwissen“ verstanden.

Die in den naturwissenschaftlichen Fächern entwickelten Testaufgaben bilden gemeinsam mit den empirischen Ergebnissen der Normierungsstudie die Grundlage für die Erarbeitung der insgesamt sechs Kompetenzstufenmodelle für die Kompetenzbereiche *Umgang mit Fachwissen* und *Erkenntnisgewinnung* in den drei Fächern. Im Rahmen des *Standard-Settings* wurden die Stufengrenzen (*Cut-Scores*) zwischen den einzelnen Kompetenzstufen gesetzt und die Kompetenzstufen inhaltlich beschrieben.

■ **Abbildung 2.6: Entwicklung von Kompetenzstufenmodellen in den Fächern Biologie, Chemie und Physik in der Sekundarstufe I**



Die Festlegung der *Cut-Scores* erfolgte mit Hilfe der *Bookmark-Methode* (Mitzel, Lewis, Patz & Green, 2001), die in vergleichbaren internationalen Studien häufig eingesetzt wird (Pant, Tiffin-Richards & Köller, 2010; Tiffin-Richards & Köller, 2010). Bei der Bookmark-Methode arbeitet eine Expertengruppe mit einem sogenannten *Ordered-Item-Booklet*, in dem eine Auswahl von Testaufgaben nach ihrer empirischen Schwierigkeit geordnet vorliegt. Die Testaufgaben werden dabei so gewählt, dass sie repräsentativ für die Gesamtheit aller entwickelten Aufgaben sind, diese also hinsichtlich wichtiger Aufgabenmerkmale, wie Antwortformat, Komplexität oder kognitivem Prozess, abbilden.

Die Entwicklung der Kompetenzstufenmodelle für die naturwissenschaftlichen Fächer erfolgte in einem iterativen Prozess. Die Aufgaben wurden von Expertinnen und Experten von der leichtesten bis zur schwersten Testaufgabe mit dem Ziel betrachtet, Merkmale zu identifizieren, die ihre Schwierigkeit bestimmen (sog. schwierigkeitsgenerierende Merkmale) und die qualitative Übergänge zwischen Kompetenzstufen kennzeichnen. Die *Cut-Scores* zwischen den Stufen setzten die einzelnen Expertinnen und Experten individuell und diskutierten diese anschließend unter erneuter Bezugnahme auf die Testaufgaben. Im nächsten Schritt erfolgte eine konsensuale Setzung der *Cut-Scores* innerhalb der Expertengruppe. Die Stufengrenzen wurden auf Basis dieser Überlegungen an jenen Stellen gesetzt, an denen eine qualitative, fachdidaktisch begründete Zunahme in den kognitiven Anforderungen sichtbar wird, die zur Lösung der Aufgabe

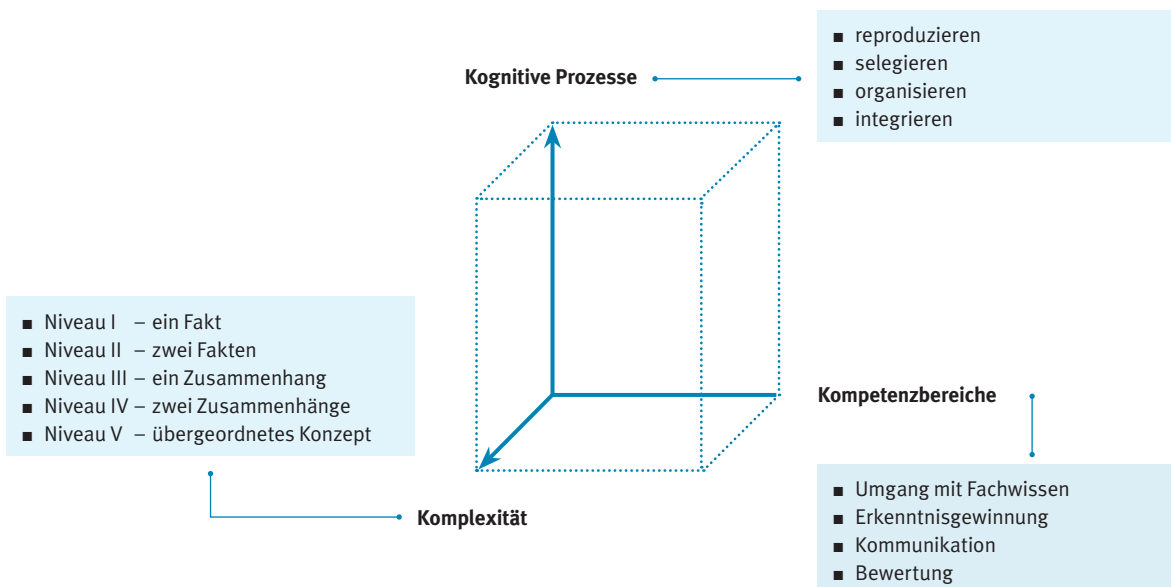
benötigt werden. Parallel zur Bestimmung der *Cut-Scores* wurden Beschreibungen der einzelnen Kompetenzstufen (Deskriptoren) in Form von Könnensbeschreibungen (*Can-do-Statements*) verfasst.

Die Verabschiedung der sechs bisher erarbeiteten Kompetenzstufenmodelle für die naturwissenschaftlichen Fächer durch die KMK erfolgte 2011. Im Herbst 2013 wurden die Modelle im Rahmen der Berichtslegung für den IQB-Ländervergleich 2012 genutzt, um zu bestimmen, inwieweit die Bildungsstandards für den MSA in den Fächern Biologie, Chemie und Physik in den Ländern erreicht werden.

Entwicklung von Bildungsstandards für die Allgemeine Hochschulreife

Das IQB wurde im Oktober 2007 von der KMK damit beauftragt, in Zusammenarbeit mit der Amtschefskommission „Qualitätssicherung in Schulen“ Bildungsstandards für die Allgemeine Hochschulreife in den Fächern Deutsch, Mathematik und fortgeführte Fremdsprache (Englisch, Französisch) zu entwickeln. Die Standardentwicklung für die Fächer Biologie, Chemie und Physik soll laut Beschluss der KMK zu einem späteren Zeitpunkt folgen. Damit erhielt das IQB erstmals den Auftrag, den Prozess der Standardentwicklung zu koordinieren – die Erarbeitung der Bildungsstandards für die Primarstufe und die Sekundarstufe I war bereits vor der Gründung des IQB durch Fachkommissionen erfolgt.

■ **Abbildung 2.7: Dreidimensionales Kompetenzstrukturmodell für die naturwissenschaftlichen Fächer** (siehe Walpuski, Kampa, Kauertz & Wellnitz, 2008)



Im Folgenden werden die Arbeitsschritte, die im IQB zur Entwicklung der Bildungsstandards für die Allgemeine Hochschulreife unternommen wurden, beschrieben. Anschließend werden die Entwicklung illustrierender Aufgaben, die zur Konkretisierung der Bildungsstandards dienen sollen, und die Erarbeitung der Hinweise zur Prüfungsdurchführung zum Erwerb der Allgemeinen Hochschulreife dargestellt. Diese drei Elemente – Bildungsstandards, illustrierende Aufgaben und Hinweise zur Prüfungsdurchführung zum Erwerb der Allgemeinen Hochschulreife – sind in den Dokumenten enthalten, die im Herbst 2012 von der KMK verabschiedet und der Öffentlichkeit vorgestellt worden sind (siehe <http://www.kmk.org/dokumentation/veroeffentlichungen-beschluesse/bildung-schule/qualitaetssicherung-in-schulen.html>).

Konzeption der Bildungsstandards für die Allgemeine Hochschulreife

Die Entwicklung der Bildungsstandards für die Allgemeine Hochschulreife begann im Dezember 2009 und wurde von einer Steuerungsgruppe begleitet, die sich aus Vertreterinnen und Vertretern der Bildungsministerien ausgewählter Länder zusammensetzte und von der Vorsitzenden des Schulausschusses der KMK, Senatsdirektorin Cornelia von Ilse (Bremen), geleitet wurde. Die mit der Erarbeitung der Bildungsstandards betrauten Arbeitsgruppen in den Fächern Deutsch, Mathematik, Englisch und Französisch setzten sich zusammen aus

- einer an das IQB abgeordneten Lehrkraft, die für die Koordination der Arbeiten im jeweiligen Fach verantwortlich war,
- wissenschaftlichen Mitarbeiterinnen und Mitarbeitern des IQB,
- Vertreterinnen und Vertretern von Schulen, Landesinstituten und Bildungsministerien der Länder sowie
- Professorinnen und Professoren der entsprechenden Fachdidaktiken.

Zwischen Februar 2010 und März 2012 wurden in jeweils etwa zwölf mehrtägigen Arbeitstreffen und in mehreren Rückmeldungs- und Überarbeitungsschleifen die in Auftrag gegebenen Bildungsstandards für die Allgemeine Hochschulreife von den Arbeitsgruppen konzipiert. Die Arbeitsgruppen für Englisch und Französisch schlossen sich zusammen und entwickelten gemeinsame „Bildungsstandards für die fortgeführte Fremdsprache“.

Bei der Konzeption der Bildungsstandards für die Allgemeine Hochschulreife sahen sich die Arbeitsgruppen am IQB vor die Herausforderung gestellt, die Anschlussfähigkeit der zu formulierenden Kompetenzerwartungen an die Bildungsstandards für den Mittleren Schulabschluss sicherzustellen. Grundsätzlich konnten die Kompetenzstrukturmodelle der Sekundarstufe I übernommen werden, wobei jedoch einige Anpassungen vorgenommen wurden, die für die Allgemeine Hochschulreife sinnvoll erschienen (siehe Abbildungen 2.8 bis 2.10). So wurde im Fach Deutsch die Struktur der Bildungsstandards für die Sekundarstufe I dahingehend erweitert, dass die Kompetenzbereiche *Sich mit Texten und Medien auseinandersetzen* und *Sprache und Sprachgebrauch reflektieren* separate Dimensionen bilden und somit ein eigenes Gewicht erhalten, was den Zielsetzungen der gymnasialen Oberstufe entspricht. Auch für die Fremdsprachen wird die *Text- und Medienkompetenz* als eigenständiger Bereich ausgewiesen. Darüber hinaus wurden in den Fremdsprachen die *Sprachbewusstheit* und die *Sprachlernkompetenz*, die substantiell an allen anderen Kompetenzen beteiligt sind, als eigene Kompetenzbereiche konzipiert. Im Fach Mathematik schließlich erfolgte eine Erweiterung der Leitideen um oberstufengerechte Inhalte. Zusätzlich zu diesen Anpassungen mussten die allgemeinen Ziele der gymnasialen Oberstufe (Vertiefung der Allgemeinbildung, Entwicklung allgemeiner Studierfähigkeit, wissenschaftspropädeutische Bildung) in Kompetenzerwartungen übersetzt werden, die sich von denen des Mittleren Schulabschlusses ausreichend abheben, ohne sie jedoch so hoch anzusetzen, dass sie im Rahmen von Regelstandards nicht mehr einlösbar wären.

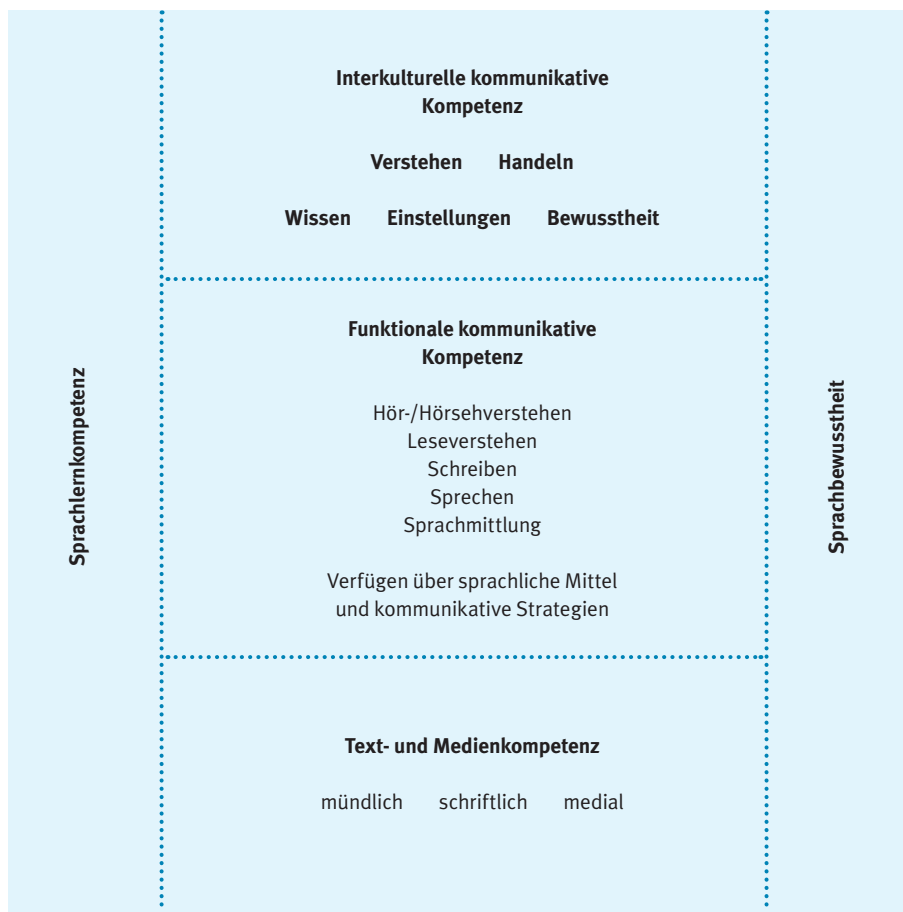
■ **Abbildung 2.8: Den Bildungsstandards für die Allgemeine Hochschulreife im Fach Deutsch zugrunde liegendes Kompetenzstrukturmodell**

Domänenspezifischer Kompetenzbereich	Prozessbezogene Kompetenzbereiche	Domänenspezifischer Kompetenzbereich
Sich mit Texten und Medien auseinandersetzen	Sprechen und Zuhören	Sprache und Sprachgebrauch reflektieren
	Schreiben	
	Lesen	

Weiterhin war durch die Teams der Standardentwickler zu prüfen, inwieweit die in den Einheitlichen Prüfungsanforderungen für die Abiturprüfung (EPA) beschriebenen Leistungserwartungen (KMK, 2002) Eingang in die neuen Abiturstandards finden konnten. Die EPA bildeten bis zur Verabschiedung der Bildungsstandards für die Allgemeine Hochschulreife den länderübergreifenden Referenzrahmen für die Ziele der gymnasialen Oberstufe und die Entwicklung von Aufgaben für die Abiturprüfung. Da die EPA in ihren letzten Fassungen von 2002 (Deutsch, Englisch, Mathematik) und 2004 (Französisch) bereits eine zumindest implizite Kompetenzorientierung aufwiesen, war es in vielen Fällen möglich, Leistungserwartungen der EPA zu Bildungsstandards für die Allgemeine Hochschulreife weiterzuentwickeln. Aus den EPA wurde auch die Differenzierung der Bildungsstandards in Bezug auf ein grundlegendes und ein erhöhtes Niveau übernommen, wobei das grundlegende Niveau mit mindestens drei, das erhöhte Niveau mit vier oder mehr

Wochenstunden Unterricht erreicht werden soll. Für die meisten Kompetenzbereiche konnten sowohl Standards für das grundlegende Niveau als auch Standards für das erhöhte Niveau formuliert werden, in einzelnen Fällen war dies jedoch nicht sinnvoll möglich. So ließ sich im Fach Deutsch für die prozessbezogenen Kompetenzbereiche *Sprechen und Zuhören*, *Schreiben* und *Lesen* keine Differenzierung vornehmen, da sie im Sinne der propädeutischen Zielsetzung der gymnasialen Oberstufe für alle Schülerinnen und Schüler in gleichem Maße unverzichtbar sind. Unterschiedliche Anforderungen ergeben sich für diese Kompetenzbereiche erst in ihrer Verschränkung mit den spezifischen Domänen *Sich mit Texten und Medien auseinandersetzen* oder *Sprache und Sprachgebrauch reflektieren*, die hinsichtlich der beiden Anforderungsniveaus differenziert wurden. Ähnliches gilt in den Fremdsprachen für die *Interkulturelle kommunikative Kompetenz*, die ebenfalls keine sinnvolle Niveaudifferenzierung zulässt.

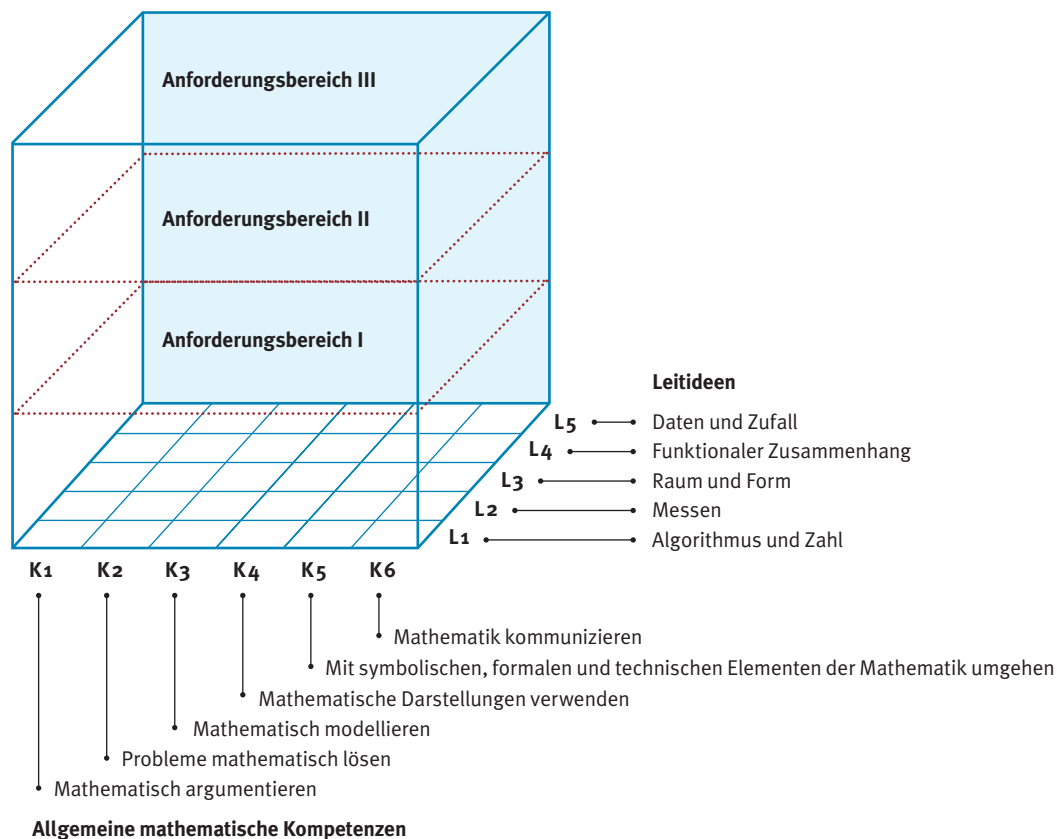
■ **Abbildung 2.9: Den Bildungsstandards für die Allgemeine Hochschulreife in den fortgeführten Fremdsprachen zugrunde liegendes Kompetenzstrukturmodell**



Im Ergebnis beschreiben die Bildungsstandards für die Allgemeine Hochschulreife in Form von Könnensbeschreibungen (*Can-do-Statements*) Kompetenzen, über die Schülerinnen und Schüler zum Zeitpunkt des Erwerbs der Allgemeinen Hochschulreife in der Regel verfügen sollen. Die Zielvorgaben wurden inhaltlich so konzipiert, dass sie gleichermaßen für alle Bildungsgänge gelten, die zur Allgemeinen Hochschulreife führen, also auch für berufliche Gymnasien und doppelqualifizierende Bildungsgänge. Lediglich die Berufsoberschulen (BOS) blieben aufgrund ihres besonderen Profils bei der Entwicklung der Bildungsstandards zunächst unberücksichtigt. Im Zusammenhang mit ihrer Implementierung soll jedoch geklärt werden, welche der Bildungsstandards auch für die BOS angemessen sind und welche veränderten bzw. zusätzlichen Zielvorgaben für diese Schulform definiert werden sollten.

Die Länder erhielten mehrfach Gelegenheit, Entwürfe der Bildungsstandards für die Allgemeine Hochschulreife zu kommentieren. Weiterhin wurden Ende 2011 vorläufige Fassungen der Dokumente im Rahmen einer gemeinsamen Fachtagung von IQB und KMK erstmals einer breiten Fachöffentlichkeit vorgestellt. Die zahlreichen Rückmeldungen der Länder und der Verbände, die bei der Anhörung vertreten waren, wurden in Synopsen dokumentiert und flossen in die Weiterentwicklung der Bildungsstandards ein. Im Juni 2012 konnte der Prozess der Entwicklung von Bildungsstandards für die Allgemeine Hochschulreife abgeschlossen werden.

■ **Abbildung 2.10: Den Bildungsstandards für die Allgemeine Hochschulreife im Fach Mathematik zugrunde liegendes Kompetenzstrukturmodell**



Entwicklung illustrierender Aufgaben

Die Bildungsstandards für die Allgemeine Hochschulreife werden mit zwei verschiedenen Arten von Beispielaufgaben illustriert. Mit *illustrierenden Abiturprüfungsaufgaben* soll ein erster Eindruck davon vermittelt werden, wie die in den Bildungsstandards definierten Anforderungen im Abitur geprüft werden könnten. Die Entwicklung dieser standardbasierten Prüfungsaufgaben für das Abitur war mit einer Reihe von Herausforderungen verbunden, da der aktuelle Diskussionsstand in der Fachdidaktik, unterschiedliche Ländertraditionen sowie der Anspruch auf Vergleichbarkeit der Abschlüsse in den Ländern miteinander in Einklang gebracht werden mussten. *Illustrierende Lernaufgaben* hingegen sollen exemplarisch verdeutlichen, welche Aufgabenstellungen dazu geeignet sein können, die jeweiligen Kompetenzen bei Schülerinnen und Schülern im Unterricht zu entwickeln. Lernaufgaben dienen nicht der Überprüfung des Erreichens von Bildungsstandards, sondern sie sollen aktive Lernprozesse anstoßen und diese durch eine Folge von gestuften Aufgabenstellungen steuern, um bei den Schülerinnen und Schülern bestimmte Kompetenzen auszubilden, zu konsolidieren oder zu vertiefen. Die illustrierenden Lernaufgaben unterscheiden sich im Hinblick auf ihr Anforderungsniveau und die gewählten Aufgabenformate deutlich voneinander, abhängig davon, welches Lernziel konkret verfolgt wird und zu welchem Zeitpunkt im Prozess der Entwicklung der jeweiligen Kompetenz die jeweilige Aufgabe zum Einsatz kommen. Während bei der Konstruktion von Lernaufgaben in der Regel von einzelnen Standards ausgegangen wurde, deren Förderung illustriert werden sollte, sind die Abiturprüfungsaufgaben so angelegt, dass damit mehrere Kompetenzbereiche und eine Vielzahl von Einzelkompetenzen vernetzt geprüft werden.

Im September 2010 begann die Entwicklung der illustrierenden Aufgaben mit Auftaktworkshops in den Fächern Deutsch, Englisch, Französisch und Mathematik. Unter Anleitung des IQB und in Zusammenarbeit mit fachdidaktischen Expertinnen und Experten fanden bis zum Frühjahr 2012 jeweils ca. zwölf mehrtägige Arbeitstreffen der Aufgabenentwicklerinnen und -entwickler statt. An diesem Prozess waren Lehrkräfte aus allgemein bildenden und beruflichen Gymnasien beteiligt. Um anhand authentischer Schülerlösungen die neu entwickelten Aufgaben optimieren zu können, wurden die Lern- und Abiturprüfungsaufgaben informell erprobt. Für die Abituraufgaben fand im März 2012 zudem eine Prüflerlesung durch insgesamt acht Expertinnen und Experten aus unterschiedlichen Ländern statt.

Zusätzlich zu den illustrierenden Aufgaben, die in den Endfassungen der Bildungsstandards für die Allgemeine Hochschulreife enthalten sind, wurden unter Federführung des IQB im Laufe des Schuljahres 2012/2013 weitere Lernaufgaben für fachspezifische Publikationen erarbeitet. In diesen Veröffentlichungen, die 2014 erschienen sind oder erscheinen sollen, werden die einzelnen Kompetenzbereiche zunächst wissenschaftlich fundiert beschrieben und der aktuelle Forschungs- und Diskussionsstand zur Entwicklung der Kompetenzen, zu zentralen Lerngegenständen und zu geeigneten Vermittlungsformen zusammengefasst. Daran anknüpfend werden Lernaufgaben präsentiert und fachdidaktisch kommentiert.

Entwicklung der Hinweise zur Prüfungsdurchführung zum Erwerb der Allgemeinen Hochschulreife

Parallel zur Standard- und Aufgabenentwicklung nahm eine weitere Arbeitsgruppe im Herbst 2011 ihre Arbeit auf, die damit betraut worden war, die Hinweise zur Durchführung der Abiturprüfung aus den EPA weiterzuentwickeln. Sie setzte sich aus Mitgliedern der Arbeitsgruppe „Gymnasiale Oberstufe“ der KMK zusammen. Die Koordinatorinnen und Koordinatoren des IQB erarbeiteten Entwürfe für die Vorgaben und nahmen an den Sitzungen der Gruppe teil. Die aus diesem Arbeitsprozess resultierenden „Hinweise zur Prüfungsdurchführung zum Erwerb der Allgemeinen Hochschulreife“ wurden in die Publikation der Bildungsstandards integriert und ersetzen die Prüfungshinweise der EPA. Bei den Veränderungen handelt es sich nicht um grundlegende Veränderungen des Bestehenden, sondern um eine Weiterentwicklung und Systematisierung der bereits in den EPA angebahnten Entwicklungen. Einzelne Neuerungen betreffen unter anderem

- im Fach Deutsch die Stärkung des propädeutischen Charakters des Unterrichts, u. a. durch materialgestütztes Schreiben als Aufgabenformat (Verfassen informierender und argumentierender Texte auf der Grundlage umfangreichen Textmaterials),
- im Fach Mathematik die Stärkung des Stellenwerts von Stochastik, u. a. durch die Vorgabe, dass keines der drei Sachgebiete (lineare Algebra, Analysis, Stochastik) über mehrere Jahre von den Prüfungsaufgaben ausgeschlossen werden darf, und
- in den Fremdsprachen die Stärkung des Mündlichen durch die verpflichtende Überprüfung des Sprechens oder Hörverstehens (in Abitur- oder Klausurprüfung).

Konzeption und Entwicklung eines Pools von Abiturprüfungsaufgaben

Ab dem Schuljahr 2016/17 sollen die Abiturprüfungen aller Länder auf den Bildungsstandards für die Allgemeine Hochschulreife basieren. Den hierfür notwendigen Implementationsprozess unterstützt das IQB zunächst mit der Entwicklung illustrierender Lernaufgaben (s. o.). Darüber hinaus wurde das IQB mit Beschluss der 342. Kultusministerkonferenz vom 20./21. Juni 2013 damit beauftragt, einen Pool von Abiturprüfungsaufgaben zu erstellen, der auf den Bildungsstandards für die Allgemeine Hochschulreife basiert. Damit wird das Ziel verfolgt, die Vergleichbarkeit der Anforderungen in den Abiturprüfungen der Länder zu gewährleisten sowie die Qualität und den Standardbezug der Aufgaben zu sichern. Der Aufgabenpool soll ab 2013 kontinuierlich aufwachsen und den Ländern ab dem Schuljahr 2016/2017 für den möglichen Einsatz im Abitur zur Verfügung stehen. Die Entwicklung und Pflege des Pools wird in den nächsten Jahren zu den zentralen Aufgaben des IQB gehören.

An diesem Arbeitsprozess sind Arbeitsgruppen (AGs) auf zwei Ebenen beteiligt:

Die AG *Abiturkommission* setzt sich aus Vertreterinnen und Vertretern der Länder und des IQB zusammen und ist für Grundsatzentscheidungen zuständig, die den Rahmen für

die Entwicklung und Nutzung des Aufgabenpools abstecken (z. B. Umgang mit Operatorenlisten, Formate der Erwartungshorizonte, allgemeine Kriterien für die Aufgabenauswahl).

Die AGs *Aufgabenbewertung*, in denen Fachexpertinnen und -experten der Länder und des IQB sowie Fachdidaktikerinnen und Fachdidaktiker zusammenarbeiten, spezifizieren die Kriterien für die Aufgabenauswahl in den einzelnen Fächern und wenden diese anschließend an. Die Kriterien beziehen sich insbesondere auf den Standardbezug der Aufgaben und die Angemessenheit ihres Anforderungsniveaus. Dabei werden sowohl die Aufgabenstellungen als auch die Erwartungshorizonte und die Vorgaben für die Korrektur und Bewertung der Schülerantworten in den Blick genommen.

Die Länder haben bereits damit begonnen, vorliegende Abiturprüfungsaufgaben beim IQB einzureichen. Diese werden vor Aufnahme in den Pool von den AGs *Aufgabenbewertung* anhand der Kriterien geprüft. Ein Teil der ausgewählten Aufgaben wird 2015 in Form einer Sammlung veröffentlicht. Diese soll insbesondere Lehrkräften einen Eindruck von den Aufgabenformaten vermitteln, die im Pool enthalten sind, und zur Implementation der Bildungsstandards für die Allgemeine Hochschulreife beitragen.



Literatur

- Blum, W., Drike-Noe, C., Leiß, D., Wiegand, B. & Jordan, A. (2005). Zur Rolle von Bildungsstandards für die Qualitätsentwicklung im Mathematikunterricht. *Zentralblatt für Didaktik der Mathematik*, 37, 267–274.
- Bremerich-Vos, A. (2009). Bildungsstandards im Fach Deutsch. Was leisten sie, was nicht? *Pädagogik*, 61, 42–45.
- Kingston, N. M., Kahl, S. R., Sweeney, K. P. & Bay, L. (2001). Setting performance standards using the body of work method. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 219–248). Mahwah, NJ: Lawrence Erlbaum.
- Klieme, E. (2004). Zur Entwicklung nationaler Bildungsstandards. Grundpositionen einer Expertise. In T. Fitzner (Hrsg.), *Bildungsstandards. Internationale Erfahrungen – Schulentwicklung – Bildungsreform* (S. 256–265). Bad Boll: Evangelische Akademie.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (Hrsg.). (2007). *Zur Entwicklung nationaler Bildungsstandards. Expertise*. Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- KMK (2002) = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2002). *Vereinbarung über Einheitliche Prüfungsanforderungen in der Abiturprüfung*. (Beschluss vom 1. 6. 1979 i. d. F. vom 24.05.2002). Neuwied: Luchterhand.
- Köller, O. (2008). Bildungsstandards – Verfahren und Kriterien bei der Entwicklung von Messinstrumenten. *Zeitschrift für Pädagogik*, 54(2), 163–173.
- Köller, O. (2009). Bildungsstandards. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (2. Auflage, S. 529–548). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mitzel, H. C., Lewis, D. M., Patz, R. J. & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (pp. 249–281). Mahwah, NJ: Erlbaum.
- Pant, H. A., Tiffin-Richards, S. P. & Köller, O. (2010). Standard-Setting für Kompetenztests im Large-Scale-Assessment. 56. *Beiheft der Zeitschrift für Pädagogik*, 175–188.
- Tiffin-Richards, S. P. & Köller, O. (2010). Comparison and synthesis of multiple standard-setting methods and panels. In C. Harsch, H. A. Pant & O. Köller (Eds.), *Calibrating standard-based assessment tasks for english as a first foreign language – Standard-setting procedures in Germany* [Volume 2] (pp. 107–112). Münster: Waxmann.
- Walpuski, M., Kampa, N., Kauertz, A. & Wellnitz, N. (2008). Evaluation der Bildungsstandards in den Naturwissenschaften. *MNU*, 61(6), 323–326.
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – Eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 17–31). Weinheim: Beltz.



03 // Ländervergleichsstudien

Zentrale Überprüfung des Erreichens der Bildungsstandards

Eine der Hauptaufgaben des IQB besteht darin, regelmäßig in Ländervergleichsstudien zu überprüfen, inwieweit die mit den Bildungsstandards definierten Kompetenzziele in den Ländern erreicht werden. Seit Bestehen des IQB wurden bislang drei Ländervergleichsstudien durchgeführt. Ihre Konzeption, Durchführung und Berichtslegung werden in diesem Kapitel vorgestellt.

Von nationalen Ergänzungsstudien zu Ländervergleichen

Seit fast zwei Jahrzehnten nimmt die Bundesrepublik Deutschland im Bereich der Sekundarstufe I und im Primarbereich an internationalen Vergleichsstudien teil (vgl. KMK, 2006). Allgemeines Ziel dieser Studien ist es, den teilnehmenden Staaten Indikatoren über die Fähigkeiten, Kenntnisse und Fertigkeiten ihrer Schülerinnen und Schüler im internationalen Vergleich zur Verfügung zu stellen und gleichzeitig Aussagen zu Entwicklungstrends zu ermöglichen, die auf der Ebene der schulischen Systeme beobachtbar sind.

In der Bundesrepublik Deutschland wurden erstmals mit TIMSS¹ 1995 (Baumert, Bos & Lehmann, 2000a, 2000b; Baumert et al., 1997) die mathematischen und naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern zum Ende der Pflichtschulzeit im internationalen Vergleich überprüft. Die internationale Schulleistungsuntersuchung PISA² (Baumert et al., 2001; Klieme et al., 2010), an der sich auch Deutschland seit 2000 beteiligt, wird im 3-Jahres-Rhythmus bei 15-jährigen Schülerinnen und Schülern mit wechselnden inhaltlichen Schwerpunkten durchgeführt: Im Jahr 2000 stand die Lesekompetenz im Zentrum, 2003 die mathematische Grundbildung und 2006 die Kompetenz in den Naturwissenschaften. Mit PISA 2009 begann dieser Zyklus zum zweiten Mal.

In der Primarstufe erfasste 2001 und 2006 PIRLS/IGLU³ (Bos, Tarelli, Bremerich-Vos & Schwippert, 2012) international vergleichend die Leseleistungen von Schülerinnen und Schülern und die Rahmenbedingungen des Lesenlernens in Schulen und Elternhäusern. In den TIMSS-Untersuchungen 2007 und 2011 (Bos, Wendt, Köller & Selter, 2012) standen mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern im Zentrum des internationalen Vergleichs.

Bis 2006 wurden zeitgleich zu PISA und IGLU repräsentative Stichproben in den Ländern der Bundesrepublik Deutschland erhoben, um einen Ländervergleich zu ermöglichen und auf Länderebene mögliche Einflussfaktoren auf die Kompetenzstände der Schülerinnen und Schüler zu analysieren, wie beispielsweise eine länderspezifische Lehrplangestaltung. Während sich an PISA-E in den Jahren 2000, 2003 und 2006 alle Länder in Deutschland beteiligten, nahmen an der Grundschulstudie IGLU-E 2001 nur die Länder Baden-Württemberg, Bayern, Brandenburg, Bremen, Hessen, Nordrhein-

Westfalen und Thüringen teil, an IGLU-E 2006 dann erstmalig alle 16 Länder. Einen innerdeutschen Ländervergleich der mathematischen und naturwissenschaftlichen Kompetenzen von Grundschulkindern – analog zu IGLUE-E – hat es in Deutschland als Ergänzung zu den internationalen Studien nicht gegeben.

Im Zuge der KMK-Gesamtstrategie zum Bildungsmonitoring im allgemein bildenden Schulsystem (KMK, 2006) wurde festgelegt, dass Ländervergleiche zukünftig auf der Grundlage der 2004/2005 verabschiedeten Bildungsstandards (siehe Kapitel 2) vom IQB durchgeführt werden und damit die bisherigen Ländervergleiche im Rahmen von IGLU und PISA ablösen. Die Ländervergleichsstudien erfolgen damit seit dem Jahr 2009 nicht mehr in Anbindung an international entwickelte Tests, sondern auf der Grundlage von Testverfahren, die gezielt die länderübergreifenden Bildungsstandards ins Zentrum der Überprüfungen rücken. Damit können fachdidaktische Traditionen des deutschen Systems stärker als bisher berücksichtigt und beispielsweise die Naturwissenschaften nach den Fächern Biologie, Chemie und Physik getrennt betrachtet werden. Bei der Testung im Fach Deutsch werden neben dem bisher in PISA-E und IGLU-E erfassten *Leseverständnis* auch *Hörverständnis* und *Rechtschreibung* als weitere Kompetenzbereiche überprüft. Zudem können weitere Kernfächer berücksichtigt werden, die in den internationalen Studien nicht getestet werden. So wurden 2009 erstmals die fremdsprachlichen Kompetenzen in den Fächern Englisch und Französisch als erster Fremdsprache erfasst. Die bildungsstandardbasierten Ländervergleichsstudien werden zeitgleich mit den internationalen Erhebungen im Bereich der Primarstufe alle fünf Jahre, im Bereich der Sekundarstufe I alle drei Jahre, alternierend für die Fächergruppen Deutsch, Englisch und Französisch einerseits und Mathematik und die Naturwissenschaften andererseits, durchgeführt. Tabelle 3.1 gibt einen Überblick über die nationalen sowie die internationalen Erhebungen mit deutscher Beteiligung im Berichtszeitraum.

Mit der Einführung eines auf den Bildungsstandards basierenden Monitorings in Deutschland ist der Vorteil verbunden, dass es auf länderübergreifend verbindlichen Zielvorgaben beruht, die sich die bildungspolitisch Verantwortlichen aller 16 Länder selbst gesetzt haben. Diese Zielvorgaben sind – wie in Kapitel 2 beschrieben – das Resultat eines komplexen Verständigungsprozesses darüber, welche Kompetenzen

¹ Das Akronym TIMSS stand zu diesem Zeitpunkt für *Third International Mathematics and Science Study*. Seit 2003 wird es in der Bedeutung *Trends in International Mathematics and Science Study* verwendet.

² Das Akronym PISA steht für *Programme for International Student Assessment*.

³ Das Akronym PIRLS steht für *Progress in International Reading Literacy Study*; im deutschen Sprachraum wird diese Studie als *Internationale Grundschul-Lese-Untersuchung* (IGLU) bezeichnet.

Schülerinnen und Schüler bis zu einem bestimmten Zeitpunkt in der Bildungslaufbahn in der Regel entwickelt haben sollen. Sie werden in den Dokumenten der KMK beschrieben und – für ausgewählte Kompetenzbereiche – vom IQB in Testaufgaben übersetzt. Anhand von Kompetenzstufenmodellen (vgl. Kapitel 2), die unter der Federführung des IQB entwickelt werden, lassen sich Aussagen darüber treffen, inwieweit die Schülerinnen und Schüler in den einzelnen Ländern die gesetzten Ziele tatsächlich erreichen. Damit ist der Grad der Verbindlichkeit im nationalen Bildungsmonitoring deutlich höher als bei den internationalen Schulleistungsstudien.

Das IQB ist von den Ländern beauftragt, die Ländervergleiche kontinuierlich vorzubereiten, sie durchzuführen, auszuwerten und die Ergebnisse in geeigneter Form der Bildungspolitik, der Bildungsverwaltung, der pädagogischen Praxis und der Öffentlichkeit vorzustellen. Ähnlich wie in den internationalen Schulleistungsstudien werden in den Berichten zu den IQB-Ländervergleichen neben den erreichten Kompetenzniveaus auch die Koppelung der Leistung an weitere Faktoren, wie beispielsweise die soziale Herkunft, den Zuwanderungshintergrund oder das Geschlecht beschrieben. Die Berichterstattung zu den IQB-Ländervergleichsstudien erfolgt jeweils ungefähr eineinhalb Jahre nach der Datenerhebung.

Die IQB-Ländervergleichsstudien Wer und was wurde getestet? Zielpopulation und untersuchte Kompetenzbereiche

Die Bildungsstandards sind abschlussbezogen konzipiert und beziehen sich auf das Ende der Primarschulzeit beziehungsweise auf den Hauptschulabschluss (HSA) oder den Mittleren Schulabschluss (MSA) in der Sekundarstufe I. In den drei im Berichtszeitraum vom IQB durchgeführten Ländervergleichsstudien wurden daher Kompetenzen von Schülerinnen und Schülern am Ende der Primarschulzeit in der 4. Jahrgangsstufe⁴ bzw. zum Ende der Sekundarstufe I in der 9. Jahrgangsstufe erfasst. Hinsichtlich des gegliederten Schulsystems in der Sekundarstufe I gibt es zwischen den Ländern der Bundesrepublik Deutschland erhebliche Schulstrukturunterschiede. Zahlreiche Länder befinden sich derzeit in einem Reformprozess, der darauf abzielt, ein Zwei-Säulen-Modell zu etablieren, welches neben dem Gymnasium nur noch eine weitere nicht gymnasiale Schulart vorsieht (vgl. Hurrelmann, 2013; Tillmann, 2012). Diese weitere Schulart bietet in der Regel mehrere Bildungsgänge an bzw. ist als integrierte oder kooperative Gesamtschule organisiert. Die nicht gymnasialen Schularten bieten neben Haupt- und Realschulbildungsgängen teilweise auch einen Gymnasialzweig an, der mit dem Abitur nach 13 – mancherorts auch nach 12 – Schuljahren abgeschlossen werden kann (vgl. Pant et al., 2013). Trotz der grundsätzlichen Tendenz zur

Tabelle 3.1: Zeitpunkt der Datenerhebungen der internationalen Schulleistungsstudien und der nationalen Ländervergleichsstudien im Berichtszeitraum

	PIRLS	TIMSS	PISA	Ländervergleichsstudien auf Basis der Bildungsstandards	
				Primarstufe	Sekundarstufe I
2007		x			
2008					
2009			x		x (Deutsch, Englisch, Französisch*)
2010					
2011	x	x		x (Deutsch, Mathematik)	
2012			x		x (Mathematik, Naturwissenschaften)
2013					

Anmerkung zu den Akronymen: PIRLS = Progress in Reading Literacy Study, TIMSS = Trends in International Mathematics and Science Study, PISA = Programme for International Student Assessment

* Für das Fach Französisch fand die Datenerhebung bereits 2008 statt.

⁴ In zwei Ländern, Berlin und Brandenburg, endet die Primarstufe regulär nach der 6. Jahrgangsstufe.

Zweigliedrigkeit bleibt die Landschaft der Schularten in Deutschland heterogen. Da die Bildungsstandards abschlussbezogen formuliert sind, führen die Schulstruktur-reformen der zurückliegenden Jahre zu Problemen bei der Definition der Zielpopulation der IQB-Ländervergleiche in der Sekundarstufe I. Um dennoch einen sinnvollen Vergleich der von den Jugendlichen in den verschiedenen Ländern erbrachten Leistungen durchführen zu können, werden in den IQB-Ländervergleichen in der Sekundarstufe I Schülerinnen und Schüler der 9. Jahrgangsstufe aus allen Bildungsgängen des allgemein bildenden Schulsystems berücksichtigt. Im Ländervergleich 2009 umfasste dies Jugendliche aus Hauptschulen, Realschulen, Gymnasien und integrierten Gesamtschulen sowie Schülerinnen und Schüler aus Schulen mit mehreren Bildungsgängen; im IQB-Ländervergleich 2012 wurden zusätzlich Förderschulen einbezogen.

Im Jahr 2009 fand der erste IQB-Ländervergleich auf der Grundlage der KMK-Bildungsstandards für die Sekundarstufe I in den Fächern Deutsch, Englisch und Französisch statt. Im Fach Deutsch wurden zentrale Aspekte der Kompetenzbereiche *Zuhören*, *Lesen* und *Rechtschreibung* getestet, für die erste Fremdsprache Englisch und Französisch wurden Kompetenzen in den Bereichen *Hörverstehen* und *Leseverstehen* erfasst. Insgesamt nahmen im Fach Deutsch über 36 000, im Fach Englisch gut 31 000 und im Fach Französisch fast 6 000 Schülerinnen und Schüler der neunten Jahrgangsstufe teil.

Die zweite Ländervergleichsstudie des IQB im Jahr 2011 überprüfte erstmalig das Erreichen der Bildungsstandards in der Primarstufe in den Fächern Deutsch und Mathematik. Im Zentrum der Erhebungen standen im Fach Deutsch Kompetenzen in den Bereichen *Lesen* und *Zuhören*. Im Fach Mathematik wurden die fünf inhaltsbezogenen Kompetenzbereiche (Leitideen) *Zahlen und Operationen*, *Raum und Form*, *Muster und Strukturen*, *Größen und Messen* sowie *Daten, Häufigkeit und Wahrscheinlichkeit* erfasst. Die Tests wurden mit ca. 27 000 Schülerinnen und Schülern der vierten Jahrgangsstufe durchgeführt.

Der dritte Ländervergleich im Jahr 2012 ermittelte Kompetenzen von Schülerinnen und Schülern in der Sekundarstufe I im Fach Mathematik sowie in den naturwissenschaftlichen Fächern Biologie, Chemie und Physik. Im Fach Mathematik wurden – wie in der Primarstufe – die fünf Leitideen *Zahl*, *Messen*, *Raum und Form*, *Funktionaler Zusammenhang* sowie *Daten und Zufall* geprüft. In den naturwissenschaftlichen Fächern wurde der Schwerpunkt auf die beiden Kompetenzbereiche *Fachwissen* und *Erkenntnisgewinnung* gelegt, die für jedes der drei Fächer separat getestet wurden. Am IQB-Ländervergleich 2012 nahmen in Deutschland insge-

samt fast 45 000 Schülerinnen und Schüler der 9. Jahrgangsstufe teil.

Mit diesen drei Studien ist der erste Zyklus der IQB-Ländervergleiche abgeschlossen. Im Jahr 2015 wird der zweite Zyklus mit dem Ländervergleich sprachlicher Kompetenzen im Bereich der Sekundarstufe I beginnen, gefolgt von den Ländervergleichsstudien zur Überprüfung der Kompetenzen in den Fächern Deutsch und Mathematik in der Primarstufe im Jahr 2016 sowie den mathematischen und naturwissenschaftlichen Kompetenzen in der Sekundarstufe I im Jahr 2018. Nach jeder Studie des zweiten Zyklus wird es dann auch möglich sein, Trendaussagen zu treffen. Im Zentrum wird dabei die Frage stehen, inwieweit es in den einzelnen Ländern gelungen ist, den Anteil der Schülerinnen und Schüler, die selbst Mindeststandards verfehlen, zu verringern bzw. den Anteil der Schülerinnen und Schüler, deren Leistungen mindestens den Anforderungen der Regelstandards entsprechen, zu erhöhen.

Kompetenztests und Fragebögen

Um zu überprüfen, ob die Schülerinnen und Schüler die in den Bildungsstandards definierten Kompetenzen erreichen, wird in den Ländervergleichsstudien eine große Anzahl von standardbasierten Testaufgaben eingesetzt. Diese Aufgaben entstammen den Itempools des IQB, die von Aufgabenentwicklerteams mit Unterstützung von fachdidaktischen Kooperationspartnerinnen und -partnern entwickelt und anschließend vom IQB in großen nationalen Erhebungen pilotiert und normiert worden sind (siehe Kapitel 2).

In Ergänzung zu den Kompetenztests wurden die Schülerinnen und Schüler, Eltern (nur Ländervergleichsstudien 2009 und 2011), Lehrkräfte sowie Schulleiterinnen und Schulleiter jeweils gebeten, einen Fragebogen zu bearbeiten. Diese Instrumente dienen unter anderem der Erfassung individueller Hintergrundmerkmale von Schülerinnen und Schülern, um Leistungsdisparitäten zu untersuchen, die mit dem Geschlecht, dem Zuwanderungshintergrund oder der sozialen Herkunft verbunden sind. Weiterhin wurden in den Fragebögen Informationen zum fachlichen Interesse oder den fachlichen Selbstkonzepten der Schülerinnen und Schüler, zu Merkmalen des Unterrichts, zur Schule und zum familiären Umfeld erfasst, um Zusammenhänge zwischen der Lernmotivation und den zur Verfügung stehenden Lerngelegenheiten einerseits und den erreichten Kompetenzen andererseits analysieren zu können. Diese Hintergrundfragebögen wurden von Mitarbeiterinnen und Mitarbeitern des IQB neu entwickelt oder aus früheren Studien – teilweise in angepasster Form – übernommen.

Vom Itempool zum Testheft: das Testdesign

Für jede Testdomäne wurden vom IQB Hunderte von Testaufgaben entwickelt, um Inhalte und Schwierigkeitsspektrum der zu erfassenden Kompetenzbereiche möglichst umfassend abzubilden. Um die Testzeit und die von den Schülerinnen und Schülern jeweils zu bearbeitende Aufgabenmenge in einem vertretbaren Rahmen zu halten, wurde – wie in den meisten internationalen Schulleistungsstudien – auch in den Ländervergleichsstudien ein *Multi-Matrix-Design* gewählt (Gonzalez & Rutkowski, 2010; Frey, Hartig & Rupp, 2009), bei dem jede Schülerin und jeder Schüler nur einen Teil der Aufgaben des gesamten Aufgabenpools bearbeitete (siehe Kapitel 2, 3. Teilschritt: Pilotierungsstudien). Die Zusammenstellung der Testhefte erfolgte in einem mehrstufigen Verfahren, bei dem zunächst die Aufgaben Aufgabenblöcken und anschließend die Aufgabenblöcke Testheften zugeordnet wurden (siehe Tabelle 3.2). Jeder Aufgabenblock kam dabei über alle Testhefte hinweg an jeder Position ungefähr gleich häufig vor. Die Erstellung eines solchen balancierten Blockdesigns ist notwendig, da die Aufmerksamkeit und Motivation der Schülerinnen und Schüler im Verlauf der Testbearbeitung in der Regel nachlässt (Robitzsch, 2009). Mit diesem Design wird angestrebt, derartige Effekte der Aufgabenposition auf die Lösungswahrscheinlichkeit auszugleichen und diese in den Analysen statistisch zu kontrollieren.

In den bisherigen IQB-Ländervergleichen kamen zwischen 81 und 94 Testheftversionen zum Einsatz. Die Zusammenstellung der Aufgaben in Blöcke und Testhefte ermöglichte es, bei der späteren Auswertung der Daten eine Verbindung zwischen allen Testheftvarianten und somit zwischen allen Items des Aufgabenpools herzustellen. Trotz unterschiedlich zusammengesetzter Testhefte konnten damit die Leistungen

aller Schülerinnen und Schüler auf einer gemeinsamen Messskala abgebildet werden.

Vielfältige Expertise: beteiligte Personen und Institutionen

Um die IQB-Ländervergleichsstudien erfolgreich durchführen zu können, müssen viele Institutionen und Einzelpersonen miteinander kooperieren. Diese Abstimmungsprozesse beginnen bereits ein halbes Jahr vor dem jeweiligen Testzeitfenster. Auf Seiten der 16 Länder müssen beispielsweise Projektkoordinatorinnen und -koordinatoren benannt und auf ihre koordinierende Tätigkeit vorbereitet werden. Die Datenschutzbeauftragten der Länder müssen alle im Rahmen der Erhebungen eingesetzten Materialien prüfen und genehmigen (dazu gehören Informationsschreiben für Befragte, Schülerfragebögen, Elternfragebögen, Lehrerfragebögen, Schulleiterfragebögen, Testleitermaterialien). Die Schulen wiederum ermöglichen vor Ort einen reibungslosen Ablauf und bereiten die Schülerinnen und Schüler sowie deren Eltern auf die Durchführung vor. Das IQB ist nicht nur für die Entwicklung der Testaufgaben verantwortlich, sondern ebenfalls für die Auswahl der geeignetsten Testaufgaben aus den Itempools, die Erstellung der Testunterlagen, Begleitfragebögen und Auswertungsanleitungen und – im Anschluss an die Datenerhebung – die Auswertung und Berichterlegung zum Ländervergleich. Die Testdurchführung selbst erfolgt in den Schulen durch externe Testleiterinnen und Testleiter, die das IEA Data Processing Center in Hamburg (DPC) zuvor intensiv geschult hat. Zusätzlich verantwortet das DPC die Logistik der Erhebungen, wie die Ziehung der Stichprobe und den Versand der Testunterlagen. Die elektronische Datenerfassung einschließlich der Kodierung der Schülerantworten wird ebenfalls vom DPC übernommen.

Tabelle 3.2: Ausschnitt des Testdesigns im IQB-Ländervergleich 2011

Testheftgruppe	Testheft-Nr.	Block 1	Block 2	Block 3	Block 4
1	THo1	D-Zuhöreno1	D-Leseno1	D-Zuhöreno2	D-Lesen32
	THo2	D-Zuhöreno1	D-Lesen31	D-Zuhöreno2	D-Lesen03
2	THo3	D-Leseno1	D-Zuhöreno4	D-Leseno2	D-Zuhöreno3
	THo4	D-Lesen21	D-Zuhöreno4	D-Lesen42	D-Zuhöreno3
3	THo5	D-Zuhöreno3	D-Leseno3	D-Leseno1	D-Zuhöreno5
	THo6	D-Zuhöreno3	D-Lesen11	D-Lesen21	D-Zuhöreno5

Anmerkung: TH = Testheft. Die Bezeichnung der Blöcke setzt sich zusammen aus einer Abkürzung für das zu testende Fach (D = Deutsch), der Bezeichnung der operationalisierten Kompetenz und einer fortlaufenden Nummer für den Aufgabenblock innerhalb dieser Kompetenz in diesem Fach.

Berichtslegung

Zu jedem Ländervergleich wird am IQB ein ausführlicher Berichtsband erstellt, der neben theoretisch einführenden Kapiteln zentrale Ergebnisse enthält (siehe Köller, Knigge & Tesch, 2010; Stanat, Pant, Böhme & Richter, 2012; Pant et al., 2013). In vertiefenden Analysen werden ferner jeweils die mit dem Geschlecht, der sozialen Herkunft und dem Zuwanderungshintergrund verbundenen Disparitäten (Leistungsunterschiede) betrachtet. Der jeweilige Stand des in den Ländern erreichten Bildungsniveaus lässt sich auf verschiedene Weise darstellen und einordnen. Analog zu Leistungsbeurteilungen bei Schülerinnen und Schülern auf Individualebene, lassen sich auch für Ergebnisse von Leistungserhebungen auf der Systemebene unterschiedliche Vergleichsmaßstäbe heranziehen (Rheinberg, 1980). Hierbei können eine soziale, eine kriteriale sowie eine ipsative Bezugsnorm unterschieden werden.

Im Rahmen der Berichterstattung der Ergebnisse von Schulleistungsstudien steht häufig die *soziale* Bezugsnorm im Mittelpunkt öffentlicher Diskussionen. Hier werden im Sinne von Länderrankings die in den Ländern erreichten Durchschnittsleistungen miteinander verglichen und gefragt, auf welchem Rangplatz in der 16-Länder-Tabelle ein bestimmtes Land liegt. Dabei sollte genau darauf geachtet werden, ob die Rangplatzunterschiede auf nur geringen, statistisch nicht signifikanten Punktdifferenzen beruhen. Solche geringen numerischen Unterschiede sollten in aller Regel nicht inhaltlich interpretiert werden. Aber auch hinter statistisch signifikanten Abweichungen zweier Länderergebnisse voneinander oder eines Länderergebnisses vom Bundesdurchschnitt stehen mitunter recht geringe absolute Punktwertdifferenzen, die in der Praxis kaum als Nachweis für die qualitative Überlegenheit oder Unterlegenheit eines Schulsystems gewertet werden sollten. Daher bedarf es eines Maßstabs, der es erlaubt, die praktische Bedeutsamkeit von durchschnittlichen Leistungsunterschieden zwischen den Ländern zu beurteilen.

Einen solchen Maßstab zur Bewertung der Bedeutsamkeit von Punktwertunterschieden stellt der durchschnittlich erwartbare Lernzuwachs, ausgedrückt als Lernzeit, dar. Auf der Grundlage von Längsschnittdaten oder repräsentativen Querschnittsdaten aufeinanderfolgender Schülerkohorten (z.B. Schülerinnen und Schüler der dritten und der vierten Jahrgangsstufe) lässt sich der Kompetenzzuwachs schätzen, der im Verlauf eines Schuljahres im Durchschnitt erreicht wird (Bremerich-Vos & Böhme, 2009; Reiss & Winkelmann, 2009). Anhand dieser Daten kann dann die Aussage getroffen werden, wie viel Lernzeit erforderlich wäre, um den mittleren Leistungsnachteil von Schülerinnen und Schülern eines Landes im Vergleich zu einem anderen Land aufzuholen. Im IQB-Ländervergleich 2011 für die Primarstufe im Fach Mathe-

matik zum Beispiel entsprach der Unterschied zwischen dem niedrigsten Ländermittelwert (Berlin) und dem höchsten Mittelwert (Bayern) einer Lernzeit von ungefähr 10 Monaten (Haag & Roppelt, 2012).

Ein solcher „Lernzeit“-Maßstab hilft zwar bei der inhaltlichen Interpretation von Unterschieden; dies sagt aber nichts darüber aus, ob das erreichte Kompetenzniveau in jedem einzelnen Land für sich gesehen zufriedenstellend ist oder nicht. Hierfür ist eine *kriteriale* Bezugsnorm erforderlich, die auf inhaltlich definierte Lernziele rekurriert, wie sie die Bildungsstandards der KMK und die auf ihnen basierenden Kompetenzstufenmodelle formulieren. Im Rahmen der bereits beschriebenen *Standard-Setting-Verfahren* (siehe Kapitel 2) werden die Kompetenzstufen dabei so festgelegt, dass die einzelnen Stufen qualitativ unterscheidbaren Anforderungen entsprechen, die in den Standards der KMK beschrieben sind. Da das Ziel der IQB-Ländervergleichsstudien darin besteht, das Erreichen der Bildungsstandards zu überprüfen, ist für die Bewertung der Ergebnisse die Verteilung der Schülerinnen und Schüler auf die kriterial beschriebenen Kompetenzstufen entscheidend. So besteht beispielsweise politisch das Ziel, dass Kompetenzstufe I (unterhalb des Mindeststandards) möglichst gar nicht oder nur sehr gering besetzt ist. Je nach Zusammensetzung der Schülerschaft wird es dabei in den einzelnen Ländern unterschiedlich schwierig sein, dieses Ziel zu erreichen. Mittelfristig wäre daher ein gesellschaftlicher und politischer Verständigungsprozess darüber wünschenswert, welcher Anteil von Schülerinnen und Schülern, die die Minimalanforderungen verfehlen, vorübergehend noch akzeptiert werden kann und mit welchen konkreten Schritten und in welchem Tempo er reduziert werden soll.

Ein weiterer wichtiger Vergleichsmaßstab für die Bewertung von Ergebnissen des Bildungsmonitorings ist die Entwicklungs- oder *ipsative* Bezugsnorm, bei der die aktuellen Landesergebnisse mit denen früherer Erhebungen verglichen werden. Hier geht es also um die Betrachtung des *Trends*, ob und inwieweit sich das mittlere Kompetenzniveau und die Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen (insgesamt und für bestimmte Teilgruppen, wie etwa Heranwachsende aus zugewanderten Familien) in den einzelnen Ländern verändert haben. Solche Trendaussagen konnten bislang nur im Rahmen der internationalen Schulleistungsstudien für Deutschland insgesamt getroffen werden, wobei sich in der Primarstufe keine nachweisbaren Veränderungen (Bos, Tarelli et al., 2012; Bos, Wendt et al., 2012), in der Sekundarstufe I hingegen signifikante Leistungszuwächse insbesondere für Schülerinnen und Schüler aus sozial schwachen und zugewanderten Familien zeigten (Klieme et al., 2010; Prenzel, Sälzer, Klieme & Köller, 2013). In

Bezug auf das Erreichen der Bildungsstandards werden zukünftig Aussagen über Veränderungen auf Länderebene zwischen 2009 und 2015 für die sprachlichen Fächer in der Sekundarstufe I, zwischen 2011 und 2016 für Deutsch und Mathematik in der Primarstufe sowie zwischen 2012 und 2018 für Mathematik und die naturwissenschaftlichen Fächer in der Sekundarstufe I möglich sein.

Zur Aussagekraft von Studien des Bildungsmonitorings

Bereits nach der Berichtslegung zum zweiten IQB-Ländervergleich wurde in Politik, Verbänden und der Öffentlichkeit kontrovers diskutiert, ob und inwieweit es sinnvoll ist, diese Art von Studien regelmäßig durchzuführen.

Die Kritik zielt dabei auf folgende vier Aspekte: (1) Ländervergleichsstudien reproduzierten lediglich bereits bekannte Erkenntnisse; (2) sie klärten die Ursachen bestehender Leistungsdifferenzen zwischen den Ländern nicht auf; (3) sie lieferten kein Steuerungs- bzw. Handlungswissen für Bildungspolitik und Schulpraxis und (4) die Ländervergleiche seien nicht „fair“, weil die Zusammensetzung der Schülerschaft sich stark zwischen den Ländern unterscheidet.

Mit dem ersten Kritikansatz wird bemängelt, die Untersuchungen würden lediglich bereits bekannte Ergebnisse replizieren (GEW, 2013). Dabei wird ausschließlich auf das Länderranking eingegangen, das in den letzten zehn Jahren in der Tat relativ stabil war. Die wichtigsten Befunde der IQB-Ländervergleichsstudien sind jedoch nicht – wie oben ausführlich erläutert – die Länderrankings, sondern die Anteile der Schülerinnen und Schüler auf den Kompetenzstufen sowie Veränderungen in den erreichten Kompetenzständen über die Zeit. Solche Trendanalysen werden in Zukunft im Rahmen der IQB-Ländervergleichsstudien möglich sein, wobei insbesondere Verschiebungen in den Verteilungen auf den kriterial verankerten Kompetenzstufen in den Fokus genommen werden. Besondere Aufmerksamkeit wird der Frage zukommen, ob es solchen Ländern, in denen substantielle Anteile von Schülerinnen und Schülern die Mindeststandards verfehlen, über die Jahre gelingt, leistungsrelevante Maßnahmen der Qualitätsentwicklung zu implementieren.

Aber auch ohne solche Trendaussagen haben die Ländervergleichsstudien des IQB Erkenntnisse generiert, die für eine Steuerung auf Systemebene genutzt werden können. So hat der Ländervergleich 2009 in der Sekundarstufe I gezeigt, dass im Fach Englisch Schülerinnen und Schüler in den östlichen Flächenländern deutlich geringere Kompetenzniveaus erreichen als Schülerinnen und Schüler in den restlichen

Bundesländern (Leucht, Frenzel & Pöhlmann, 2010), im Ländervergleich 2012 weisen die gleichen Länder dagegen deutliche Kompetenzvorteile im Fach Mathematik und in den naturwissenschaftlichen Fächern auf (Roppelt, Penk, Pöhlmann & Pietsch, 2013; Schroeders et al., 2013). Weiterhin beinhalten die Berichte über die IQB-Ländervergleichsstudien immer auch Analysen zu Faktoren, die für die Qualitätsentwicklung relevant sind, wie etwa der potenzielle Zugang von Kindern mit geringen Kompetenzen im *Lesen* und *Hörverstehen* zu Angeboten der Sprach- und Leseförderung in Primarschulen (Stanat, Weirich & Radmann, 2012), zu Aspekten der Aus- und Weiterbildung von Lehrkräften (Richter, Kuhl, Reimers & Pant, 2012; Richter, Kuhl, Haag & Pant, 2013), zur Rolle unterschiedlicher Stundenkontingente in den naturwissenschaftlichen Fächern (Schroeders, Siegle, Weirich & Pant, 2013) sowie zu motivationalen Schülermerkmalen (Jansen, Schroeders & Stanat, 2013).

Der zweite Kritikansatz zielt darauf, dass die Studien zum Bildungsmonitoring keine Erkenntnisse darüber liefern würden, worauf die Kompetenzunterschiede zwischen den Ländern ursächlich zurückzuführen sind. Diese Beobachtung kann schwer bestritten werden. Die deutschen Länder und ihre Bildungssysteme unterscheiden sich in einer Vielzahl von Merkmalen, die für Lehr-Lern-Prozesse und deren Erträge relevant sein können, und es ist praktisch unmöglich, die Effekte dieser einzelnen Faktoren zu isolieren (Stanat & Lüdtke, 2007). Zudem weisen die Ergebnisse internationaler Schulleistungsstudien darauf hin, dass in sehr unterschiedlich ausgestalteten Bildungssystemen (z.B. Finnland und Japan) ähnlich gute Leistungen erzielt werden können. Diese Erfolge lassen sich nicht auf wenige Einzelmerkmale zurückführen, sondern basieren auf dem Zusammenspiel zahlreicher Faktoren, die auch Kontextbedingungen schulischer Arbeit einschließen (z.B. die gesellschaftliche Wertschätzung von Bildung oder bestimmten Schulfächern). Die genauere Bestimmung solcher Konstellationen kann nicht im Rahmen von Studien zum Bildungsmonitoring geleistet werden, sondern bedarf ausformulierter Forschungsprogramme, die qualitative Analysen schulischer Systeme (z.B. Stevenson & Nerison-Low, 1997) ebenso einschließen können wie vergleichende Videoanalysen von Unterricht (z.B. Klieme & Baumert, 2001).

Der dritte, häufig geäußerte Kritikansatz bemängelt, die Studien des Bildungsmonitorings würden keine Aussagen darüber erlauben, welche Maßnahmen ergriffen werden sollten, um die Qualität der Lernprozesse zu optimieren. Auch dies können Bildungsmonitoringstudien aufgrund ihrer Anlage nicht leisten. Zudem unterscheiden sich die Problemlagen und deren potenzielle Ursachen in den verschiedenen Ländern teilweise erheblich (so dürfte zum Beispiel das schwa-

che Abschneiden von Schülerinnen und Schülern der Sekundarstufe I im Bereich *Lesen* in Brandenburg und Bremen unterschiedliche Ursachen haben), so dass es auf die Frage, welche Maßnahmen zu ergreifen sind, keine allgemeingültigen Antworten gibt. Vielmehr ist es erforderlich, ein kohärentes Gesamtkonzept der Qualitätsentwicklung zu formulieren, das auf die spezifischen Problemlagen des jeweiligen Landes abgestimmt ist und von dem theoretisch begründet angenommen werden kann, dass es zu einer Verbesserung der Qualität von Lehr-Lern-Prozessen im Unterricht führen wird. Der Beitrag der Studien zum Bildungsmonitoring besteht dabei darin, spezifische Stärken und Schwächen zu identifizieren, die bei der Entwicklung des jeweiligen Konzepts zur Qualitätsentwicklung besonders in den Blick zu nehmen sind, und langfristige Wirkungen in Trendanalysen zu evaluieren.

Auch in die Ausgestaltung des Qualitätskonzepts sollten selbstverständlich Erkenntnisse der empirischen Bildungsforschung eingehen, die Hinweise auf Effekte (oder auch auf ausbleibende Wirkungen und potenzielle Nebeneffekte) von Maßnahmen geben. Hierfür sind jedoch nicht primär Ergebnisse zum Bildungsmonitoring relevant, sondern verschiedene andere Arten von Studien, wie etwa Untersuchungen zur Unterrichtsqualität, Interventionsstudien zur Überprüfung der Wirksamkeit von spezifischen Fördermaßnahmen oder Analysen zur Gestaltung effektiver Weiterbildung für Lehrkräfte. Zu manchen Fragen, wie etwa der nach Merkmalen guten Unterrichts, existiert bereits ein solider Forschungsstand (vgl. z.B. Helmke, 2009; Klieme & Rakoczy, 2008; Seidel & Shavelson, 2007), während die Befundlage

zu anderen Fragestellungen, wie zum Beispiel der Effektivität von verschiedenen Ansätzen der Sprachförderung, noch deutlich weniger befriedigend ist (vgl. z.B. Paetsch, Wolf, Stanat & Darsow, im Druck). Aber selbst ein relativ klarer Erkenntnisstand zur Effektivität bestimmter Maßnahmen kann bildungspolitische Entscheidungen nicht ersetzen, die immer auch andere Erwägungen, wie etwa Fragen der Akzeptanz und der Finanzierbarkeit, einbeziehen müssen. Dabei wäre es allerdings wünschenswert, dass die Grundlage der jeweiligen Entscheidung transparent gemacht wird.

Ein vierter Kritikpunkt verweist darauf, dass sich die Bundesländer wirtschaftlich und soziokulturell stark unterscheiden und es daher unfair sei, diese zu vergleichen. Stattdessen sollten die Vergleiche auf Regionen bezogen werden, die sich in dieser Hinsicht ähnlicher sind. Um abschätzen zu können, inwieweit das Erreichen schulischer Erträge von ökonomischen und soziokulturellen Kontextbedingungen abhängt, kann es durchaus sinnvoll sein, die Ländervergleiche durch regionsbezogene Vergleiche zu ergänzen. Ein Ersetzen wäre jedoch insofern problematisch, als die bildungspolitische Verantwortung nun einmal gesetzlich in die Zuständigkeit der Länder fällt. Zudem würden die Länder damit aus der Pflicht genommen werden, die selbst gesetzten Ziele für zentrale Erträge schulischer Bildungsprozesse zu erreichen. Auch wenn es angesichts schwieriger sozialer Lagen in manchen Ländern deutlicher mühevoller ist als in anderen, die Kompetenzerwartungen der Bildungsstandards zu erfüllen, so sollten doch alle Länder an ihrem eigenen Anspruch festhalten, dass möglichst alle Schülerinnen und Schüler wenigstens die Mindeststandards erreichen.



Literatur

- Baumert, J., Bos, W. & Lehmann, R. H. (Hrsg.). (2000a). *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (Bd. 1., Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit). Opladen: Leske + Budrich.
- Baumert, J., Bos, W. & Lehmann, R. H. (Hrsg.). (2000b). *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (Bd. 2., Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe). Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W. et al. (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Lehmann, R. H., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O. & Neubrand, J. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Bos, W., Tarelli, I., Bremerich-Vos, A. & Schwippert, K. (Hrsg.). (2012). *IGLU 2011 – Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Wendt, H., Köller, O. & Selter, C. (Hrsg.). (2012). *TIMSS 2011 – Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bremerich-Vos, A. & Böhme, K. (2009). Lesekompetenzdiagnostik – die Entwicklung eines Kompetenzstufenmodells für den Bereich Lesen. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 219–249). Weinheim: Beltz.
- Frey, A., Hartig, J. & Rupp, A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and practice. *Educational Measurement: Issues and Practice*, 28, 39–53.

- GEW (2012) = Gewerkschaft für Erziehung und Wissenschaft.** *Neuer_Test_altbekannte_Ergebnisse*. Zugriff am 18.03.2013 unter URL: http://www.gew.de/Neuer_Test_altbekannte_Ergebnisse.html
- Gonzalez, E. & Rutkowski, L. (2010).** Principles of multiple matrix booklet designs and parameter recovery in large-scale assessments. *IERI Monograph Series*, 3, 125–156.
- Haag, N. & Roppelt, A. (2012).** Der Ländervergleich im Fach Mathematik. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 117–127). Münster: Waxmann.
- Helmke, A. (2009).** *Unterrichtsqualität und Lehrerprofessionalität – Diagnose, Evaluation und Verbesserung*. Seelze-Velber: Kallmeyer.
- Hurrelmann, K. (2013).** Das Schulsystem in Deutschland: Das „Zwei-Wege-Modell“ setzt sich durch. *Zeitschrift für Pädagogik*, 59, 455–468.
- Jansen, M., Schroeders, U. & Stanat, P. (2013).** Motivationale Schülermerkmale in Mathematik und den Naturwissenschaften. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 347–366). Münster: Waxmann.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O. & Prenzel, M. (Hrsg.). (2010).** *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Klieme, E. & Baumert, J. (2001).** Identifying national cultures of mathematics education: Analysis of cognitive demands and differential item functioning in TIMSS. *European Journal of Psychology of Education*, 16, 383–400.
- Klieme, E. & Rakoczy, K. (2008).** Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54, 222–237.
- KMK (2006) = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). (2006).** *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. München: Wolters Kluwer.
- Köller, O., Knigge, M. & Tesch, B. (Hrsg.). (2010).** *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.
- Leucht, M., Frenzel, J. & Pöhlmann, C. (2010).** Der Ländervergleich im Fach Englisch. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 97–104). Münster: Waxmann.
- Paetsch, J., Wolf, K. M., Stanat, P. & Darsow, A. (im Druck).** Sprachförderung von Kindern und Jugendlichen aus Zuwandererfamilien. *Zeitschrift für Erziehungswissenschaft*.
- Pant, H.A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013).** *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (Hrsg.). (2013).** *PISA 2012. Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann.
- Reiss, K. & Winkelmann, H. (2009).** Kompetenzstufenmodelle für das Fach Mathematik im Primarbereich. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 120–141). Weinheim: Beltz.
- Rheinberg, F. (1980).** *Leistungsbewertung und Lernmotivation*. Göttingen: Hogrefe.
- Richter, D., Kuhl, P., Reimers, H. & Pant, H. A. (2012).** Aspekte der Aus- und Fortbildung von Lehrkräften in der Primarstufe. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 237–250). Münster: Waxmann.
- Richter, D., Kuhl, P., Haag, N. & Pant, H.A. (2013).** Aspekte der Aus- und Fortbildung von Mathematik- und Naturwissenschaftslehrkräften im Ländervergleich. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 367–390). Münster: Waxmann.
- Robitzsch, A. (2009).** Methodische Herausforderungen bei der Kalibrierung von Leistungstests. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Wather (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 42–106). Weinheim: Beltz.
- Roppelt, A., Penk, C., Pöhlmann, C. & Pietsch, E. (2013).** Der Ländervergleich im Fach Mathematik. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 123–140). Münster: Waxmann.
- Schroeders, U., Hecht, M., Heitmann, P., Jansen, M., Kampa, N., Klebba, N. et al. (2013).** Der Ländervergleich in den naturwissenschaftlichen Fächern. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 141–158). Münster: Waxmann.
- Schroeders, U., Siegle, T., Weirich, S. & Pant, H.A. (2013).** Der Einfluss von Kontext- und Schülermerkmalen auf die naturwissenschaftlichen Kompetenzen. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 331–346). Münster: Waxmann.
- Seidel, T. & Shavelson, R. J. (2007).** Teaching effectiveness research in the last decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77, 454–499.
- Stanat, P., Pant, H. A., Böhme, K. & Richter, D. (Hrsg.). (2012).** *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011*. Münster: Waxmann.
- Stanat, P., Weirich, S. & Radmann, S. (2012).** Sprach- und Leseförderung. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 251–276). Münster: Waxmann.
- Stanat, P. & Lüdtke, O. (2007).** Internationale Schulleistungsvergleiche. In G. Trommsdorf & H. J. Kornadt (Hrsg.), *Enzyklopädie der Psychologie: Kulturvergleichende Psychologie, Band 3: Kulturelle Determinanten des Erlebens und Verhaltens* (S. 279–347). Göttingen: Hogrefe.
- Stevenson, H. W. & Nerison-Low, R. (1997).** *To sum it up: Case studies of education in Germany, Japan, and the United States*. Washington, D.C.: U.S. Government Printing Office.
- Tillmann, K.-J. (2012).** Das Sekundarschulsystem auf dem Weg in die Zweigliedrigkeit. *Pädagogik*, 64, 8–12.



04 // Vergleichsarbeiten

Ein wichtiger Bestandteil der Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring

Im Rahmen der Gesamtstrategie zum Bildungsmonitoring wurde im Jahr 2006 von der Kultusministerkonferenz (KMK) beschlossen, Lehrkräften regelmäßig Rückmeldungen darüber zu geben, inwieweit ihre Schülerinnen und Schüler über die in den Bildungsstandards definierten Kompetenzen verfügen. Hierzu dienen die länderübergreifend einheitlichen Vergleichsarbeiten (VERA), die jedes Jahr flächendeckend durchgeführt werden.

Was ist VERA?

Vergleichsarbeiten sind schriftliche Tests, die jahrgangsbezogen feststellen, in welchem Maße Schülerinnen und Schüler zu einem bestimmten Zeitpunkt die angezielten Kompetenzen bereits entwickelt haben. Diese Tests sind flächendeckend. Dies bedeutet, dass VERA in der dritten und achten Jahrgangsstufe so gut wie aller allgemein bildenden Schulen und Klassen in Deutschland durchgeführt wird (VERA 3 bzw. VERA 8). Eine Ausnahme bildet bis 2014 noch Baden-Württemberg, wo in der achten Jahrgangsstufe anstelle von VERA ein landeseigenes Verfahren der Kompetenzmessung zur Anwendung kommt. Darüber hinaus führt eine Gruppe von Ländern (Hamburg, Hessen, Mecklenburg-Vorpommern, Sachsen und Thüringen sowie Südtirol) zusätzlich Vergleichsarbeiten in der Jahrgangsstufe 6 durch (VERA 6). In einigen Bundesländern werden die VERA-Tests nicht als Vergleichsarbeiten, sondern als „Lernstandserhebungen“ (Hessen, Nordrhein-Westfalen), „KERMIT – Kompetenzen ermitteln“ (Hamburg) oder „Kompetenztests“ (Sachsen, Thüringen) bezeichnet. Vergleichsarbeiten orientieren sich inhaltlich an den Bildungsstandards und werden im Rahmen von VERA 3 und VERA 8 durch den Einsatz von normierten Items (sog. Ankeritems) an die Metrik der Bildungsstandards angekoppelt. Damit wird es möglich, Rückmeldungen zum Stand des Erreichens der Bildungsstandards zur Verfügung zu stellen. Das IQB ist seit 2008 bzw. 2009 dafür zuständig, die Aufgaben für VERA 3 und VERA 8 zu entwickeln, zu erproben und begleitende didaktische Materialien für die Unterrichtspraxis zu erstellen. Seit 2011 koordiniert das IQB zudem die Entwicklung von Aufgaben für VERA 6 im Fach Deutsch.

Ziele von VERA

Die Kultusministerkonferenz hat 2012 in einer Vereinbarung zur Weiterentwicklung von VERA als zentrale Funktion von Vergleichsarbeiten die Unterrichts- und Schulentwicklung definiert. Lehrkräfte erhalten mit VERA eine an objektiven Maßstäben orientierte Momentaufnahme des Leistungsstands ihrer Schülerinnen und Schüler. Die Ergebnisse der Vergleichsarbeiten geben Hinweise darauf, in welchen Bereichen die Schülerinnen und Schüler die erfassten Kompetenzen bereits aufgebaut haben und in welchen Bereichen ihre Kompetenzen noch gestärkt werden sollten. Die Ergebnissrückmeldungen für Lehrkräfte werden von den einzelnen Ländern erstellt und enthalten in der Regel Angaben (1) zur Leistung jeder einzelnen Schülerin und jedes einzelnen Schülers, (2) zur mittleren Leistung der eigenen Klasse, (3) zur

mittleren Leistung der Parallelklassen an der eigenen Schule und (4) zur mittleren Leistung der eigenen Schule. Einige Länder vergleichen in den Rückmeldungen außerdem die mittlere Leistung der jeweiligen Schule mit der mittleren Leistung von Schulen der jeweiligen Schulform im eigenen Bundesland, teilweise unter Berücksichtigung der Zusammensetzung der Schülerschaft. Die VERA-Testungen finden deutlich vor Abschluss der Bildungsetappen statt, auf die sich die Bildungsstandards beziehen, d. h. etwa ein Jahr vor dem Ende der Primarstufe¹ bzw. zwei Jahre vor dem Ende der Sekundarstufe I. Im Sinne eines Frühwarnsystems sollen Lehrkräfte und Schulleitungen damit Hinweise auf möglichen Handlungsbedarf erhalten, die es ihnen ermöglichen, Angebote zur gezielten Förderung der Schülerinnen und Schüler zu entwickeln und umzusetzen.

Die Vergleichsarbeiten dienen primär dazu, mit Hilfe von Leistungsrückmeldungen Lehrkräfte bei der Entwicklung eines kompetenzorientierten, auf das Erreichen der Bildungsstandards ausgerichteten Unterrichts zu unterstützen. Die Tests der Vergleichsarbeiten zielen *nicht* darauf ab, den gerade durchgenommenen Unterrichtsstoff oder die aktuell zu behandelnden Inhalte des Curriculums abzufragen. Es wird in der Vereinbarung der KMK betont, dass VERA weder zur Benotung noch für eine verbindliche Prognose des Schulerfolgs an weiterführenden Schulen geeignet ist. Damit die Vergleichsarbeiten ihren primären Zweck erfüllen können, die Weiterentwicklung eines kompetenzorientierten Unterrichts zu unterstützen, soll gemäß der KMK-Vereinbarung ferner davon abgesehen werden, VERA-Ergebnisse der einzelnen Schulen in Form von *Ranking*-Tabellen zu veröffentlichen. Die Schulaufsicht und die Schulinspektion sollen VERA-Ergebnisse nur nach klaren Regelungen einsehen können, um zu verhindern, dass die Ergebnisse als Instrument der Kontrolle verwendet werden. Wie die internationale Forschung zeigt, können Leistungserhebungen, die mit erheblichen Konsequenzen für die einzelne Lehrkraft oder Schule verbunden sind (sog. „high-stakes tests“), einer Verfälschung der Ergebnisse etwa durch Hilfestellungen bei der Testbearbeitung Vorschub leisten.

Um die Schul- und Unterrichtsentwicklung zusätzlich zu unterstützen, stellt das IQB jedes Jahr außerdem ausführliche fachdidaktische Kommentare zur Verfügung, die die VERA-Aufgaben hinsichtlich der zur ihrer Bearbeitung erforderlichen Kompetenzen einordnen, typische Fehler und deren mögliche Ursachen beschreiben sowie Hinweise zur Weiterarbeit im Unterricht geben (für Aufgaben- und Materialbeispiele siehe <http://www.iqb.hu-berlin.de/vera/aufgaben>).

¹ Dies gilt nicht für die Länder Berlin und Brandenburg, in denen die Primarstufe in der Regel sechs Schuljahre umfasst.

Das pädagogische Potenzial, das mit VERA für Lehrkräfte und Schulen verbunden ist, basiert unter anderem auf

- der Kompetenzorientierung der Testaufgaben und Ergebnismeldungen, die auch den Unterricht verstärkt prägen soll,
- dem „Blick nach außen“, den die multiplen Vergleichsmöglichkeiten für den Lernstand der eigenen Klasse eröffnen,
- den Rückmeldungen über den relativen Leistungsstand der Schülerinnen und Schüler einer Klasse, die zur Weiterentwicklung der diagnostischen Kompetenzen von Lehrkräften beitragen können,
- den Informationen über die Kompetenzprofile der Klassen und den didaktischen Kommentaren, die bei der Planung und Umsetzung pädagogischer Interventionen und Fördermaßnahmen genutzt werden können, und
- die Möglichkeiten der Nutzung der Ergebnismeldungen für kooperative Unterrichtsentwicklung im Kollegium.

VERA unterscheidet sich in mehrfacher Hinsicht von den internationalen Schulleistungsstudien (wie PISA, TIMSS und IGLU/PIRLS) und den Erhebungen, die das IQB durchführt, um das Erreichen der Bildungsstandards im Ländervergleich zu überprüfen (siehe Kapitel 3). Diese Unterschiede betreffen vor allem die verfolgten Ziele (*Wozu sollen die Ergebnisse dienen?*) und die Evaluationsebene (*Wer wird evaluiert?*). Einige der wichtigsten Unterschiede zwischen den Schulleistungsstudien und VERA sind in Tabelle 4.1 zusammengefasst.

Beteiligte Jahrgangsstufen und Fächer

VERA 3, VERA 6 und auch VERA 8 werden jeweils im Frühjahr jeden Jahres in der entsprechenden Jahrgangsstufe durchgeführt. Die Ergebnisse werden zumeist noch vor den Sommerferien durch die auswertenden Institute der Länder an die Schulen übermittelt.

In der Primarstufe haben sich alle 16 Länder auf die Durchführung von Vergleichsarbeiten am Ende der Jahrgangsstufe 3 in Deutsch und Mathematik verständigt (VERA 3). Im Fach Deutsch werden dabei jedes Jahr die *Lesekompetenz* sowie ein weiterer Kompetenzbereich (*Zuhören, Schreiben, Rechtschreibung* oder *Sprache und Sprachgebrauch untersuchen*) getestet; im Fach Mathematik werden in jedem Durchgang zu zwei der fünf inhaltlichen Kompetenzbereiche (1) *Zahlen und Operationen*, (2) *Raum und Form*, (3) *Muster und Strukturen*, (4) *Größen und Messen* sowie (5) *Daten, Häufigkeit und Wahrscheinlichkeit* Aufgaben gestellt. Welche beiden Bereiche dies sind, wird vor jedem Testdurchgang von den Ländern gemeinsam mit dem IQB festgelegt. Im Fach Mathematik wird ein Testheft, im Fach Deutsch werden zwei Testhefte eingesetzt. Die Testzeit umfasst im Fach Deutsch verteilt auf zwei Testtage insgesamt 70–80 Minuten, die Erhebung im Fach Mathematik erfolgt an einem weiteren Testtag mit 60 Minuten.

Tabelle 4.1: Unterschiede zwischen stichprobenbasierten Schulleistungsstudien und Vergleichsarbeiten

	Internationale Schulleistungsstudien (PISA, PIRLS/IGLU, TIMSS)	Nationale Schulleistungsstudien (Ländervergleiche des IQB)	Vergleichsarbeiten (VERA 3 und VERA 8)
Design	Stichprobenerhebung	Stichprobenerhebung	Vollerhebung: alle Schülerinnen und Schüler einer Jahrgangsstufe
Häufigkeit	alle 3–5 Jahre	alle 5 Jahre (Primarstufe) bzw. alle 3 Jahre (Sekundarstufe I)	jährlich
Hauptziel	Systemmonitoring	Systemmonitoring	Unterrichts-/Schulentwicklung
Evaluationsebene	Staaten	Länder in der Bundesrepublik Deutschland	Schulen, Lerngruppen bzw. Klassen
Durchführung	externe Testleiter/innen	externe Testleiter/innen	in der Regel Lehrkräfte
Auswertung	zentral	zentral	dezentral durch Lehrkräfte sowie Landesinstitute
Ergebnismeldung	nach ca. 1,5 Jahren	nach ca. 1,5 Jahren	Sofortrückmeldungen im Anschluss an die Dateneingabe; differenzierte Rückmeldungen mit multiplen Vergleichswerten nach wenigen Wochen

In der Sekundarstufe I wird VERA am Ende der achten Jahrgangsstufe (VERA 8) in den Fächern Deutsch, Mathematik sowie erste Fremdsprache in 15 der 16 Länder durchgeführt (bis 2014 noch ohne Baden-Württemberg). Die erste Fremdsprache ist für die Schülerinnen und Schüler in den meisten Ländern Englisch, in einigen Ländern aber auch teilweise Französisch. In Deutsch und der ersten Fremdsprache sind jeweils zwei der inhaltlichen Kompetenzbereiche Gegenstand der Vergleichsarbeiten (immer *Lesekompetenz* sowie entweder *Zuhören*, *Schreiben*, *Rechtschreibung* oder *Sprache und Sprachgebrauch untersuchen*); in Mathematik werden mit (1) *Zahl*, (2) *Messen*, (3) *Raum und Form*, (4) *Funktionaler Zusammenhang* und (5) *Daten und Zufall* alle fünf inhaltlichen Kompetenzbereiche erfasst. Für VERA 8 entwickelt das IQB drei unterschiedlich schwierige Testheftvarianten, die sich grob an den Leistungsniveaus von Haupt-, Real- bzw. Gymnasialschülerinnen und -schülern orientieren. Es bleibt den Ländern überlassen, welche Testheftvarianten sie jeweils einsetzen. Die Testzeit in der 8. Jahrgangsstufe beläuft sich pro Fach auf 80 Minuten, wobei für jedes Fach ein Testtag vorgesehen wird.

In der sechsten Jahrgangsstufe werden in den an VERA 6 teilnehmenden Ländern (s. o.) im Fach Deutsch jedes Jahr die Kompetenzbereiche *Zuhören*, *Lesen* sowie *Sprache und Sprachgebrauch untersuchen* getestet. Darüber hinaus umfasst VERA 6 Erhebungen in den beiden Fächern Mathematik und Englisch, für die zurzeit nicht das IQB die Aufgabenentwicklung koordiniert. Die Testzeit umfasst für die Fächer Deutsch, Mathematik und Englisch jeweils an einem eigenen Testtag 80 Minuten.

Das IQB und die Länder: Wer ist wofür bei VERA zuständig?

Zwischen dem IQB einerseits und den Landesinstituten, Qualitätsagenturen bzw. zuständigen Fachabteilungen der Ministerien andererseits gibt es in Bezug auf VERA eine klare Arbeitsteilung (siehe Tabelle 4.2).

Neben der Ziehung der Stichprobe für die Pilotierung der VERA-Aufgaben sind die Länder für die Durchführung des VERA-Tests zuständig. Sie organisieren die Vorbereitung, den Ablauf der Erhebungen, die Auswertung der Daten und die Ergebnisrückmeldung jeweils in eigener Verantwortung auf der Grundlage länderspezifischer Regelungen. In den Ländern wird der Druck und Versand der Testhefte organisiert, werden Internetportale zur Eingabe der Ergebnisse entwickelt und Schulen mit Informationsveranstaltungen und Materialien auf VERA vorbereitet. Auch der Testumfang und die eingesetzten Testaufgaben können durch die Länder den landesspezifischen Bedürfnissen und Besonderheiten angepasst werden.

Das IQB hingegen ist im Rahmen von VERA für die Entwicklung von Aufgaben und fachdidaktischen Kommentaren zuständig. Die Testaufgaben werden länderübergreifend von Lehrkräften erarbeitet, von Fachdidaktikerinnen und Fachdidaktikern aus verschiedenen Hochschulen inhaltlich überprüft und bewertet sowie von wissenschaftlichen Testspezialisten am IQB anhand von Daten aus Pilotierungsstudien mit jeweils mehreren hundert Schülerinnen und Schülern im Hinblick auf ihre Schwierigkeit und psychometrische Eignung überprüft. In VERA werden nur solche Aufgaben eingesetzt,

Tabelle 4.2: Zuständigkeiten im Rahmen von VERA

In der Verantwortung ... des IQB	der Länder
• Aufgabenentwicklung • Aufgabenerprobung (Pilotierung) • Ermittlung der Aufgabenschwierigkeiten (Skalierung) • Zusammenstellung der Testhefte • Erarbeitung der didaktischen Materialien begleiten	• Stichprobenziehung für die Aufgabenerprobung (Pilotierung) • Druck und Distribution der Testhefte • Testdurchführung • Korrektur und Dateneingabe • statistische Auswertung • Rückmeldungsgestaltung • Rückmeldung • Unterstützung der Schulen durch flankierende Maßnahmen nach der Ergebnisrückmeldung

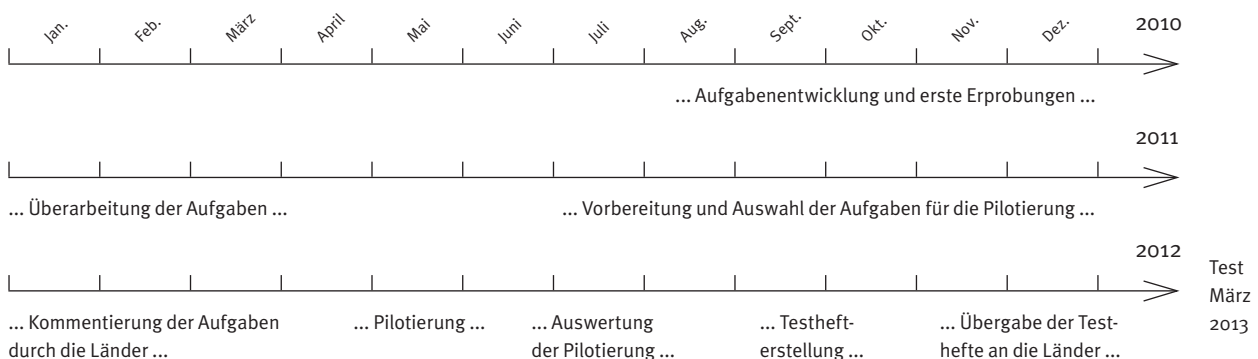
die sich in den Pilotierungen bewährt haben. Diese didaktisch und psychometrisch geprüften Aufgaben stellt das IQB in den Testheften zusammen und übermittelt diese gemeinsam mit ausführlichen didaktischen Kommentierungen den jeweils zuständigen Einrichtungen, die für die Administration von VERA in den Ländern verantwortlich sind.

Um die erforderlichen Abstimmungen mit den Ländern und einen reibungslosen Ablauf der VERA-Testungen zu ermöglichen, wird am IQB bereits gut zwei Jahre vor der Testung damit begonnen, die erforderlichen Materialien vorzubereiten. Der zeitliche Ablauf wird in der folgenden Grafik am Beispiel der VERA-Testung im Jahr 2013 veranschaulicht (siehe Abbildung 4.1).

Beispielaufgaben für VERA 3, VERA 6 und VERA 8

Zur Veranschaulichung werden im Folgenden drei VERA-Aufgaben dargestellt: Eine Mathematik-Aufgabe aus VERA 3, eine Deutsch-Aufgabe aus VERA 6 und eine Englisch-Aufgabe aus VERA 8. Zusätzlich werden in Auszügen einige didaktische Kommentare abgebildet, die jeder Lehrkraft zusammen mit den Aufgaben zugänglich gemacht werden. Nach jedem VERA-Durchgang werden die eingesetzten Aufgaben und didaktischen Kommentare in der Regel auf den Internetseiten des IQB veröffentlicht (<http://www.iqb.hu-berlin.de/vera/aufgaben>).

■ **Abbildung 4.1: Vorbereitende Arbeitsprozesse für die VERA 8-Testung im Jahr 2013**

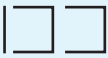
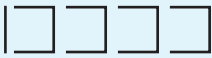


Aufgabenbeispiel VERA 3 Mathematik

Im Folgenden ist eine typische Mathematik-Aufgabe aus dem Inhaltsbereich *Muster und Strukturen* für VERA 3 abgebildet. Diese Aufgabe umfasst vier Items, die in ihrer Schwierigkeit zunehmen.

Beispiel für eine richtige Schülerlösung

Niki legt mit Holzstäbchen ein Muster und schreibt die passenden Aufgaben darunter.

1.	2.	3.	4.	5.
				
$1 \cdot 3 + 1 = 4$	$2 \cdot 3 + 1 = 7$	$3 \cdot 3 + 1 = 10$	$4 \cdot 3 + 1 = 13$	$5 \cdot 3 + 1 = 16$

- a) Zeichne die 3. Figur.
- b) Schreibe die 4. Aufgabe auf.
- c) Zeichne die 5. Figur.
Schreibe die 5. Aufgabe darunter.
- d) Für jede weitere Figur legt Niki immer 3 Hölzchen dazu.

Didaktischer Kommentar zur abgebildeten Mathematik-Aufgabe

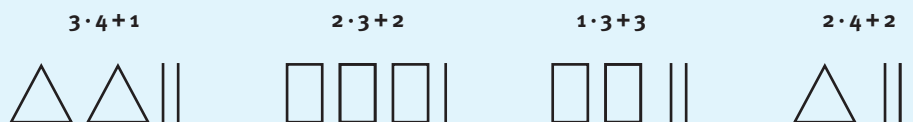
In den didaktischen Kommentaren erhalten Lehrkräfte nach Abschluss der Testungen für diese Aufgabe zunächst folgende Kommentare:

- Die Kinder müssen das Bildungsgesetz einer geometrischen und einer analogen arithmetischen Folge erkennen.
- Bei Teilaufgabe a besteht die Anforderung darin, die geometrische Folge um das nächste Glied zeichnerisch fortzusetzen.
- Bei Teilaufgabe b besteht die Anforderung darin, die nächste Gleichung in der Folge zu finden und aufzuschreiben. Dabei können die Kinder das Bildungsgesetz aus den vorangehenden Gleichungen ableiten oder zur Unterstützung auch die darüber befindliche geometrische Struktur heranziehen.
- Bei Teilaufgabe c muss sowohl die geometrische Folge um die nächste Figur zeichnerisch fortgesetzt als auch die nächste Gleichung in der Folge aufgeschrieben werden.
- In Teilaufgabe d besteht die Anforderung darin, das Bildungsgesetz der geometrischen Folge mit Hilfe einer im Wesentlichen vorgegebenen Teilformulierung zu erfassen und zu beschreiben.

Anschließend werden in den didaktischen Kommentaren Anregungen für den Unterricht gegeben:

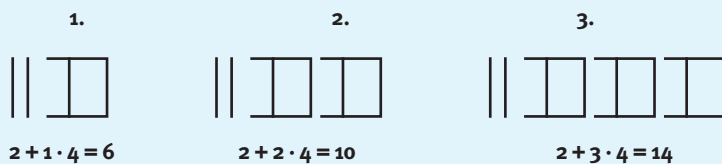
Vor dem Fortsetzen oder Ergänzen von Mustern dieser Art sollte der Zusammenhang zwischen Bild und zugehöriger Gleichung geklärt werden. Dazu wäre es möglich, verschiedene Gleichungen vorzugeben. Die Kinder müssen herausfinden, welche Gleichung zu welcher bildhaften Darstellung passt und dies begründen.

Beispiel: Welche Aufgabe passt zu welchem Bild? Ordne zu.



Die Muster können mit Hölzchen nachgelegt werden, um die Aufgaben zu finden bzw. zu erklären.

Ein Beispiel für eine ähnliche geometrische Folge kann sein:



- Die Kinder legen zuerst mit Hölzchen die vierte (und fünfte) Figur.
- Sie beschreiben das Bildungsgesetz der geometrischen Folge.
- Sie finden die Gleichung für die vierte Aufgabe.
- Sie beschreiben den Zusammenhang von geometrischen Figuren und Gleichungen.

Schließlich können die Kinder dazu animiert werden, selbst ähnliche geometrische Folgen zu erfinden und sie darzustellen. Diese legen sie anderen Kindern vor, die das Bildungsgesetz dann beschreiben und die Gleichungen dazu finden.

Aufgabenbeispiel VERA 8 Englisch

Die Aufgabe Vince Ford wurde im Jahr 2012 in VERA 8 im Fach Englisch im Bereich *Leseverstehen* eingesetzt. Es handelt sich dabei um ein Interview zum Thema Jugendliteratur. Die sieben Teilaufgaben unterscheiden sich in ihrer Schwierigkeit. Bei diesem Beispiel handelt es sich um eine geschlossene Zuordnungsaufgabe, eine richtige Schülerlösung ist dargestellt.

Aufgabe Vince Ford: Interview with a NZ children's author

Read the children's questions to a New Zealand author (0 to 7) and match them with the answers (a to k). There are two more answers than you need. There is an example at the beginning (0).



Name: Vince Ford
Date of birth: 12 April 1970
Place of birth: Eltham, Taranaki



0. What is your favourite food?
 1. Do you have a nickname, and if so, what is it?
 2. What do you do in your free time?
 3. What was the subject you hated most at school?
 4. What was the book you most loved as a child?
 5. What person from the past would you most like to meet?
 6. Do you have a special place where you write your books?
 7. What advice would you give to young writers?
- a) Write as much as you can as often as you can and write the truth – from your heart.
 - b) I play sports – I love touch rugby and kayaking, and I'm learning to surf now. I also dive for crayfish and kina.
 - c) Geography – it suited the way I thought.
 - d) Fairly quiet. I was a goody goody at primary school but became more laid back as I got older. I was always a bit of an individual.
 - e) Yes. My office looks out over Makarori Beach. It's beautiful.
 - f) I can't remember the title – it was about a hippopotamus race.
 - g) Yes – it's V. My name kept getting shortened. Now that's all that's left.
 - h) Math – it bored me.
 - i) Passionfruit crème brûlée. Next is oatmeal pancakes, banana and bacon.
 - k) Someone from the Clovis culture. They lived in the late Ice Age in America. They made stone spear points and hunted mammoths. I'm researching a book set in that period.

0	1	2	3	4	5	6	7
i	g	b	h	f	k	e	a

Picture & text: <http://christchurchcitylibraries.com/Kids/ChildrensAuthors/VinceFord.asp>, © Vince Ford

Didaktischer Kommentar zur abgebildeten Englisch-Aufgabe

In den didaktischen Kommentaren finden Lehrkräfte zunächst folgende Erläuterungen zu Aufgabenmerkmalen und zum Thema:

Die Aufgabe Vince Ford überprüft das detaillierte Leseverstehen auf unterschiedlichen Niveaus: Schülerinnen und Schüler, die die Teilaufgaben 1 und 2 richtig lösen, können vertraute Wörter (z. B. Bezeichnungen für Freizeitaktivitäten) und elementare Wendungen (z. B. *I love...*, *that's all...*, *I play...*) in einfachen alltäglichen Mitteilungen verstehen, was dem Niveau A1 (gemäß Gemeinsamer Europäischer Referenzrahmen) entspricht. Die Teilaufgaben 3, 4, 6 und 7 liegen auf dem Niveau A2, da kurze, einfache Texte zu einem vertrauten konkreten Thema verstanden werden sollen. So müssen z. B. positive und negative Wertungen voneinander unterschieden oder ein Ratschlag in seiner kommunikativen Funktion erkannt werden. Lernende, die Teilaufgabe 5 richtig lösen, sind in der Lage, einen einfachen argumentativen Text zu verstehen. Dies entspricht Niveau B1.

Anschließend wird auf mögliche Schwierigkeiten und Weiterarbeit eingegangen:

Die Aufgabe ist eine Multiple-Matching-Aufgabe, bei der den Fragen eines Interviews entsprechende Antworten zuzuordnen sind. Die Interviewfragen an den neuseeländischen Kinderbuchautor Vince Ford sind, wie auch seine Antworten, überwiegend kurz und unkompliziert. Inhaltlich entsprechen sie dem Erfahrungshorizont der Schülerinnen und Schüler. So wird der Autor z. B. nach seinem Spitznamen, seinen Freizeitaktivitäten oder seinem Lieblingsbuch gefragt. Das hierfür erforderliche Vokabular dürfte bereits aus dem Anfangsunterricht bekannt sein.

Der Text enthält nur wenige seltenere Begriffe, wie z. B. *kina* (Maori für „Seeigel“) in der Antwort a. Das Wort wird im Zusammenhang mit dem Verb *dive* und in einer Reihung mit *crayfish* verwendet. Dadurch ist es möglich zu erschließen, dass es etwas mit der Wasserwelt zu tun haben muss. Zudem ist es für die richtige Lösung nicht erforderlich, dass *kina* verstanden wird. Das Wort steht erst am Ende der Antwort, doch bereits im Satz davor wird deutlich, dass Vince Ford hier auf die Frage 2 nach seinen Freizeitaktivitäten antwortet. Gleich am Anfang nennt er verschiedene Sportarten, die den Schülerinnen und Schülern vertraut sein dürften.

Problematischer könnte hingegen der Begriff *Clovis* sein. Er ist sicher nicht nur Lernenden des achten Jahrgangs unbekannt. Zwar enthält die Antwort Sachinformationen über die *Clovis*-Kultur. Schwächere Lernende könnten sich jedoch dennoch durch fehlendes Wissen zu diesem Begriff so verunsichern lassen, dass sie nicht über den Anfang der Äußerung hinaus weiterlesen. Eine weitere Schwierigkeit ergibt sich auch durch die geringe Überlappung hinsichtlich des Fokus von Frage 5 und der richtigen Antwort k: Der Autor wird nach einer historischen Person gefragt, die er gern treffen würde. Seine Antwort *someone from the Clovis culture* nimmt aber Bezug auf eine Personengruppe aus der Eiszeit. Diese Information ist nicht ohne weiteres mit der aufgebauten Leseerwartung in Einklang zu bringen. Da die Antwort eine recht ausführliche Erläuterung und Begründung für Vince Fords Interesse an den Menschen der Clovis-Kultur darbietet, ist sie überdies länger als die Antworten zu den übrigen Fragen. Aus den genannten Gründen ergibt sich, dass die Teilaufgabe empirisch deutlich schwieriger ist als die anderen Teilaufgaben der Aufgabe.

Zur Lösung der Aufgabe können folgende Strategien angewandt werden:

- Vorwissen aktivieren und eine Leseerwartung aufbauen
- Anforderungen der Aufgabe ableiten
- angemessenes Vorgehen zur Lösung der Aufgabe wählen
- auf Schlüsselbegriffe konzentrieren
- auf wesentliche Informationen konzentrieren
- Bedeutung aus dem Kontext erschließen
- auf die Struktur der Fragen und Antworten achten

Für die Weiterarbeit im Unterricht werden Hinweise gegeben, denen Lehrkräfte folgen können, wenn Schülerinnen und Schüler die Aufgabe erfolgreich gelöst haben, aber auch Vorgehensweisen, wie Schülerinnen und Schüler bei Problemen mit der Aufgabe sinnvoll gefördert werden können.

Aufgabenbeispiel VERA 6 Deutsch

Die abgebildete Aufgabe enthält eine eingebettete Erzählung, die durch die Ankündigung einer Briefmarke und durch das Fußballgeschehen von 1908 und 2008 gerahmt wird. Zu dieser Aufgabe werden den Schülerinnen und Schülern insgesamt 12 Teilaufgaben zur Bearbeitung vorgelegt, die sich jeweils in ihrer Schwierigkeit und den geprüften Bildungsstandards aus dem Kompetenzbereich *Leseverstehen* unterscheiden. Die abgebildete Schülerlösung ist jeweils eine korrekte Lösung der Aufgabe. Die umfangreichen didaktischen Kommentare können aus Platzgründen an dieser Stelle nicht abgedruckt werden. Sie können auf den Schulservern der Kultusministerien der teilnehmenden Länder eingesehen werden.

Lies den folgenden Text.

Tor für Deutschland

- 1 *100 Jahre deutsche Fußball-Länderspiele*
In Basel fand 1908 das erste deutsche Fußball-Länderspiel statt – dort, wo am 7. Juni die Europameisterschaft 2008 angepfiffen wird. Deutschland würdigt beide Sportereignisse mit Sonderbriefmarken, die am 13. März bzw. am 10. April 2008 erscheinen.
- 5 Zahlreiche Fahrgäste drängten am 4. April 1908 auf den Bahnsteig des Frankfurter Bahnhofs, als der Zug nach Basel einlief. Unter ihnen war auch der 19-jährige Fritz Becker, der von allen nur „s’Beckerle“ genannt wurde. In seiner Manteltasche befand sich eine Einladung vom Deutschen Fußball-Bund (DFB), die er erst drei Tage zuvor erhalten hatte: Mit zehn anderen deutschen Fußballern sollte er bei einem „Fußball-Länderspiel“ gegen eine Schweizer Mannschaft antreten – welch eine verrückte Idee!
- 10 *Englische Fußblümmelei*
Fritz Becker war neugierig und abenteuerlustig. Also hatte der Schüler zugesagt – obwohl ihm das Fußballspielen eine ganze Menge Ärger einbrachte. Sein Direktor von der Oberrealschule Frankfurt hatte ihn erst vor kurzem dabei erwischt, wie er ohne schulische Genehmigung bei den Frankfurter Kickers Fußball gespielt hatte. Drei Stunden Karzer* hatte er deshalb aufgebrummt bekommen. Denn sein Direktor war der Ansicht, dass Fußball kein Sport für den Schüler einer höheren deutschen Lehranstalt war, sondern
15 eine „englische Fußblümmelei, bei der ein Haufen von Rohlingen mit den Beinen gegen einen Ball treten.“
- Bitte Smoking mitbringen*
Für die Länderspielreise war Becker sogar bereit, finanzielle Opfer auf sich zu nehmen. Denn der DFB bat in der Einladung nicht nur um die Mitnahme eigener Fußballschuhe, sondern auch eines Smokings für
20 das anschließende Bankett. Solch einen feinen Zwirn besaß der 19-Jährige nicht. Er musste ihn sich für 12 Mark leihen – viel Geld für einen Schüler.
Im Hotel „Metropol“ trafen sich die elf deutschen Nationalspieler am Tag vor dem Spiel zum ersten Mal. Zusammen gekickt hatten sie noch nie. Aus heutiger Sicht ebenso unvorstellbar ist die Tatsache, dass es keinen Trainer gab. Die Taktik bestimmte Mannschaftskapitän Arthur Hiller vom 1. FC Pforzheim.
- 25 *Deutsche Elf ist Mitfavorit*
Das Spiel der deutschen Nationalmannschaft gegen die Schweiz am 5. April ging mit 3:5 verloren. Und doch war es der Start einer Erfolgsgeschichte. Denn der Deutsche Fußball-Bund sollte zu einem der erfolgreichsten Fußball-Verbände der Welt werden. Deutsche Herren-Nationalmannschaften wurden drei Mal Weltmeister und drei Mal Europameister. Fußballer aus allen Spielergenerationen begeisterten die
30 Fans in aller Welt. Fußball „made in Germany“ war und ist mittlerweile wieder ein Qualitätssiegel.

- 35 Und so zählt die deutsche Mannschaft auch zu den Favoriten bei der Europameisterschaft 2008, die vom 7. bis 29. Juni in Österreich und in der Schweiz stattfinden wird. In Basel könnte sich damit ein Kreis schließen. Die Stadt, in der vor 100 Jahren das erste deutsche Fußball-Länderspiel angepfiffen wurde, ist Gastgeber eines der beiden EM-Halbfinals. Millionen Fans drücken der Mannschaft von Joachim Löw die Daumen, dass sie dort zu ihrem vorletzten Spiel bei der EM antreten wird, um dann in das Finale in Wien einzuziehen.

Schlechtes Ende eines historischen Tages

- 40 Doch zurück zum ersten Länderspiel 1908: Für Fritz Becker blieb die Fußball-Premiere in schlechter Erinnerung, und das, obwohl er zwei Tore, darunter das erste deutsche Länderspieltor überhaupt, schoss. Denn der 5. April 1908, dieser für alle deutschen Fußballfans historische Tag, endete für den jungen Spieler denkbar ungünstig: Beim Bankett schüttete ihm der Schweizer Torwart versehentlich Senf auf den geliehenen Smoking. Die Reinigung kostete 48 Mark. Fritz Becker fuhr nie wieder zu einem Länderspiel.

- 44 Die Sonderbriefmarke zur Fußball-EM 2008 erscheint am 13. März in der Serie „Für den Sport“.

* Karzer: Gefängniszelle in Schulen

© Postfrisch Das Philatelie-Journal, Deutsche Post, März/April 2008

Teilaufgabe 1.1

Welche Bezeichnungen ersetzen im Text den Namen Fritz Becker?
Schreibe zwei Bezeichnungen wörtlich aus dem Text heraus.

- 1) der Beckerle
- 2) der 19-jährige

Teilaufgabe 1.2

Vervollständige die Aussage. Fritz Becker steht im Mittelpunkt des Textes, weil er ...

das erste deutsche Länderspieltor schoss.

Teilaufgabe 1.3

Fritz Becker hatte Ärger mit seinem Direktor, weil er ...

- A ☐ sich in der Schule zu wenig anstrenge.
- B ☐ als Schüler einer höheren Lehranstalt Sport trieb.
- C ☒ ohne schulische Erlaubnis bei den Frankfurter Kickers spielte.
- D ☐ zu den Rabauken der Schule gehörte.

Teilaufgabe 1.12

Schreibe das Wort aus dem Text auf, auf das sich „feiner Zwirn“ bezieht.

auf Smoking



05 // Forschung

Als wissenschaftliche Einrichtung erfüllt das IQB nicht nur Serviceaufgaben für die Länder, sondern führt auch zahlreiche Projekte im Bereich der empirischen Bildungsforschung durch. Die im IQB bearbeiteten Forschungsthemen reichen von methodischen Herausforderungen der Kompetenzmessung bis zur Bestimmung von Faktoren, die den Erfolg von Lehr-Lern-Prozessen beeinflussen. Im Rahmen dieser Forschungsaktivitäten erhalten die wissenschaftlichen Mitarbeiterinnen und Mitarbeiter des IQB die Möglichkeit, sich weiterzuqualifizieren.

Struktur kognitiver Kompetenzen

Zum Kerngeschäft des IQB gehört die Messung von Kompetenzen. Hierfür ist es zunächst erforderlich, die zu messenden Kompetenzkonstrukte und ihre Struktur theoretisch zu definieren und in Form eines Kompetenzmodells zu beschreiben. Ein solches Modell macht präzise Angaben darüber, was aus theoretischer Sicht Bestandteil des jeweiligen Kompetenzkonstrukts ist. Es bildet die Grundlage für die Entwicklung von Testaufgaben, die zur Erfassung des Konstrukts eingesetzt werden können. Anschließend lässt sich empirisch prüfen, ob das Konstrukt durch die entwickelten Aufgaben angemessen operationalisiert wird.

War die Aufgabenentwicklung erfolgreich, müssten sich bestimmte Zusammenhangsmuster zeigen. So sollten die anhand der Aufgabenlösungen gebildeten Skalenwerte, die theoretisch sehr eng verwandte Kompetenzen abbilden (im Fach Deutsch z. B. *Lesen* und *Zuhören*), höher korrelieren als Skalenwerte, die sich auf konzeptuell unterschiedlichere Kompetenzen beziehen (z. B. *Lesen* und *Rechtschreibung*).

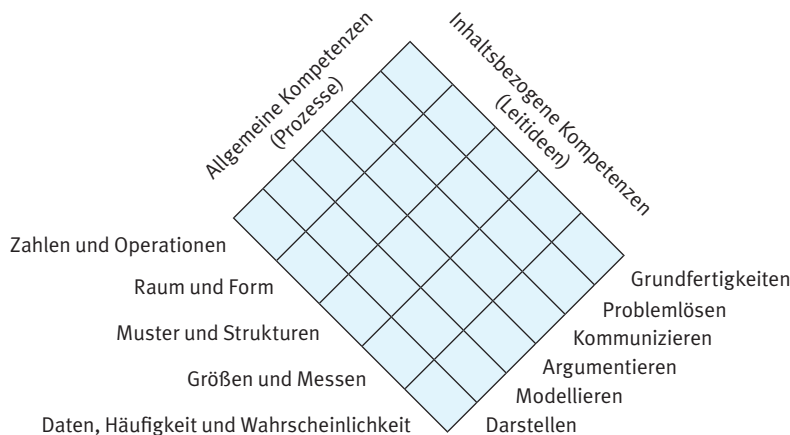
Zur genauen Beschreibung von Kompetenzen ist zudem die Klärung der Frage wichtig, welche theoretisch sinnvollen Bestandteile ein insgesamt homogenes Konstrukt bilden und welche Bestandteile eher getrennt untersucht werden sollten – also die Frage nach dem Aufbau und der Zusammensetzung von Kompetenzen. Hierzu werden *Struktur-* oder *Dimensionsanalysen* durchgeführt und *Kompetenzstrukturmodelle* empirisch überprüft. Dies umfasst auch die Unterscheidung von Teilkompetenzen, die zwar eng verwandt oder sehr ähnlich sind, die aber aus bestimmten (beispielsweise didaktischen) Gründen empirisch getrennt voneinander erfasst werden sollen. Zentral ist hierbei, inwieweit es sich bei den untersuchten Kompetenzen um unterscheidbare kognitive Fähigkeiten handelt und wie diese zusammenhängen.

Wenn sich die theoretisch angenommenen Zusammenhänge empirisch bestätigen, kann davon ausgegangen werden, dass die entwickelten Aufgaben die Konstrukte angemessen operationalisieren. Auf diese Weise ist gewährleistet, dass die Kompetenzen, die in den Ländervergleichsstudien und Vergleichsarbeiten untersucht werden, theoretisch und empirisch fundiert sind, und dass anhand der Daten möglichst präzise Aussagen darüber getroffen werden können, in welchen Kompetenzbereichen Schülerinnen und Schüler Stärken oder Schwächen aufweisen.

Am IQB wird im Forschungsbereich „Struktur kognitiver Kompetenzen“ zum einen untersucht, welche Zusammenhänge zwischen den verschiedenen Kompetenzbereichen innerhalb eines Faches bestehen und ob diese Kompetenzbereiche klar voneinander abgegrenzt werden können. Zum anderen wird erforscht, ob alle Teilkompetenzen innerhalb eines bestimmten Kompetenzbereichs ein homogenes Konstrukt bilden oder ob weitere Differenzierungen angezeigt sind. Ferner werden die Beziehungen zwischen der allgemeinen kognitiven Leistungsfähigkeit von Schülerinnen und Schülern, ihrem deklarativen Wissen und bestimmten fächerspezifischen Kompetenzbereichen analysiert.

In einer der Studien wurde die Struktur mathematischer Kompetenzen in der Primarstufe (vgl. KMK, 2005) untersucht (Winkelmann, Robitzsch, Stanat & Köller, 2012). In den Bildungsstandards der Kultusministerkonferenz werden für das Fach Mathematik im Primarbereich sechs prozessbezogene und fünf inhaltsbezogene Kompetenzen unterschieden (KMK, 2005 und Abbildung 5.1). Die prozessbezogenen mathematischen Kompetenzen beschreiben kognitive Operationen, die Schülerinnen und Schüler in allen Inhaltsbereichen der Mathematik anwenden müssen. Konkret handelt es sich bei den prozessbezogenen Kompetenzen um *Argumen-*

■ **Abbildung 5.1: Modell mathematischer Kompetenz der Bildungsstandards im Fach Mathematik für die Primarstufe**



tieren, Problemlösen, Modellieren, Kommunizieren, um das Technische Arbeiten und um Technische Grundfähigkeiten. Die inhaltsbezogenen Kompetenzen sind eng mit den klassischen mathematischen Stoffgebieten wie beispielsweise Arithmetik, Geometrie, Algebra und Stochastik verwandt. Konkret werden die Bereiche *Zahlen und Operationen, Größen und Messen, Raum und Form, Muster und Strukturen* sowie *Daten, Häufigkeit und Wahrscheinlichkeit* unterschieden. In der Studie von Winkelmann et al. (2012) wurde unter anderem untersucht, ob sich das komplexe Modell mathematischer Kompetenz mit sowohl prozessbezogenen als auch inhaltsbezogenen Dimensionen empirisch stützen lässt oder ob sich ein eindimensionales Modell mathematischer Kompetenz, in dem alle Dimensionen integriert und zu einem Globalfaktor zusammengefasst werden, als überlegen erweist.

Für die Dimensionsanalysen unserer Studie wurden Daten von rund 17 000 Schülerinnen und Schülern der dritten und vierten Jahrgangsstufe genutzt, die in den Jahren 2006 und 2007 an den Erhebungen zur Pilotierung und Normierung des IQB-Itempools für das Fach Mathematik in der Primarstufe teilgenommen haben. Aufgrund der großen Itemmengen (660 Items in der Pilotierung und 458 Items in der Normierung) bearbeiteten die Schülerinnen und Schüler jeweils nur kleine Teilmengen von Aufgaben (Multi-Matrix-Design). Für jedes Item lagen Antworten von mindestens 300 Schülerinnen und Schülern vor.

Geprüft wurden ein- und mehrdimensionale Rasch-Modelle, die anhand von Güteindizes verglichen werden können. Es zeigte sich, dass ein Fünf-Faktoren-Modell, das die fünf inhaltsbezogenen Kompetenzen abbildet, die empirischen Daten am besten beschreibt. Es erwies sich als überlegen sowohl im Vergleich zu einem eindimensionalen Modell, das einen Globalfaktor mathematischer Kompetenz abbildet, als auch im Vergleich zu einem sechsdimensionalen Modell prozessbezogener Kompetenzen.

Für die Bildungsstandards im Fach Mathematik im Primarbereich lassen sich die fünf inhaltsbezogenen Kompetenzen also analytisch trennen und beschreiben. Ob es allerdings in jedem Fall sinnvoll ist, diese separat zu messen und zu berichten, hängt von den Zielen der jeweiligen Untersuchung und den Möglichkeiten der eingesetzten Messinstrumente ab. So ist zur Beschreibung von Leistungsunterschieden auf der Ebene von Bildungssystemen, wie beispielsweise im Rahmen der IQB-Ländervergleiche, zumeist ein Globalwert ausreichend, der die Kompetenzausprägung im Fach Mathematik mit einem einzelnen Wert abbildet (vgl. Roppelt & Reiss, 2012).

Der Bildungspolitik und Schulpraxis können Ergebnisse von Strukturanalysen Anhaltspunkte dafür liefern, in welchen Bereichen eine differenzierte Förderung sinnvoll ist. Die empirische Diskriminierbarkeit der in den Bildungsstandards für das Fach Mathematik im Primarbereich definierten inhaltsbezogenen Dimensionen der Kompetenz zeigt, dass sich die Relevanz dieser Bereiche empirisch bestätigen lässt und über eine rein didaktische Veranschaulichung hinausgeht. Es ist im Unterricht also wichtig, diese spezifischen Kompetenzen gezielt zu fördern, um bei den Schülerinnen und Schülern eine solide Grundlage für das Mathematiklernen in den weiterführenden Schulen zu entwickeln und eine umfassende Basis für die lebenslange Auseinandersetzung mit den mathematischen Anforderungen des täglichen Lebens zu schaffen.

Heterogenität schulischer Kompetenzen

Ein weiterer Schwerpunkt der Forschungsaktivitäten des IQB liegt auf Analysen der Heterogenität schulischer Kompetenzen und möglicher Ursachen für ihre Entstehung. Im Mittelpunkt steht dabei die Frage, welche Faktoren den Lern- und Bildungserfolg von Schülerinnen und Schülern beeinflussen. Um Hinweise auf Ansatzpunkte für die Optimierung schulischer Kompetenzentwicklung zu erhalten, ist es erforderlich, solche Bedingungsfaktoren auf unterschiedlichen Ebenen zu bestimmen. Diese reichen von der Ebene des Schulsystems und dessen Steuerung bis hin zu Merkmalen des Unterrichts, der Lehrer-Schüler-Interaktion und der Schülerinnen und Schüler. In den vom IQB durchgeführten Schulleistungsstudien werden daher auch soziale und kulturelle Merkmale von Schülerinnen und Schülern, wie das Geschlecht oder die Herkunftssprache, erfasst. Unter anderem wird untersucht, inwieweit solche Hintergrundmerkmale beeinflussen, wie erfolgreich Schülerinnen und Schüler bestimmte Aufgabenstellungen bewältigen.

In einem Projekt des Forschungsbereichs „Heterogenität schulischer Kompetenzen“ haben Wissenschaftlerinnen und Wissenschaftler des IQB und der Humboldt-Universität zu Berlin in Kooperation mit dem Institut für Schulqualität der Länder Berlin und Brandenburg (ISQ) untersucht, ob Kinder nicht deutscher Herkunftssprache in den Vergleichsarbeiten der Grundschule (VERA 3) aufgrund von sprachlichen Anforderungen des Mathematiktests benachteiligt sind und inwieweit sprachliche Merkmale der Aufgaben eine mögliche Benachteiligung in den Mathematikleistungen erklären können (Haag, Heppt, Stanat, Kuhl & Pant, 2013).

Ob eine Testaufgabe bestimmte Schülergruppen benachteiligt, kann im Rahmen von Item-Response-Modellen untersucht werden. Dabei wird angenommen, dass die Lösung einer Aufgabe allein von der Fähigkeit (z. B. der Mathematikkompetenz) abhängt. Sonstige Eigenschaften der Person, die nicht zum gemessenen Konstrukt zählen, wie beispielsweise die Herkunftssprache oder das Geschlecht, sollten die Lösung der Aufgabe dagegen nicht beeinflussen. Die systematische Benachteiligung von Schülergruppen durch einzelne Testaufgaben wird anhand des sogenannten *differential item functioning* (DIF) untersucht. Dabei wird für jede einzelne Aufgabe bestimmt, ob DIF vorliegt oder nicht. Eine Aufgabe weist dann DIF auf, wenn es für eine bestimmte Gruppe von Schülerinnen und Schülern trotz gleicher Fähigkeit schwieriger ist, diese Aufgabe zu lösen als für alle anderen Schülerinnen und Schüler.

In unserer Studie haben wir untersucht, ob Schülerinnen und Schüler mit Zuwanderungshintergrund durch die sprachlichen Anforderungen von Mathematikaufgaben benachteiligt sind und welche sprachlichen Merkmale der Aufgaben die Benachteiligung erklären können.

Die Bestimmung sprachlicher Merkmale der Mathematikaufgaben basierte dabei auf dem Konzept der Bildungssprache. Unter Bildungssprache wird diejenige Sprache verstanden, die in der Schule verwendet wird, um Wissen zu erwerben und weiterzugeben. Es wird angenommen, dass sich diese in zentralen Merkmalen von der Alltagssprache unterscheidet (Bailey, 2007; Bailey & Butler, 2003; Cummins, 2000; Schleppegrell, 2004). Aus linguistischer Sicht zeichnet sich Bildungssprache unter anderem durch das gehäufte Auftreten komplexer grammatischer Strukturen sowie durch einen spezialisierten Wortschatz aus (z. B. Schleppegrell, 2004). Der bildungssprachliche Wortschatz lässt sich ausdifferen-

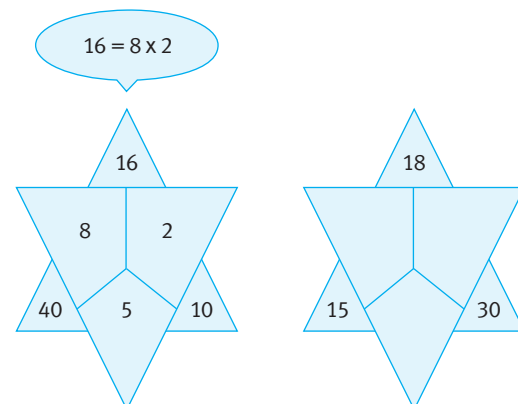
zieren in einen allgemeinen bildungssprachlichen Wortschatz, der fächerübergreifend eingesetzt wird (z. B. „vervollständigen“, „beschreiben“, „eine Übersicht erstellen“), und den fachspezifischen bildungssprachlichen Wortschatz, der zumeist nur in einem bestimmten Fach Verwendung findet (z. B. „multiplizieren“, „Rechenzeichen“, „ohne Rest teilen“). Es wird angenommen, dass der fachspezifische Wortschatz stärker als andere Merkmale der Bildungssprache im Unterricht erläutert wird und daher allen Kindern, unabhängig von ihrem Sprachhintergrund, in ähnlichem Maße zugänglich sein sollte. Für den allgemeinen Wortschatz hingegen wird erwartet, dass er Kindern nicht deutscher Herkunftssprache größere Probleme bereitet als Kindern deutscher Herkunftssprache (z. B. Komor & Reich, 2008). Für unsere Studie haben wir ein aus dem anglo-amerikanischen Sprachraum stammendes Kodierschema für die deutsche Sprache adaptiert (vgl. Bailey & Butler, 2003), um bildungssprachliche Merkmale von Testaufgaben zu erfassen (siehe Haag et al., 2013).

Die Datengrundlage der Studie bildete die im Frühjahr 2010 in Berlin durchgeführte Haupterhebung der Vergleichsarbeiten in der Grundschule (VERA 3). Es nahmen insgesamt 21 618 Drittklässlerinnen und Drittklässler teil, von denen nach Einschätzung der Lehrkräfte ca. ein Drittel zuhause vorwiegend eine andere Sprache als Deutsch sprach. Alle 56 im Fach Mathematik eingesetzten Testitems wurden von zwei geschulten Kodiererinnen hinsichtlich ihrer bildungssprachlichen Merkmale eingeschätzt. Unter anderem wurde die Länge der Aufgaben (Anzahl der Wörter), der Umfang des allgemeinen und mathematikspezifisch bildungssprachlichen Wortschatzes (siehe Abbildung 5.2 für ein Beispiel) sowie die Häufigkeit von Nominal- und Präpositionalphrasen erfasst. Es zeigte sich, dass diese vier Merkmale in den Testaufgaben im Vergleich zu anderen bildungssprachlichen Merkmalen am häufigsten auftraten.

■ Abbildung 5.2: Bildungssprachlicher Wortschatz in einer Aufgabe aus VERA 3 Mathematik

Immer zwei Zahlen im Inneren des Sterns werden **malgenommen**. Die **Ergebnisse** stehen in den Außenzacken. Setze die **fehlenden** Zahlen in den rechten Stern **ein**.

- fachspezifischer bildungssprachlicher Wortschatz
- allgemeiner bildungssprachlicher Wortschatz



Zudem deutete die Kodierung darauf hin, dass die sprachlichen Anforderungen der Aufgaben insgesamt eher niedrig waren, sich zwischen den Items jedoch unterschieden. Während einige Items relativ komplexe sprachliche Merkmale aufwiesen, waren andere weitgehend sprachfrei. Im Anschluss an die Kodierung wurde für jedes der 56 Testitems ein DIF-Wert berechnet. Dieser Wert beschreibt, wie stark sich die Lösungshäufigkeit bei Schülerinnen und Schülern mit deutscher und nicht deutscher Herkunftssprache für die jeweilige Aufgabe unterscheidet. Insgesamt fanden sich nur geringe Unterschiede in den Itemschwierigkeiten für Kinder deutscher und Kinder nicht deutscher Herkunftssprache unter Berücksichtigung ihrer mathematischen Kompetenz. So konnte bei lediglich zwei Items ein moderater DIF zum Nachteil der Kinder nicht deutscher Herkunftssprache festgestellt werden. Dies bedeutet, dass nur zwei Items für diese Schülerinnen und Schüler überproportional schwieriger waren als für Gleichaltrige deutscher Herkunftssprache.

In einem zweiten Schritt wurden die DIF-Kennwerte mit den bildungssprachlichen Merkmalen der Aufgaben in Beziehung gesetzt. Es zeigte sich, dass nahezu alle sprachlichen Merkmale mit dem DIF der Items zusammenhingen. Das Ergebnismuster deutet darauf hin, dass Aufgaben für Kinder nicht deutscher Herkunftssprache schwieriger sind, je mehr bildungssprachliche Merkmale sie enthalten, während dies für Kinder deutscher Muttersprache nicht in dem Maße der Fall ist. Für den mathematikspezifischen Wortschatz ließ sich dieser Zusammenhang dagegen nicht feststellen. Dies entspricht der Annahme, dass alle Kinder den mathematikspezifischen Wortschatz in vergleichbarer Weise im Unterricht erwerben, wohingegen andere bildungssprachliche Merkmale vermutlich seltener thematisiert und daher eher außerschulisch erlernt werden müssen, wofür Kinder deutscher Herkunftssprache mehr Gelegenheiten haben dürften (z. B. Komor & Reich, 2008).

Die Ergebnisse der Studie weisen darauf hin, dass sich bildungssprachliche Merkmale für Kinder deutscher und Kinder nicht deutscher Herkunftssprache unterschiedlich stark auf die Wahrscheinlichkeit auswirken, eine Mathematikaufgabe zu lösen. Eine substanzielle Benachteiligung von Kindern nicht deutscher Herkunftssprache, die auf die Testaufgaben zurückzuführen ist, konnte jedoch ausgeschlossen werden. Dass diese Kinder in Mathematik deutlich geringere Leistungen als Kinder deutscher Herkunftssprache erzielen, scheint also nicht auf sprachliche Anforderungen der Testaufgaben zurückzuführen sein.

Die sprachlichen Anforderungen des Unterrichts dagegen dürften zu den Leistungsunterschieden beitragen, die mit dem Sprachhintergrund der Kinder verbunden sind. Um die

Benachteiligung von Kindern nicht deutscher Herkunftssprache zu reduzieren, sollten daher systematisch Lerngelegenheiten für den Erwerb bestimmter bildungssprachlicher Merkmale – wie etwa dem allgemeinen bildungssprachlichen Wortschatz – geschaffen werden. Hierzu ist es erforderlich, Lehrkräfte für die Schwierigkeiten zu sensibilisieren, die Schülerinnen und Schüler nicht deutscher Herkunftssprache mit bildungssprachlichen Merkmalen haben, die der Alltagssprache zwar oberflächlich ähneln, im schulischen Kontext jedoch unterschiedliche Bedeutungen haben (wie beispielsweise die Verben „zeichnen“ vs. „kennzeichnen“ und „stellen“ vs. „erstellen“). Zudem sollten komplexere grammatikalische Strukturen im Unterricht thematisiert, erläutert und eingeübt werden.

In einem Folgeprojekt wird im Rahmen einer experimentellen Studie derzeit untersucht, inwieweit sich sprachliche Vereinfachungen von Mathematikaufgaben auf die Lösungswahrscheinlichkeiten für Kinder deutscher und nicht deutscher Herkunftssprache auswirken. Darüber hinaus wird geprüft, ob sich ähnlich differenzielle Muster in den Lösungswahrscheinlichkeiten auch für Testaufgaben im Fach Deutsch zeigen.

Methodische Fragen der Kompetenzdiagnostik

Gemäß dem Auftrag, die Bildungsstandards zu operationalisieren und ihr Erreichen zu überprüfen, wird am IQB eine Reihe großer Schulleistungstudien (*Large-Scale-Assessments*) durchgeführt und ausgewertet. Damit verbunden sind zahlreiche methodische Herausforderungen, denen angemessen zu begegnen ist, um valide Schlussfolgerungen aus den Studien ziehen zu können. Einige der daraus erwachsenden praktischen Fragestellungen werden im Forschungsbereich „Methodische Fragen der Kompetenzdiagnostik“ bearbeitet. Das IQB trägt mit Publikationen in diesem Bereich zur nationalen und internationalen Methodenentwicklung im Bereich des *Large-Scale-Assessments* bei.

Ein Teil der Studien beschäftigt sich mit der Entwicklung von Kompetenzstufenmodellen und der Festlegung von Schwellenwerten zwischen Kompetenzstufen. Einen zweiten Teilbereich bilden Arbeiten, die sich mit der Validität und Ökonomie der Durchführung von *Large-Scale-Assessments* befassen. Hierzu zählt beispielsweise die Schätzung der Bearbeitungszeiten von Items mit Hilfe statistischer Modelle, um Studiendesigns bezüglich der Anzahl der eingesetzten Items zu optimieren. Darüber hinaus wird in diesem Forschungsbereich untersucht, welchen Einfluss die Position eines Items innerhalb eines Testheftes auf seine Schwierigkeit hat. Ergebnisse dieser Analysen werden im folgenden Abschnitt exemplarisch vorgestellt.

Itempositionseffekte in Large-Scale-Assessments

Empirische Studien, die das Erreichen der Bildungsstandards überprüfen, verwenden in der Regel Modelle, die auf der Item-Response-Theorie (IRT) beruhen. Unter den verschiedenen IRT-Modellen wird dabei häufig das Rasch-Modell (Embretson & Reise, 2000; Hambleton, Swaminathan & Rogers, 1991) gewählt, da es gut zu interpretierende Ergebnisse liefert und eine Reihe von vorteilhaften Eigenschaften besitzt. Die Einfachheit des Modells ist jedoch notwendigerweise mit relativ strikten Annahmen verbunden, die in empirischen Anwendungen oft nur näherungsweise erfüllt sind. Das Rasch-Modell impliziert, dass die Wahrscheinlichkeit einer (korrekten) Antwort auf ein Item ausschließlich von der Schwierigkeit dieses Items und der Fähigkeit der Person, die dieses Item zu lösen versucht, abhängt. Tatsächlich kann diese Wahrscheinlichkeit aber noch von weiteren Faktoren auf Item- oder Personenseite beeinflusst werden, wie zum Beispiel der Position des Items im Testheft, der mittleren Schwierigkeit des Testhefts oder spezifischen Eigenschaften der Person, etwa ihrem Geschlecht oder Zuwanderungshintergrund.

Formal stellt die Nichtbeachtung solcher Faktoren eine Fehlspezifikation dar. Um die Auswirkungen von Modellverletzungen durch spezielle Itemeigenschaften auf die Messgüte und -präzision zu minimieren, werden komplexe, ausbalancierte Testdesigns verwendet. Ausbalancieren bedeutet dabei, mit Hilfe des Testdesigns sicherzustellen, dass die Verteilung nicht im Modell berücksichtigter Faktoren (z. B. die Position der Items) über die durch das Modell berücksichtigten Faktoren (z. B. Items) hinweg gleich ist. Sofern die Interaktion von Effekten vernachlässigt werden kann, ist in einem balancierten Design trotz des Auftretens von beispielsweise Itempositionseffekten keine Beeinträchtigung der Messgüte von Item- oder Personenparametern zu erwarten.

Um zu überprüfen, inwieweit einfache Modelle wie das Rasch-Modell unverzerrte Schätzungen von Parametern erlauben, ist es notwendig, in komplexeren IRT-Modellen zu prüfen, ob die oben beschriebenen Voraussetzungen für valide Schätzungen (z. B. keine Interaktion der Parameter) gegeben sind. Eine Möglichkeit hierfür bieten *linear logistische Testmodelle*, die es erlauben, Zufallseffekte auf der Item- und Personenseite zu modellieren. Mit Hilfe eines solchen Modells haben wir untersucht, ob sich in der Primarstufe für den Kompetenzbereich *Lesen* im Fach Deutsch im IQB-Ländervergleich 2011 Itempositionseffekte zeigen und wie stark diese gegebenenfalls ausfallen (Weirich, Hecht & Böhme, im Druck).

Hierfür wurde aus der Gesamtstichprobe von über 25 000 Schülerinnen und Schülern eine Teilstichprobe von 4 000 Kindern gezogen, deren Testhefte bezüglich der bearbeiteten Items und Itempositionen vollständig balanciert waren. Die insgesamt 80 Items waren in insgesamt neun Blöcken gruppiert. Da jeder Block eine nominale Bearbeitungszeit von 20 Minuten hatte, resultierten für ein Testheft mit 80 Minuten Bearbeitungszeit vier mögliche Positionen, an denen ein Aufgabenblock im Testheft platziert werden kann. Ein vollständig balanciertes Design bedeutet dabei, dass jeder der neun Blöcke an jeder der vier Positionen mit derselben Häufigkeit auftritt, in diesem Falle 200-mal.

Für den Kompetenzbereich *Lesen* ergab sich ein nicht linearer Itempositionseffekt, das heißt die mittlere Schwierigkeit der Items stieg nicht gleichmäßig mit aufsteigender Blockposition an. Wurden den Kindern Items im zweiten von vier Blöcken zur Beantwortung vorgelegt, war ihre Schwierigkeit im Mittel um 10 Punkte höher als wenn diese Items im ersten Aufgabenblock enthalten waren. Auf der Skala des IQB-Ländervergleichs (Mittelwert $M = 500$, Standardabweichung $SD = 100$) entspricht eine solche Differenz einer zehntel Standardabweichung oder ungefähr einem Achtel der Breite einer Stufe im Kompetenzstufenmodell *Lesen*. Für den dritten Block fand sich eine im Mittel um 6 Punkte größere Schwierigkeit als für den ersten Block (vgl. Abbildung 5.3). Dieser im Vergleich zum zweiten Block geringere Unterschied lässt sich damit erklären, dass für die Kinder zwischen dem zweiten und dritten Block eine zehnminütige Pause vorgesehen war, in der sie sich ausruhen konnten. Für den vierten und letzten Block ergab sich eine im Mittel sogar um 24 Punkte größere Itemschwierigkeit (ca. ein Drittel der Breite einer Kompetenzstufe). Diese Effekte waren dabei weitgehend stabil über alle Items, hingen also nicht von Merkmalen des Items (Multiple-Choice-Item, offenes oder halboffenes Itemformat; Schwierigkeit des Items) ab.

Abbildung 5.3 illustriert den Anstieg der Itemschwierigkeit in Abhängigkeit von der Position für ein hypothetisches Item, das eine Itemschwierigkeit von 500 Punkten aufweist, wenn es im ersten der vier Blöcke eines Testheftes positioniert ist. Legt man für dieses Item eine Lösungsquote von 50 Prozent für die erste Blockposition zugrunde, so würden nur noch 47,5 Prozent der Kinder das Item lösen, wenn es an zweiter Blockposition auftritt. Für die dritte Blockposition ergäbe sich eine Lösungsquote von 48,5 Prozent und für die vierte nur noch von 44 Prozent.

Um die Zuverlässigkeit dieser Schätzungen zu überprüfen, wurde in einem zweiten Teil der Untersuchung mit einer Simulationsstudie getestet, inwieweit das Modell Positionseffekte mit einer vorab definierten Effektstärke tatsächlich

identifizieren kann. In einer solchen Simulationsstudie werden zunächst zufällige Daten auf der Grundlage eines Modells mit vorgegebenen Parametern erzeugt. Anschließend werden die generierten Daten analysiert und überprüft, inwieweit die in der Analyse geschätzten Modellparameter die vorgegebenen Modellparameter replizieren. Dabei sind vor allem zweierlei Größen von Interesse: Zum einen, mit welcher prozentualen Häufigkeit ein Modell fälschlicherweise einen Effekt anzeigt, der nicht existiert – wie oft also das Modell Positionseffekte ermittelt, obwohl die Daten ohne einen solchen Effekt erzeugt worden sind. Zum anderen, mit welcher prozentualen Häufigkeit ein Modell einen Effekt findet, der in den Daten tatsächlich vorhanden ist. Für die Simulationsstudie kam dasselbe Testdesign zum Einsatz wie für den IQB-Ländervergleich 2011. Es zeigte sich, dass das Modell Positionseffekte tatsächlich gut identifizieren kann: Gibt es diese Effekte nicht, zeigt das Modell (irrtümlicherweise) nur in 5 Prozent der Fälle einen Effekt an, was genau der nominellen Irrtumswahrscheinlichkeit entspricht. Sind hingegen solche Effekte vorhanden, findet das Modell sie auch in 40 bis 80 Prozent der Fälle, abhängig von der simulierten Effektstärke.

Eine Besonderheit der zuletzt genannten Befunde ist, dass diese Ergebnisse nur auftreten, wenn man vollständig ausbalancierte Designs verwendet. In lediglich teilweise balancierten Designs (in denen z. B. jedes Item an mehreren, nicht aber allen Positionen im Testheft auftritt) sinkt die Teststärke deutlich. Dies veranschaulicht, dass balancierte Designs nicht nur notwendig sind, um valide Itemparameter trotz des

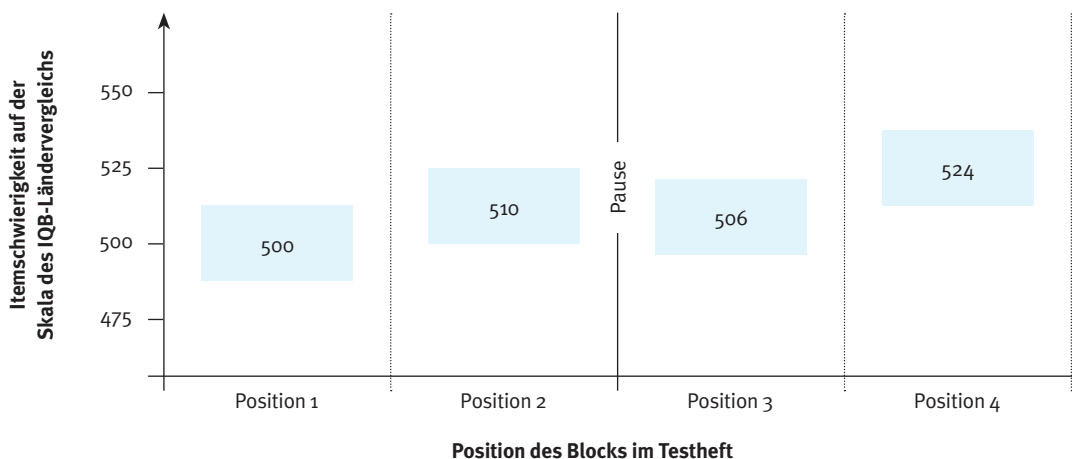
Auftretens von Kontexteffekten zu gewährleisten, sondern auch, um überhaupt schätzen zu können, ob und in welcher Weise solche Positionseffekte in relevantem Maße auftreten.

Zusammenfassend konnte die Studie also zum einen zeigen, dass Itempositionseffekte existieren und zum anderen ihre Größe quantifizieren. Die Ergebnisse verdeutlichen, dass die Schwierigkeit eines Items nicht ausschließlich von seinen inhärenten Eigenschaften abhängt, sondern auch vom Kontext des Items, etwa von seiner Position im Testheft. Dies ist zum Beispiel bei der Interpretation von Kompetenzstufen zu beachten, in deren Beschreibungen Kontexteffekte unberücksichtigt bleiben. Vorsicht ist insbesondere dann geboten, wenn keine ausbalancierten Studiendesigns zum Einsatz kommen, weil dann von der willkürlich gewählten Position eines Items im Testheft abhängen kann, ob dieses beispielsweise auf Kompetenzstufe III oder IV verortet wird. Die Studie gibt zum anderen aber auch Hinweise darauf, wie mit der Herausforderung von Itempositionseffekten umzugehen ist. Neben einem ausbalancierten Studiendesign und einer nicht zu langen Gesamtdauer des Tests wirken offenbar auch Pausen den Auswirkungen von Itempositionseffekten entgegen.

Entwicklung des R-Pakets „eat“

Neben den methodischen Forschungsprojekten entwickelt eine Arbeitsgruppe am IQB Prozeduren, um die komplexen Datenanalysen, die im Institut regelmäßig durchzuführen sind, zu strukturieren und zu vereinfachen. Die Routinen

■ **Abbildung 5.3: Nichtlinearer Anstieg der Itemschwierigkeit in Abhängigkeit von der Position des Items in einem Testheft mit vier Blöcken mit je 20-minütiger Bearbeitungszeit**



Anmerkung: Die Darstellung bezieht sich auf ein hypothetisches Item mit dem Referenzwert von 500 an der ersten Blockposition des Testhefts.

werden in der auf statistische Analysen ausgerichteten Programmiersprache *R* entwickelt und als R-Paket namens *eat* veröffentlicht¹. Das Paket unterstützt und automatisiert einerseits die Datenaufbereitung, die gerade bei großen Datensätzen nicht nur aufwändig, sondern auch fehleranfällig ist. Andererseits werden die technischen Schritte bei der Datenskalierung mit der IRT-Software *Conquest* (Wu, Adams, Wilson & Haldane, 2007) vereinfacht, indem das Paket als Schnittstelle für die Ein- und Ausgabe von *Conquest* dient und einen einfachen Zugang zu typischen Anwendungsfällen bietet. Darüber hinaus bündelt *eat* statistische Funktionen aus anderen R-Paketen zur Analyse von *Large-Scale-Assessment*-Daten. Das Paket hat sich sowohl bei den Auswertungen der IQB-Ländervergleiche der Jahre 2011 und 2012 als auch bei den Analysen der jährlichen Pilotierungsstudien zu VERA bewährt.

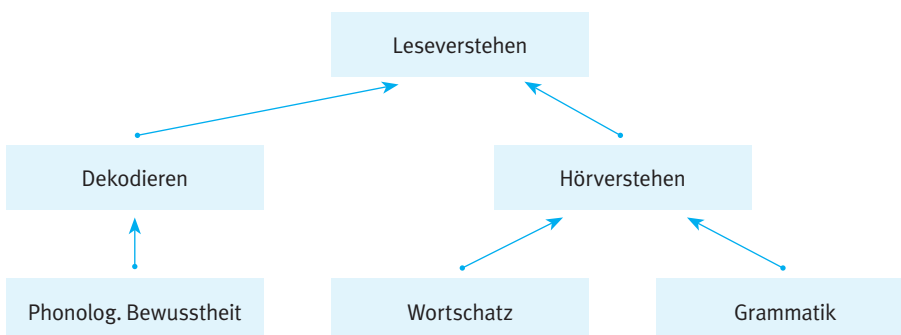
Erwerb und Förderung von Sprach- und Lesekompetenz

Die Schulleistungsstudien im Rahmen von PISA haben wiederholt belegt, dass ca. ein Fünftel der 15-Jährigen nur über sehr basale Lesefertigkeiten verfügt und davon auszugehen ist, dass diese Jugendlichen Schwierigkeiten haben werden, einfache Anforderungen an das Leseverständnis zu bewältigen, die sich ihnen in Alltags- und Ausbildungssituationen stellen. Dieser Befund hat sich im IQB-Ländervergleich 2009 für die Schülerinnen und Schüler der neunten Jahrgangsstufe bestätigt. Der Anteil an Jugendlichen mit schwachen Leseleistungen ist in der Gruppe der Schülerinnen und Schüler nicht deutscher Herkunftssprache noch deutlich größer. Diesen Kindern und Jugendlichen fehlt es oft an Lerngelegenhei-

ten für den Erwerb von Kompetenzen in der Instruktionssprache Deutsch, die für eine erfolgreiche Schullaufbahn notwendig sind. Ausgehend von dieser Befundlage wird in verschiedenen Projekten des Forschungsbereichs „Erwerb und Förderung von Sprach- und Lesekompetenz“ (BeFo, BiSpra, KuL, NEPS, DELKO) erforscht, wie sich Sprach- und Lesekompetenzen in den verschiedenen Bildungsinstitutionen (Kindergarten, Grundschule und Sekundarstufe I) entwickeln, wie man diese erfassen und gezielt fördern kann.

Im von der Deutschen Forschungsgemeinschaft (DFG) geförderten DELKO-Projekt (DEterminanten der Lesekompetenz) wurde in Kooperation mit Wissenschaftlerinnen und Wissenschaftlern der Universität Würzburg untersucht, welche spezifischen Kompetenzen mit der Fähigkeit, Texte verstehend zu lesen, zusammenhängen und ob sich dieses Zusammenhangsgefüge für Schülerinnen und Schüler deutscher und nicht deutscher Herkunftssprache unterscheidet (Marx et al., in Vorbereitung; Marx & Roick, 2012; Marx & Stanat, 2012). Nach dem *Simple View of Reading* (Hoover & Gough, 1990) wird das Leseverständnis vor allem durch die Dekodierfähigkeit, d. h. durch die effiziente Erkennung einzelner geschriebener Wörter, sowie durch das Hörverstehen beeinflusst (vgl. Abbildung 5.4). Diese Teilfertigkeiten wiederum werden dem Modell zufolge durch weitere Kompetenzen determiniert. Während für die Ausbildung der Dekodierfähigkeit vor allem die phonologische Bewusstheit (die Fähigkeit, lautliche Strukturen von Sprache zu erkennen) und das Arbeitsgedächtnis relevant sind, hängt das Hörverstehen primär von grammatikalischen Kenntnissen und vom Wortschatz ab.

■ Abbildung 5.4: Modell des Leseverstehens nach dem *Simple View of Reading* (Hoover & Gough, 1990)



¹ Die Entwicklungsversion des Pakets ist unter <https://r-forge.r-project.org/projects/eat/> abrufbar.

Das Hauptaugenmerk der DELKO-Studie lag auf der Frage, inwieweit Jugendliche deutscher und nicht deutscher Herkunftssprache in diesen Determinanten des Leseverstehens unterschiedliche Stärken und Schwächen aufweisen und worauf die im Durchschnitt geringere Lesekompetenz von Schülerinnen und Schülern nicht deutscher Herkunftssprache primär zurückzuführen ist. Da es im deutschsprachigen Raum an Verfahren zur Erfassung zentraler sprachlicher Kompetenzen von Jugendlichen in den Bereichen Grammatik, Hörverstehen und phonologische Fähigkeiten mangelt (vgl. Ehlich, 2005), zielte die Studie zudem darauf ab, entsprechende Testverfahren zu entwickeln und zu erproben. Untersucht wurde – dem *Simple View of Reading* folgend – die Rolle der phonologischen Bewusstheit, des Wortschatzes, der grammatikalischen Kompetenzen, der Dekodierfähigkeit, des Hörverstehens sowie (als weitere zentrale Determinante der Lesekompetenz) des Arbeitsgedächtnisses auf die Leseleistungen von Jugendlichen mit deutscher bzw. nicht deutscher Herkunftssprache. Damit sollte auch geprüft werden, ob sich ein theoretisches Modell, das für die Lesekompetenz von Heranwachsenden in einer Erstsprache entwickelt worden ist, auch auf Lesekompetenz in einer Zweitsprache anwenden lässt.

Für die DELKO-Studie wurde im Rahmen der PISA-2009-Erhebung eine zusätzliche Stichprobe von 15-Jährigen gezogen. Da die Studie vor allem auf die Erklärung schwacher Leseleistungen abzielte, konzentrierte sich die Stichprobenziehung auf Haupt-, Real- und Gesamtschulen. Diese Schülerinnen und Schüler bearbeiteten neben besonders einfachen PISA-Aufgaben auch Testverfahren zur Erfassung der genannten Determinanten der Lesekompetenz. Insgesamt nahmen an der Studie 479 Jugendliche (mittleres Alter 16,1 Jahre) teil, davon 179 mit deutscher Herkunftssprache, 186 mit türkischer Herkunftssprache und 114 mit einer anderen Herkunftssprache als Deutsch oder Türkisch. Die Ergebnisse des PISA-Lesekompetenztests bestätigen, dass es sich bei den Jugendlichen um sehr schwache Leserinnen und Leser handelt, die erhebliche Schwierigkeiten haben, einfache Texte zu verstehen. Im Mittel erreichten die Jugendlichen nicht die Kompetenzstufe II des PISA-Lesetests, die als Voraussetzung für einen erfolgreichen Übergang in eine Berufsausbildung angesehen wird. Dabei war die Lesekompetenz der Jugendlichen mit nicht deutscher Herkunftssprache im Durchschnitt noch geringer ausgeprägt als die der Jugendlichen deutscher Herkunftssprache.

Um das Bedingungsgefüge schwacher Leseleistungen für die verschiedenen Sprachgruppen auf Unterschiede zu prüfen, wurden Strukturgleichungsmodelle angewendet. Die Ergebnisse zeigen zunächst, dass sich die Gruppen in den Bereichen Leseverstehen, Hörverstehen, Dekodieren, Wortschatz

und Grammatik unterscheiden. Die Kompetenzen der Schülerinnen und Schüler deutscher Herkunftssprache waren in allen Bereichen höher ausgeprägt als die der Jugendlichen nicht deutscher Herkunftssprache. Diese Kompetenzunterschiede verschwanden für nicht deutschsprachige Jugendliche mit einer anderen Herkunftssprache als Türkisch weitgehend, wenn die soziale Herkunft der Schülerinnen und Schüler statistisch kontrolliert wurde. Die geringeren sprachlichen Kompetenzen dieser Jugendlichen scheinen demnach vornehmlich dadurch bedingt zu sein, dass sie im Durchschnitt aus sozial schwachen Familien stammen. Für Jugendliche mit Türkisch als Herkunftssprache zeigten sich hingegen auch unter Berücksichtigung ihrer sozialen Herkunft im Vergleich zu deutschsprachigen Jugendlichen weiterhin schwächere Leistungen in den betrachteten Kompetenzbereichen. Jugendliche türkischer Herkunftssprache weisen demnach persistierende Leistungs Nachteile in allen sprachlichen Determinanten der Lesekompetenz auf, die sich nicht allein auf die oftmals niedrigere soziale Herkunft zurückführen lassen.

Um weiterhin zu überprüfen, ob das postulierte Lesekompetenzmodell nach dem *Simple View of Reading* auch auf das Lesen in einer Zweitsprache übertragbar ist, wurde mit Mehrgruppenanalysen überprüft, inwieweit sich das Bedingungsmodell der Lesekompetenz in Abhängigkeit vom Sprachhintergrund der Jugendlichen unterscheidet. Dabei zeigte sich, dass unter statistischer Kontrolle der sozialen Herkunft der Jugendlichen kaum Unterschiede in den Zusammenhängen der Determinanten der Lesekompetenz zwischen den Sprachgruppen zu erkennen waren. In allen Gruppen war im Strukturgleichungsmodell sowohl der Einfluss des Dekodierens als auch der des Hörverstehens auf das Lesen signifikant. Der geschätzte Einfluss des Hörverstehens war dabei in allen Sprachgruppen ausgeprägter als der Einfluss des Dekodierens (Marx et al., in Vorbereitung). Dieses Muster entspricht der mehrfach bestätigten Annahme, dass mit zunehmendem Alter und Automatisierung des Dekodierens die Bedeutung des Hörverstehens für das Leseverständnis zunimmt.

Zusammenfassend weisen die Ergebnisse der DELKO-Studie darauf hin, dass die Zusammenhangsstruktur der untersuchten sprachlichen Determinanten mit der Lesekompetenz bei leseschwachen Jugendlichen deutscher und nicht deutscher Herkunftssprache vergleichbar ist. Somit scheint es prinzipiell möglich zu sein, Modelle der Lesekompetenz, die für das Lesen in einer Erstsprache konzipiert worden sind, auf das Lesen in einer Zweitsprache zu übertragen. Allerdings scheinen einige der sprachlichen Determinanten des Leseverstehens für Jugendliche nicht deutscher Herkunftssprache am Ende der Regelschulzeit noch eine besondere Herausforderung darzustellen, auch wenn man berücksichtigt, dass sie

oft aus sozial schwachen Familien stammen. Fördermaßnahmen, die auf eine Verbesserung des Leseverstehens besonders schwacher Leser abzielen, sollten demnach auch bei älteren Schülerinnen und Schülern sprachliche Vorläuferfähigkeiten des Lesens fokussieren. Die Befunde der DELKO-Studie deuten darauf hin, dass dabei für nicht deutschsprachige Jugendliche mit einer anderen Herkunftssprache als Türkisch eine Förderung der Wortschatzkenntnisse ausreichend sein könnte, während für Jugendliche mit türkischer Herkunftssprache eine umfassendere sprachliche Förderung erforderlich zu sein scheint.

Implementation

Die im Forschungsbereich „Implementation“ angesiedelten Studien untersuchen, ob und inwiefern die von der KMK eingeführten Bildungsstandards und die darauf basierenden Schulleistungsstudien zu Veränderungen im Verhalten und Erleben von schulischen Akteuren und des Unterrichts führen. Die Arbeiten am IQB hierzu lassen sich in zwei Teilbereiche gliedern: Im ersten Bereich gehen wir der Frage nach, wie die Bildungsstandards in der schulischen Praxis implementiert und wie Evaluationsdaten von Lehrkräften und Schulleitungen genutzt werden. Dazu gehören das Projekt „Schulen als Steuerungsakteure im Bildungssystem“ (StaBil) sowie Analysen zum Grad der kognitiven Auseinandersetzung mit den Bildungsstandards. Der zweite Teilbereich umfasst Studien, die Unterstützungsangebote zur Auseinandersetzung mit den Bildungsstandards umsetzen und evaluieren. Hierzu zählt die Evaluation des länderübergreifenden Projektes for.mat („Fortbildungskonzepte und -materialien zur kompetenz- bzw. standardbasierten Unterrichtsentwicklung“) sowie die Evaluation eines Fortbildungsprojektes am Landesinstitut für Schule und Medien Berlin-Brandenburg (LISUM). Nachfolgend werden Ergebnisse des Projekts StaBil exemplarisch für die Arbeit in diesem Forschungsschwerpunkt vorgestellt.

Das Verbundprojekt „StaBil“ untersuchte in Berlin und Brandenburg, wie Schulen Daten aus der Schulinspektion, der internen Evaluation, den Vergleichsarbeiten und zentralen Prüfungen nutzen und welche institutionellen Bedingungen die Nutzung einschränken oder unterstützen. Die Finanzierung erfolgte durch das Bundesministerium für Bildung und Forschung im Rahmen des Forschungsschwerpunktes Steuerung im Bildungssystem (SteBis) für den Zeitraum von 2010 bis 2013. Insgesamt bestand das Projekt aus vier Teilstudien, von denen zwei am IQB und zwei an der Freien Universität Berlin verortet waren. Die am IQB durchgeführten Teilstudien untersuchten, welche Zusammenhänge zwischen schulischer Evaluationspraxis und Schülerleistungen bestehen

und wie sich die Gestaltung von Rückmeldeformaten auf das Verständnis der zurückgemeldeten Ergebnisse auswirkt.

Die erste Teilstudie analysierte die schulische Evaluationspraxis im Rahmen einer flächendeckenden Online-Befragung von Schul- und Fachkonferenzleitungen in Berlin und Brandenburg. An der Erhebung im Schuljahr 2011/2012 beteiligten sich insgesamt 329 Schulleiterinnen und Schulleiter (Rücklauf: 27%) sowie 342 Fachkonferenzleiterinnen und Fachkonferenzleiter (Rücklauf: 14%) aus insgesamt 470 allgemein bildenden Schulen. Die ersten Analysen basieren auf Daten der Schulleitererhebung und beziehen sich auf die Einschätzung und Nutzung von Ergebnissen der schulinternen Evaluation, Schulinspektion sowie Vergleichsarbeiten (Wurster, Richter, Schliesing & Pant, 2013).

Die Ergebnisse weisen darauf hin, dass sich die Verfahren hinsichtlich des wahrgenommenen Aufwandes und Nutzens sowie der wahrgenommenen Diagnosegüte unterscheiden. Die Rückmeldungen aus der schulinternen Evaluation werden im Vergleich zu den beiden anderen betrachteten Verfahren als nützlicher und valider eingeschätzt, auch wenn die Durchführung des Verfahrens mit einem höheren Aufwand verbunden ist. Den mit der Schulinspektion und den Vergleichsarbeiten verbundenen Aufwand schätzen Schulleiterinnen und Schulleiter geringer ein, jedoch halten sie auch den Nutzen des Verfahrens für geringer. Im Ländervergleich halten Brandenburger Schulleiterinnen und Schulleiter die Evaluationsverfahren insgesamt für nützlicher und valider als Berliner Schulleiterinnen und Schulleiter. In beiden Ländern gibt die überwiegende Mehrheit der Befragten an, die Rückmeldungen in der Schulkonferenz und Lehrer-/Gesamtkonferenz zu besprechen.

Die Schulleitungen wurden zudem darüber befragt, für welche Zwecke sie die Ergebnisse von Schulinspektionen und Vergleichsarbeiten verwenden. Die Ergebnisse beider Evaluationsverfahren werden nach Auskunft der überwiegenden Mehrheit der Schulleitungen zur Optimierung der Lehr- und Lern-Prozesse im Unterricht genutzt. Darüber hinaus berichtet etwa die Hälfte aller befragten Schulleitungen, dass die Ergebnisse der Schulinspektion auch zur Weiterentwicklung der Schulkultur, zur Verbesserung des Schulmanagements, zur Personalentwicklung und zur schulischen Qualitätsentwicklung beitragen. Die Ergebnisse der Vergleichsarbeiten tragen aus der Perspektive der Schulleitung weiterhin stark zur kollegialen Unterrichtsentwicklung sowie teilweise zur Arbeit am Schulprogramm, zum Ausbau schulinterner Strukturen und zur Personalentwicklung bei. Damit macht die Studie deutlich, dass ein Großteil der befragten Berliner und Brandenburger Schulleitungen Evaluationsdaten für die Schul- und Unterrichtsentwicklung nutzt.

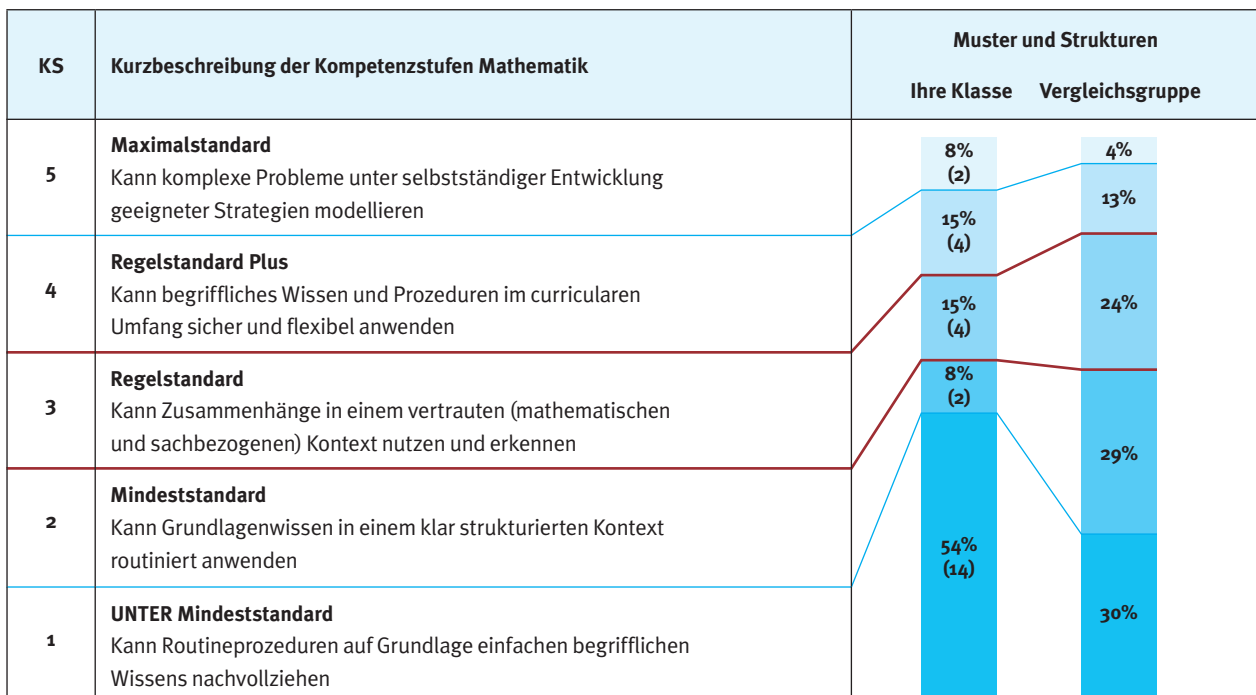
Die zweite am IQB angesiedelte Teilstudie untersuchte, wie Rückmeldungen aus Vergleichsarbeiten gestaltet sein sollten, damit sie von Lehrkräften verstanden werden. Zunächst wurden Gruppendiskussionen mit Lehrkräften an Brandenburger Grundschulen durchgeführt, um aus Sicht der Anwender Erkenntnisse über wichtige Gestaltungsmerkmale der Rückmeldungen zu gewinnen. Unter Berücksichtigung der Ergebnisse aus den Gruppendiskussionen und Befunden der Feedback-Forschung wurden anschließend verschiedene Rückmeldeformate entwickelt, die in einer experimentellen Studie von Lehrkräften hinsichtlich ihrer Verständlichkeit und Nutzbarkeit bewertet wurden. Die entwickelten Rückmeldeformate unterschieden sich in der *Textdarbietung* (mit vs. ohne Text), in der *Grafikanzahl* (drei vs. fünf) und in der *Darbietung von Lernaufgaben* (mit vs. ohne), die für die Weiterarbeit im Unterricht genutzt werden können. Diese Merkmale wurden jeweils miteinander kombiniert, sodass insgesamt acht verschiedene Rückmeldeformate zum Einsatz kamen. Mit einem Verständnistest wurde geprüft, inwieweit die Lehrkräfte die präsentierten Informationen erfasst haben. Ein Beispiel für die Grafiken, die in den untersuchten Rückmeldungen zur Anwendung kamen, ist in Abbildung 5.5 dargestellt. Die experimentelle Studie wurde im Rahmen einer Online-Befragung realisiert, an der insgesamt 412 Grundschullehrkräfte des Landes Nordrhein-Westfalen teilnahmen.

Die Studie zeigt, dass das Verstehen der Rückmeldung nicht vom erklärenden Text in den Rückmeldungen abhängt. Hinsichtlich der Grafikanzahl zeigte sich ein kleiner, statistisch bedeutsamer Effekt: Beispielrückmeldungen mit drei Grafiken wurden besser verstanden als solche mit fünf Grafiken. Die Ergebnisse weisen darauf hin, dass die für diese Studie ausgewählten Grafiken für Lehrkräfte insgesamt gut verständlich sind und ohne erklärende schriftliche Informationen auskommen. Dabei scheint der informative Gehalt von drei Grafiken auszureichen, um die Beispielrückmeldungen umfassend verstehen zu können. Ein Effekt der Darbietung von Lernaufgaben auf das Verständnis der Ergebnissrückmeldung konnte erwartungskonform nicht nachgewiesen werden.

Technologiebasierte Kompetenzdiagnostik

In der pädagogisch-psychologischen Diagnostik lässt sich eine langsame Verschiebung von herkömmlichen Papier-Stift-Testungen hin zu Testungen mit elektronischen Medien erkennen. Neben dem klassischen Computer kommen dabei auch vermehrt Tablet-PCs oder Smartphones zum Einsatz und auch das Internet wird einbezogen (Schroeders, Wilhelm & Schipolowski, 2010). Für diese Entwicklung spielen nicht nur administrative und ökonomische Überlegungen, wie zum

■ **Abbildung 5.5: Kompetenzstufenverteilung in Mathematik, die in einer Beispielrückmeldung im Rahmen der experimentellen Studie präsentiert wurde**





Beispiel die Einsparung von Testmaterial, eine Rolle, sondern auch inhaltliche Gesichtspunkte, wie etwa die Möglichkeit, Testverfahren durch Multimediainhalte anzureichern und zu optimieren. In der technologiebasierten Kompetenzdiagnostik lassen sich zwei Forschungsstränge voneinander abgrenzen: Zum einen wird versucht, bereits bestehende Messinstrumente möglichst unverändert auf elektronische Medien zu übertragen, um eine Testung zu realisieren, die mit der herkömmlichen Methode vergleichbar ist. Zum anderen wird angestrebt, durch interaktive Aufgaben oder die Verwendung von Videos die Kompetenzmessung effizienter zu gestalten oder inhaltlich anzureichern (Schroeders, Bucholtz, Formazin & Wilhelm, 2013). Mit dieser Forschungsstrategie ist meist auch die Hoffnung verbunden, mit Papierverfahren nicht messbare Konstrukte wie komplexes Problemlösen einer Messung zugänglich zu machen (Wilhelm & Schroeders, 2008). An dieser groben Zweiteilung des Forschungsfelds – der Vergleichbarkeit von Testleistungen und der Anreicherung der Messung – orientieren sich auch die Studien, die im Bereich der technologiebasierten Kompetenzdiagnostik am IQB angesiedelt sind. Exemplarisch soll an dieser Stelle ein Forschungsprojekt zur Vergleichbarkeit von Testleistungen genauer vorgestellt werden.

Durch die Digitalisierung von schulischen Lehrinhalten, den Einbezug des Computers in Lernprozesse und die Vernetzung von Personen und Wissensbeständen hat sich die Art und Weise, mit der Schülerinnen und Schüler Informationen sammeln, verarbeiten und bewerten, in den letzten Jahrzehnten grundlegend geändert. In Klassenzimmern kommen vermehrt Tablet-Computer oder Notebooks zum Einsatz und das Internet ist gegenüber traditionellen Informationsquellen wie einem Lehrbuch vielseitiger, insbesondere für Jugendliche häufig zugänglicher und oft auch ansprechender. Je prominenter der Einsatz von Computern, Tablets und Smartphones im Alltag und in der Schule werden, desto größer ist auch die Notwendigkeit, Computer im Rahmen von Schulleistungsstudien einzusetzen, damit sie schulischen wie außerschulischen Lernprozessen gerecht werden. Der Transfer auf ein neues Medium ist dabei weniger eine Innovation, die zukünftigen Entwicklungen vorweggreift, als eine Anpassung an tiefgreifende technologische und gesellschaftliche Veränderungen. Implizit ist mit einer Migration von Testaufgaben auf den Computer neben einer höheren Standardisierung bei der Testadministration auch die Hoffnung auf eine Kostenersparnis verbunden, mit der jedoch nur unter bestimmten Bedingungen zu rechnen ist (Farcot & Latour, 2009).

Aus der Perspektive des Bildungsmonitorings ist die prinzipielle Vergleichbarkeit von Leistungen über verschiedene Messinstrumente hinweg auch für die Validität von Trendaussagen wichtig. Wenn Kompetenzniveaus im Längsschnitt betrachtet werden sollen, muss sichergestellt werden, dass trotz eines etwaigen Wechsels des Testmediums Schülerleistungen, die auf einem Computer erzielt werden, mit Schülerleistungen, die zu einem früheren Zeitpunkt durch eine Papier-Stift-Testung ermittelt wurden, auf einer gemeinsamen Metrik verortet werden können. Anders formuliert: Die Aussagekraft der Messinstrumente sollte über die Zeit hinweg stabil sein und Aussagen erlauben, die unabhängig vom verwendeten Testmedium sind (Schroeders, 2009).

Am IQB wurde zur Äquivalenz von Leistungsmessungen über verschiedene Medien eine Reihe von Studien durchgeführt. Zentrale Fragen dieser auf die Äquivalenz von Testbedingungen fokussierenden Forschung lauten: Wird die Messung durch das Medium verändert oder ist die Messung über verschiedene Administrationsmodi invariant? Inwieweit können Testleistungen, die einerseits als Papier-Stift-Testung und andererseits computerisiert erhoben wurden, miteinander verglichen beziehungsweise ineinander umgerechnet werden? Ausgangspunkt war die umfangreiche Forschungsliteratur zum Vergleich von Papier-Stift- vs. Computer-Testungen, die zeigt, dass die Unterschiede in vielen Fällen nur gering sind (z. B. Mead & Drasgow, 1993). Computerisierte Messungen können jedoch nicht nur auf Notebook-Computern, sondern auch auf anderen elektronischen Medien wie Handhelds, Smartphones oder Tablet-PCs dargeboten werden. Diese neuen Medien eignen sich aufgrund ihres geringen Gewichts und der drahtlosen Datenübertragung besonders für den mobilen Einsatz im Klassenraum. Deshalb wurde in einer Studie der Frage nachgegangen, ob sich Smartphones zur Leistungsmessung eignen, und ob eine Messung der Fähigkeit zum schlussfolgernden Denken, die auf drei unterschiedlichen Medien – Smartphones, Notebooks und Papier-Stift – dargeboten werden, durch das Medium beeinflusst wird.

Um diese Frage zu beantworten, bearbeiteten 157 Schülerinnen und Schüler verschiedenartige Aufgaben zum schlussfolgernden Denken – einer Basisfähigkeit, die vielen kognitiven Prozessen zugrunde liegt – nacheinander auf allen drei Medien. Dabei wurden die Reihenfolge, in der die Medien bearbeitet wurden, und die Schwierigkeit der eingesetzten Aufgaben ausbalanciert. Die Aufgaben decken die drei Inhaltsbereiche *sprachlich*, *numerisch* und *figural* ab. Auf den Notebooks und den Smartphones kam dieselbe browsergesteuerte Programmierung zum Einsatz, so dass im Vergleich zur Papier-Stift-Variante in den computerisierten Bedingungen gewisse Aspekte der Messung ähnlicher

waren. Dazu gehörte beispielsweise, dass es keine Möglichkeit gab, bereits beantwortete Fragen nachträglich zu korrigieren, oder dass jeweils immer nur eine Aufgabe (anstatt mehrerer) präsentiert wurde. Andere Faktoren wie die Größe des Bildschirms oder die Art der Antworteingabe (Maus vs. Stylus/Eingabestift) unterschieden sich hingegen auch innerhalb der computerisierten Testungen, also zwischen der Notebook- und der Smartphone-Bedingung. Mit Hilfe geeigneter statistischer Verfahren, wie der konfirmatorischen Faktorenanalyse, konnte anschließend gezeigt werden, dass erstens das Testmedium nur einen geringen Einfluss auf die interindividuellen Unterschiede in der gemessenen Fähigkeit zum schlussfolgernden Denken hat. Zweitens liefern die Daten Hinweise darauf, dass die Messungen in den computerisierten Bedingungen untereinander nicht stärker zusammenhängen als die testmedienübergreifende Messung allgemein. Durch gezielte Modellvergleiche konnte die Vielzahl der Faktoren, die in früheren Studien zu testmedienspezifischen Unterschieden beitrugen, auf Unterschiede in den Wahrnehmungs- und motorischen Anforderungen bei der Antworteingabe eingeschränkt werden (Schroeders & Wilhelm, 2010).

Obwohl sich die Studienergebnisse mit der Mehrzahl der in der Literatur berichteten Befunde decken, muss dennoch die Vergleichbarkeit von Leistungen im Einzelfall – also im Grunde genommen auch bei Mehrfachtestung innerhalb eines Mediums – kritisch überprüft werden. Ob und wie sich Schwierigkeiten bei spezifischen Messinstrumenten und Hardware-Realisierungen manifestieren, ist nach derzeitigem Forschungsstand nicht zuverlässig zu prognostizieren. Deshalb ist es notwendig, entsprechende Äquivalenzstudien durchzuführen und identifizierte Unterschiede mittels geeigneter statistischer Analysemethoden zu berücksichtigen. Mit der Studie konnte jedoch auch eindrucksvoll belegt werden, dass sich Smartphones, die für die Weiterentwicklung technologiebasierter Erhebungen besonders interessant sind, prinzipiell zur Leistungsmessung eignen und auf hohe Akzeptanz bei den Studienteilnehmerinnen und Studienteilnehmern stießen.

Moderne Medien besitzen großes Potenzial, Bildungsprozesse nachhaltig zu verändern. Computer, Smartphones oder Tablets sind per se aber weder nützlicher als ein Buch, noch vereinfachen sie zwangsläufig Lernprozesse. Damit dies tatsächlich geschieht, müssen Konzepte erarbeitet werden, die die richtige Balance ihres Einsatzes mit Blick auf Motivation, Wissensvermittlung, Anwendung und Übung der Lerninhalte sowie Lernstandsüberprüfung finden. Die Computernutzung wird sich in den nächsten Jahren weiter verändern; neue Medien werden noch stärker Einzug in den Schulalltag halten. Eingedenk der derzeit häufig rezeptiv ausgerichteten

Mediennutzung von Jugendlichen und der bildungspolitischen Bedeutung moderner Medien muss es das Ziel sein, aus möglichst vielen Jugendlichen, die Medien zur reinen Unterhaltung nutzen, junge Erwachsene zu machen, die Medien instrumentell verwenden, um Lernaufgaben zu lösen, Bildungs Herausforderungen zu meistern und ihren schulischen und außerschulischen Interessen nachzugehen (Wilhelm & Schroeders, 2010).



Literatur

- Bailey, A. L. (Ed.). (2007). *The language demands of school. Putting academic english to the test*. New Haven: Yale University Press.
- Bailey, A. L. & Butler, F.A. (2003). *An evidentiary framework for operationalizing academic language for broad application to K-12 education. A design document*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Bitan-Friedlander, N., Dreyfus, A. & Milgrom, Z. (2004). Types of ‚teachers in training‘: The reactions of primary school science teachers when confronted with the task of implementing an innovation. *Teaching and Teacher Education*, 20(6), 607–619.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. et al. (2004). Mathematische Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47–92). Münster: Waxmann.
- Brunner, M. (2006). *Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität*. Dissertation, Humboldt-Universität zu Berlin.
- Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon: Multilingual Matters LTD.
- Ehlich, K. (2005). Sprachaneignung und deren Feststellung bei Kindern mit und ohne Migrationshintergrund: Was man weiß, was man braucht, was man erwarten kann. In K. Ehlich (Hrsg.), *Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund* (S. 3–75). Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, N.J.: L. Erlbaum Associates.
- Farco, M. & Latour, T. (2009). Transitioning to computer-based assessments: a question of costs. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment* (pp. 108–116). JRC Scientific and Technical Report 23306 EN.
- Fullan, M. (2000). The return of large-scale reform. *Journal of Educational Change*, 1, 5–28.
- Haag, N., Heppt, B., Stanat, P., Kuhl, P. & Pant, H. A. (2013). Second language learners' performance in mathematics: Disentangling the effects of academic language features. *Learning and Instruction*, 28, 24–34.
- Hall, G. E. & Hord, S. M. (2001). *Implementing change: patterns, principles, and potholes*. Needham Heights: Allyn and Bacon.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, Calif: Sage Publications.
- Hoover, W.A. & Gough, P.B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127–160.
- Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 139–190). Opladen: Leske + Budrich.
- KMK (2005). *Bildungsstandards im Fach Mathematik für den Primarbereich. Beschluss der Kultusministerkonferenz vom 15.10.2004*. München: Luchterhand.
- Köller, O. (1998). *Zielorientierungen und schulisches Lernen*. Münster: Waxmann.
- Komor, A. & Reich, H. H. (2008). Semantische Basisqualifikationen. In K. Ehlich, U. Bredel & H. H. Reich (Hrsg.), *Referenzrahmen zur altersspezifischen Sprachaneignung* (Bd. 29/I, S. 49–61). Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Lundberg, I. (2002). The child's route into reading and what can go wrong. *Dyslexia*, 8, 1–13.
- Marx, A. & Roick, T. (2012). Prädiktoren des Hörverstehens bei Jugendlichen deutscher und Jugendlichen nichtdeutscher Herkunftssprache. *Zeitschrift für Pädagogische Psychologie*, 26(2), 121–134.
- Marx, A. & Stanat, P. (2012). Reading comprehension of immigrant students in Germany: research evidence on determinants and target points for intervention. *Reading and Writing: An Interdisciplinary Journal*, 25(8), 1929–1945.
- Marx, A., Stanat, P., Roick, T., Segerer, R., Schneider, W. & Marx, P. (in Vorbereitung). *Determinants of reading in first-language and second-language students*.
- Mead, A. D. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449.
- Pant, H. A., Vock, M., Pöhlmann, C. & Köller, O. (2008a). Eine modellbasierte Erfassung der Auseinandersetzung von Lehrkräften mit den länderübergreifenden Bildungsstandards. In E.-M. Lankes (Hrsg.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 245–260). Münster: Waxmann.
- Pant, H. A., Vock, M., Pöhlmann, C. & Köller, O. (2008b). Offenheit für Innovationen: Befunde aus einer Studie zur Rezeption der Bildungsstandards bei Lehrkräften und Zusammenhänge mit Schülerleistungen. *Zeitschrift für Pädagogik*, 54(6), 827–845.
- Rogers, E. M. (1995). *Diffusion of Innovations*. New York: The Free Press.
- Roppelt, A. & Reiss, K. (2012). Beschreibung der im Fach Mathematik untersuchten Kompetenzen. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 34–48). Münster: Waxmann.

- Schlepppegrell, M. J. (2004).** *The language of schooling. A functional linguistics perspective.* Mahwah, NJ: Erlbaum.
- Schlepppegrell, M. J. (2012).** Academic language in teaching and learning. *The Elementary School Journal*, 112, 409–418.
- Schroeders, U. (2009).** Testing for equivalence of test data across media. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment. Lessons learned from the PISA 2006 computer-based assessment of science (CBAS) and implications for large scale testing.* JRC Scientific and Technical Report (pp. 164–170). Luxembourg: Office for Official Publications of the European Communities.
- Schroeders, U., Bucholtz, N., Formazin, M. & Wilhelm, O. (2013).** Modality specificity of comprehension abilities in the sciences. *European Journal of Psychological Assessment*, 29(1), 3–11.
- Schroeders, U. & Wilhelm, O. (2010).** Testing reasoning ability with handheld computers, notebooks, and paper and pencil. *European Journal of Psychological Assessment*, 26(4), 284–292.
- Schroeders, U., Wilhelm, O. & Schipolowski, S. (2010).** Internet-based ability testing. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced Methods for Conducting Online Behavioral Research* (pp. 131–148). Washington, DC: American Psychological Association.
- Weirich, S., Hecht, M. & Böhme, K. (im Druck).** Modeling Item Position Effects Using Generalized Linear Mixed Models. *Applied Psychological Measurement*.
- Wilhelm, O. & Schroeders, U. (2008).** Computerized ability measurement: *Some substantive dos and don'ts.* In F. Scheuermann & A. G. Pereira (Eds.), *Towards a research agenda in computer-based assessment. Challenges and needs for European Educational Measurement.* JRC Scientific and Technical Report (pp. 76–84). Luxembourg: Office for Official Publications of the European Communities.
- Wilhelm, O. & Schroeders, U. (2010).** Bildung in Zeiten neuer Medien aus kognitionspsychologischer Sicht. In J.-D. Gauer & J. Kraus (Hrsg.), *Bildung und Unterricht in Zeiten von Google und Wikipedia* (S. 27–45). Sankt Augustin: Konrad-Adenauer-Stiftung.
- Winkelmann, H., Robitzsch, A., Stanat, P. & Köller, O. (2012).** Mathematische Kompetenzen in der Grundschule. Struktur, Validierung und Zusammenspiel mit allgemeinen kognitiven Fähigkeiten. *Diagnostica*, 58(1), 15–30.
- Wu, M., Adams, R. J., Wilson, M. R. & Haldane, S. (2007).** *ACER ConQuest 2.0 - Generalized item response modelling software.* Camberwell: ACER.
- Wurster, S., Richter, D., Schliesing, A. & Pant, H. A. (2013).** Nutzung unterschiedlicher Evaluationsdaten an Berliner und Brandenburger Schulen. Rezeption und Nutzung von Ergebnissen aus Schulinspektion, Vergleichsarbeiten und interner Evaluation im Vergleich. *Die Deutsche Schule*, 12. Beiheft, 19–50.



06 // Der Beitrag des IQB zum wissenschaftlichen Austausch Das FDZ und die Koordinierungs- stelle

Das Forschungsdatenzentrum am IQB sowie die Koordinierungsstelle für die Zusammenarbeit zwischen dem IQB und dem Zentrum für Internationale Bildungsvergleichsstudien (ZIB) schaffen Synergien für die erfolgreiche Bearbeitung wissenschaftlicher Fragestellungen der Schulleistungsforschung.

Das Forschungsdatenzentrum (FDZ) am IQB

In den letzten 15 Jahren sind in Deutschland im Rahmen von nationalen und internationalen Schulleistungsstudien (z. B. PISA, IGLU/PIRLS, TIMSS, DESI und den Ländervergleichsstudien des IQB) umfangreiche Datensätze erhoben worden, die Kompetenzdaten und Hintergrundinformationen enthalten. Diese Datensätze bieten ein erhebliches Potenzial für wissenschaftliche Analysen, das nur dann umfassend genutzt werden kann, wenn Wissenschaftlerinnen und Wissenschaftler unterschiedlicher Disziplinen Zugang zu den Daten erhalten. Um dies zu gewährleisten, wurde 2007 das Forschungsdatenzentrum (FDZ) am IQB eingerichtet, das drei Aufgaben erfüllt. Das FDZ stellt erstens die Datensätze großer Bildungsstudien in Deutschland der *Scientific Community* für Re- und Sekundäranalysen zur Verfügung, es bietet zweitens Veranstaltungen zur methodischen Weiterbildung des wissenschaftlichen Nachwuchses an und es setzt sich drittens für den Ausbau der Forschungsdateninfrastruktur im Bildungsbereich ein.

In den letzten Jahren ist es gelungen, das FDZ zu einer häufig genutzten und anerkannten Einrichtung der deutschen Bildungsforschung und der informationellen Forschungsdateninfrastruktur zu entwickeln. Im Jahr 2009 ist die Arbeit des FDZ von einer wissenschaftlichen Gutachterkommission extern evaluiert worden, die eine Verstärkung des Angebots empfohlen hat. Nachdem das FDZ zunächst ausschließlich vom Bundesministerium für Bildung und Forschung (BMBF) finanziert wurde, ist es seit Oktober 2011 an die von Bund und Ländern kofinanzierte Stelle zur Koordinierung der wissenschaftlichen Aktivitäten des IQB und des Zentrums für Internationale Bildungsvergleichsstudien (ZIB) angekoppelt.

Archivierung, Dokumentation und Bereitstellung von Daten

Die Hauptaufgabe des FDZ am IQB besteht darin, die Daten aus nationalen und internationalen Bildungsstudien zu dokumentieren und zu archivieren, um diese für wissenschaftliche Re- und Sekundäranalysen verfügbar zu machen. Damit wird sichergestellt, dass die mit erheblichem finanziellen und intellektuellen Aufwand erzeugten Datensätze von Wissenschaftlerinnen und Wissenschaftlern verschiedener Fachrichtungen, wie etwa der Erziehungswissenschaft, der Psychologie, der Soziologie, den Fachdidaktiken und der Bildungsökonomie, vertiefend ausgewertet und wissenschaftlich optimal genutzt werden können. Das FDZ folgt in seiner Arbeit den Richtlinien des Rats für Sozial- und Wirtschaftsdaten (RatSWD) und ist ständiges Mitglied des RatSWD-Ausschusses Forschungsdateninfrastruktur (FDI), in dem zurzeit etwa 25 Forschungsdaten- und Datenservicezentren aus Deutschland vertreten sind.

Verfügbare Datensätze im FDZ am IQB

Am FDZ sind aktuell Datensätze aus zwölf Studien verfügbar (siehe Tabelle 6.1). Insbesondere in der Aufbauphase des FDZ stieß die Idee, der *Scientific Community* Datensätze zur Verfügung zu stellen, auf Seiten der Datenproduzenten zunächst auf Zurückhaltung, teilweise auch auf Widerstände. Um einerseits das Vertrauen der Datenproduzenten zu erhöhen und um andererseits dem Bedürfnis potenzieller Datennutzer nach möglichst offenem Datenzugang zu entsprechen, entwickelte das FDZ Regelungen, die den Anliegen beider Seiten gerecht werden sollten. Diese sehen unter anderem vor, dass Datenproduzenten bei der Übergabe von Datensätzen Forschungsfragen angeben können, denen im Rahmen von laufenden Qualifikationsarbeiten aktuell noch nachgegangen wird. Solche „gesperrten“ Fragestellungen dürfen innerhalb eines festgelegten Zeitraums nicht durch andere Wissenschaftlerinnen und Wissenschaftler mit dem jeweiligen Datensatz bearbeitet werden. Ferner unterstützt das FDZ potenzielle Datengeber auf Wunsch bei der Dokumentation ihrer Daten und orientiert sich dabei an den allgemeinen Richtlinien des *Council of European Social Science Data Archives* (CESSDA).

Die im FDZ archivierten Datensätze beinhalten die Ergebnisse von Kompetenzmessungen bei Kindern und Jugendlichen, vor allem in sprachlichen, mathematischen und naturwissenschaftlichen Bereichen. Zusätzlich wurden je nach Studie unter anderem lernrelevante Einstellungen und motivationale Merkmale sowie soziale, zuwanderungsbezogene und kulturelle Hintergrundmerkmale der Schülerinnen und Schüler erfasst. Neben den Test- und Befragungsergebnissen auf Schülerebene liegen zumeist auch Daten von Lehrkräften sowie teilweise von Eltern und Schulleitungen vor. Dadurch wird es möglich, individuelle, familiäre und schulische Bedingungen von Schülerleistungen gemeinsam und vertiefend zu analysieren. Die am FDZ bislang vorhandenen Datensätze sind überwiegend querschnittlich angelegt, mit den PISA-Daten zum Beispiel können jedoch zum Beispiel anhand der drei aktuell verfügbaren Wellen auch Trendanalysen durchgeführt werden. Längsschnittlich ausgerichtete Forschungsfragen auf Schülerebene lassen sich derzeit mit Daten der Studien ELEMENT und DESI sowie in Kürze mit den Daten des BIKS-Projekts untersuchen.

Tabelle 6.1: Verfügbare Datensätze im FDZ des IQB (Stand: Dezember 2013)

Studie	Federführung	Erhebungs- zeitraum	Anzahl Schülerinnen und Schüler	Jahrgangs- stufe	Beteiligte Bundesländer	Kompetenzbereiche bzw. Fächer
DESI	Prof. Dr. Eckhard Klieme (DIPF)	2003/2004	11 000	9	alle Länder	Deutsch, Englisch
ELEMENT	Prof. Dr. Dr. Rainer Lehmann (HU Berlin)	2003-2005	5 000	4-6	Berlin	<i>Lesen</i> , Mathematik, Englisch
IGLU	Prof. Dr. Wilfried Bos (TU/IfS Dortmund)	2001	10 500	4	IGLU-I: alle Länder IGLU-E: BW, BY, BB, HB, HE, NW, TH	<i>Lesen</i> , Mathematik, Naturwissenschaften
	Prof. Dr. Wilfried Bos (TU/IfS Dortmund)	2006	7 900			
					IGLU-I + E: alle Länder	
IQB- Länder- vergleich Sprachen	Prof. Dr. Olaf Köller (IQB Berlin)	2008/2009	39 500	9	alle Länder	<i>Lesen, Hören, Schreiben</i> in Deutsch, Englisch und Französisch
MARKUS	Prof. Dr. Reinhold S. Jäger und Prof. Dr. Andreas Helmke (Universität Koblenz-Landau)	2000	40 000	8	Vollerhebung Rheinland-Pfalz	Mathematik
PISA	Prof. Dr. Jürgen Baumert (MPIB Berlin)	2000	36 000	9- bzw. 15-Jährige	alle Länder	<i>Lesen</i> , Mathematik, Naturwissenschaften
	Prof. Dr. Manfred Prenzel (IPN Kiel)	2003	46 000			
	Prof. Dr. Manfred Prenzel (IPN Kiel)	2006	39 500			
	Prof. Dr. Eckhard Klieme (DIPF, Frankfurt a. M.)	2009	10 000			
QUASUM	Prof. Dr. Dr. Rainer Lehmann (HU Berlin)	1999	12 500	5 und 9	Brandenburg	Mathematik
TIMSS	Prof. Dr. Wilfried Bos (IFS Dortmund)	2007	5 200	4	alle Länder	Mathematik und Naturwissenschaften

Anmerkungen: DESI = Deutsch-Englisch-Schülerleistungen-International, ELEMENT = Erhebung zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin, IGLU = Internationale Grundschul-Leseuntersuchung, MARKUS = Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen (Schülerleistungen), Unterrichtsmerkmale, Schulkontext, PISA = Programme for International Student Assessment, QuaSUM = Qualitätsuntersuchungen an Schulen zum Unterricht in Mathematik, TIMSS = Trends in International Mathematics and Science Study 2007.

Zugang zu den Daten des FDZ am IQB

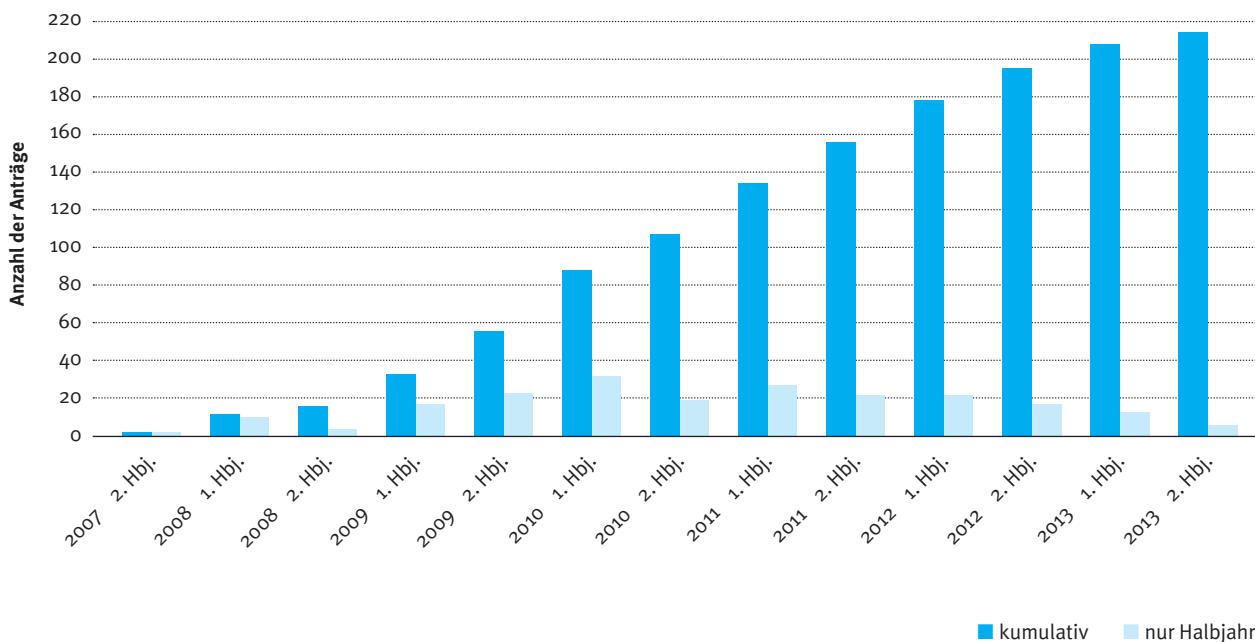
Interessierte Wissenschaftlerinnen und Wissenschaftler erhalten auf Antrag über ein Online-Formular Zugang zu den Daten der gewünschten Studie. Wird kein Zugang zu den Länderkennungen beantragt, kann das FDZ den Antrag eigenständig prüfen und genehmigen. In diesem Fall werden die Datensätze als *Scientific-Use-Files* auf CD-ROM zur Verfügung gestellt. Sind neuartige und bislang nicht publizierte Vergleiche zwischen den Ländern der Bundesrepublik Deutschland vorgesehen, so kann das FDZ nicht selbst über die Genehmigung des Antrags entscheiden. Die Antragstellerinnen und Antragsteller werden in diesem Fall aufgefordert, ihr Analysevorhaben so genau zu beschreiben, dass dieses in einem wissenschaftlichen Peer-Review begutachtet werden kann. Die Vorgaben hierfür orientieren sich an den Verfahren für DFG-Anträge. Werden die Forschungsvorhaben in den Gutachten befürwortet, genehmigt der Vorstand (bzw. ab 2014 das Kuratorium) des IQB den Antrag. Im Falle uneinheitlicher Begutachtungsergebnisse legt der IQB-Vorstand der KMK einen Entscheidungsvorschlag vor.

Bei Anträgen mit Bundesländervergleichen werden die Daten den Antragstellern entweder vor Ort im FDZ in Berlin oder über ein Fernrechensystem zur Verfügung gestellt. Für das Fernrechnen wird am FDZ ein webbasiertes *Job-Submission-System* des Instituts zur Zukunft der Arbeit (IZA) in Bonn eingesetzt. So können auch Forscherinnen und Forscher, die nicht im Raum Berlin ansässig sind, auf besonders geschützte Daten Zugriff erhalten. Mit dieser Vorgehensweise ist es möglich, die Berechnungen der Antragstellerinnen und Antragsteller nachzuvollziehen und sicherzustellen, dass die vertraglichen Vereinbarungen zwischen den Eigentümern der Daten und dem FDZ, das die Daten bereitstellt, eingehalten werden.

Entwicklung der Antragszahlen seit Einrichtung des FDZ

Nach der Einrichtung des FDZ im Jahr 2007 stieg die Anzahl der Datennutzungsanträge deutlich an. Seit 2010 wurden jährlich etwa 40 Neuanträge gestellt. Insgesamt gingen seit Einrichtung des FDZ 214 Datennutzungsanträge ein (siehe Abbildung 6.1).

■ Abbildung 6.1: Entwicklung der Datennutzungsanträge im FDZ



Am häufigsten werden die Datensätze der PISA-Erhebungen 2000, 2003 und 2006 beantragt, gefolgt von der IGLU-Studie 2001 und den ELEMENT-Daten (siehe Abbildung 6.2). Auch der IQB-Ländervergleich Sprachen in der Sekundarstufe I wird seit seiner Bereitstellung im Dezember 2011 stark nachgefragt.

Durch die Übersetzung der FDZ-Homepage ins Englische und die Präsentation des FDZ auf zahlreichen internationalen Tagungen ist es gelungen, auch internationale Wissenschaftlerinnen und Wissenschaftler auf das am IQB vorhandene Datenangebot aufmerksam zu machen. Insgesamt sind bislang 28 Anträge mit Beteiligung von Wissenschaftlerinnen und Wissenschaftlern an ausländischen Institutionen gestellt worden. Die Erstantragsteller aus dem Ausland sind überwiegend in der Schweiz tätig, gefolgt von den USA, ferner Schweden, den Niederlanden, Chile, Italien, Luxemburg und Österreich.

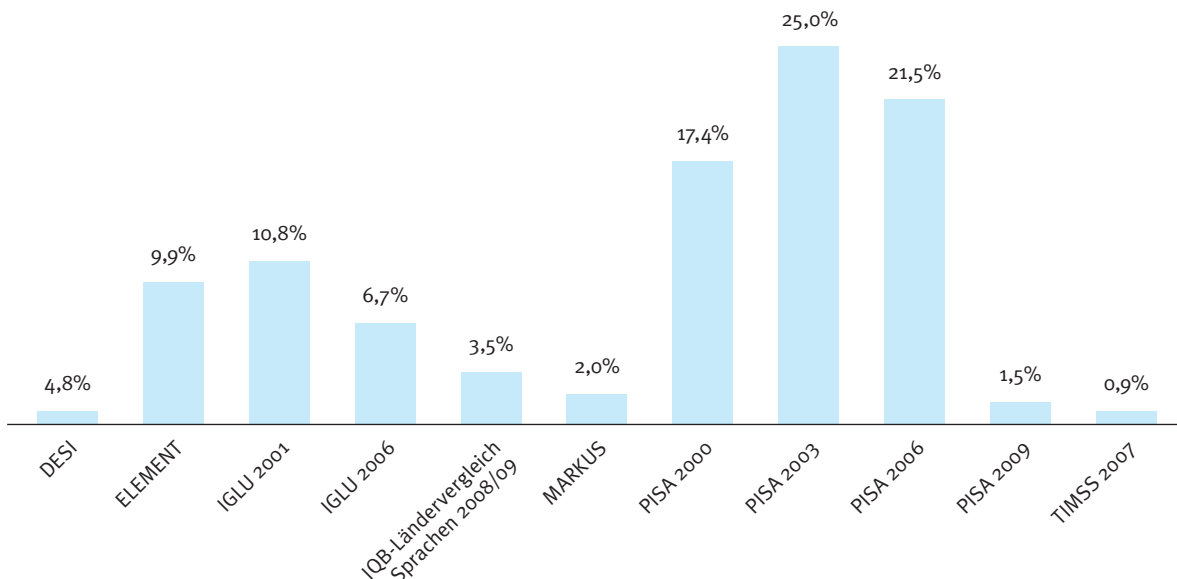
An 192 der bisher 214 gestellten Anträge auf Datenzugang waren keine am IQB beschäftigten Personen beteiligt, was für eine breite Wahrnehmung und Akzeptanz des FDZ innerhalb der *Scientific Community* spricht. Der in Zusammenarbeit mit dem IZA in Bonn angebotene Fernrechenservice wurde im Rahmen von insgesamt 70 Anträgen genutzt.

Weiterbildung des wissenschaftlichen Nachwuchses

Die Weiterbildung des akademischen Nachwuchses bildet einen weiteren zentralen Aufgabenbereich des FDZ. Im Frühjahr sowie im Herbst wird jährlich jeweils eine viertägige FDZ-Akademie zu Methoden der empirischen Bildungsforschung mit zwei parallelen Workshops pro Tag durchgeführt. Die Auswahl der Themen orientiert sich an aktuellen Strömungen und Neuerungen in der empirischen Bildungsforschung. Es wird dabei zum einen angestrebt, Workshops zu bereits etablierten Verfahren anzubieten, andererseits aber auch Inhalte zu behandeln, die innerhalb der *Scientific Community* noch wenig verbreitet sind. Die bislang angebotenen Workshops beschäftigten sich unter anderem mit den folgenden statistisch-methodischen Verfahren:

- Mehrebenenanalysen mit HLM, R oder Mplus
- Propensity Score Matching
- Multiple Imputation mit MICE
- Strukturgleichungsmodelle der Veränderungsmessung
- Logistische Regression
- IRT-Skalierung

■ **Abbildung 6.2: Verteilung der Anträge auf Datenzugang nach den im FDZ vorhandenen Datensätzen (N = 214)**



Anmerkungen: Es können mehrere Studien pro Antrag angefragt werden. DESI = Deutsch-Englisch-Schülerleistungen-International, ELEMENT = Erhebung zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin, IGLU = Internationale Grundschul-Leseuntersuchung, MARKUS = Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen (Schülerleistungen), Unterrichtsmerkmale, Schulkontext, PISA = Programme for International Student Assessment, QuaSUM = Qualitätsuntersuchungen an Schulen zum Unterricht in Mathematik, TIMSS = Trends in International Mathematics and Science Study 2007.

Das FDZ veranstaltet zudem regelmäßig Schulungen zur Unterstützung wissenschaftlicher Auswertungen der am FDZ verfügbaren Datensätze. So wurden Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftler beispielsweise im Umgang mit den Daten aus PISA und IGLU geschult. Diese Workshops sollen den Teilnehmerinnen und Teilnehmern inhaltliche Besonderheiten der Untersuchungen sowie mögliche Fragestellungen, die sich mit Daten der jeweiligen Studie bearbeiten lassen, näherbringen. Die Daten-Workshops gliedern sich dabei jeweils in einen Theorieteil und einen Praxisteil. Auf diese Weise erhalten die Teilnehmerinnen und Teilnehmer die Möglichkeit, das Gelernte unmittelbar anzuwenden und mit den jeweiligen Daten eigene Analysen durchzuführen.

Im Rahmen der FDZ-Akademien und der Daten-Workshops werden im FDZ insgesamt etwa 300 Personen pro Jahr fortgebildet. Als Dozentinnen und Dozenten der Workshops können regelmäßig ausgewiesene Expertinnen und Experten gewonnen werden. Die Evaluationen der FDZ-Veranstaltungen sind nahezu durchgängig äußerst positiv und werden bei der Auswahl von Themen sowie Dozentinnen und Dozenten für zukünftige Veranstaltungen berücksichtigt. Das FDZ bietet damit sowohl externen als auch internen Nachwuchsforscherinnen und -forschern ein anregungsreiches und anspruchsvolles Umfeld für die wissenschaftliche Weiterqualifikation im Bereich der quantitativen Methoden empirischer Bildungsforschung.

Öffentlichkeitsarbeit

Um die Möglichkeiten sichtbar zu machen, die die Datensätze am FDZ für wissenschaftliche Analysen bieten, ist eine effektive Öffentlichkeitsarbeit von zentraler Bedeutung. Sie soll den Bekanntheitsgrad des FDZ bei potenziellen Nutzerinnen und Nutzern erhöhen und zur breiteren Wahrnehmung des Datenangebots durch die *Scientific Community* beitragen. Neben Informationsbroschüren in Konferenztaschen nationaler und internationaler Tagungen im Bereich der Bildungsforschung (z. B. AEPF, ECER) verschickt das FDZ seit 2008 regelmäßig einen FDZ-Newsletter. Dieser informiert über neu verfügbare Datensätze und kommende Fortbildungsveranstaltungen. Des Weiteren stellt das FDZ seine Arbeit auf wissenschaftlichen Tagungen mit Informationsständen vor (z. B. DGPs, DGfE) und nutzt andere einschlägige Gelegenheiten, sich einem nationalen und internationalen Publikum in Vorträgen zu präsentieren (z. B. bei Veranstaltungen der GESIS, IASSIST).

Ausblick

Das FDZ am IQB baut das Datenangebot für die *Scientific Community* kontinuierlich aus; in Zukunft werden voraussichtlich auch Daten zur Verfügung gestellt werden können, die über die schulischen Bildungsetappen hinausgehen. Das FDZ strebt außerdem an, potenzielle Datengeber frühzeitig bei der Dokumentation ihrer Studien zu unterstützen. Dabei ist es sinnvoll, die Arbeiten zur Dokumentation und Vorbereitung der Datenübergabe nach Möglichkeit schon vor der Datenerhebung zu beginnen und gegebenenfalls durch den Besuch von Workshops zum Datenmanagement zu optimieren. Auch hierbei möchte das FDZ die Datenproduzenten im Bildungsbereich künftig stärker unterstützen. Darüber hinaus wird eine Intensivierung der Kooperation mit anderen Forschungsdatenzentren im In- und Ausland angestrebt.

Das FDZ am IQB hat es sich zum Ziel gesetzt, künftig in noch stärkerem Maße auch Daten bereitzustellen, anhand derer sich Fragestellungen zu aktuellen Herausforderungen im Bildungsbereich untersuchen lassen. Dazu könnten beispielsweise Daten gehören, die Analysen zu spezifischen Fördermaßnahmen etwa für Schülerinnen und Schüler mit Zuwanderungshintergrund, zum Ausbau von Ganztagschulen oder zur inklusiven Beschulung von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf ermöglichen. Vertiefende wissenschaftliche Analysen zur Umsetzung und Wirksamkeit solcher Interventionen und schulstruktureller Veränderungen sind nicht zuletzt für die Bildungspolitik und Bildungsverwaltung von Bedeutung.

Die Koordinierungsstelle zwischen dem IQB und dem Zentrum für Internationale Vergleichsstudien (ZIB)

Die Koordinierungsstelle zwischen dem Zentrum für Internationale Vergleichsstudien (ZIB) und dem IQB hat die Aufgabe, die Zusammenarbeit zwischen diesen beiden Institutionen zu unterstützen. Das ZIB stellt das internationale Pendant zum IQB dar, das für die nationalen Schulleistungstudien und die damit verbundene Forschung zuständig ist. Zentrale Aufgabe des 2010 gegründeten ZIB ist die Durchführung von internationalen Vergleichsstudien in Deutschland, insbesondere PISA. Gleichzeitig ist das ZIB in der Forschung und Nachwuchsförderung im Bereich *Educational Measurement* und *International Large-Scale-Assessment* aktiv und stärkt die internationale Sichtbarkeit der deutschen Bildungsforschung auf diesem Gebiet. Das ZIB ist ein Verbund der Technischen Universität München (TUM), des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) sowie des Leibniz-Instituts für die Pädago-

gik der Naturwissenschaften und Mathematik (IPN). Es wird gemeinsam vom Bundesministerium für Bildung und Forschung (BMBF) und den Ländern finanziert.

Mit Beginn des Jahres 2012 hat die Koordinierungsstelle ihre Arbeit am IQB aufgenommen. Aus der Verwandtschaft der Tätigkeitsfelder zwischen ZIB und IQB ergeben sich zahlreiche Abstimmungsbedarfe und Kooperationsmöglichkeiten, die für die Vergleichbarkeit der Erhebungen und die Weiterentwicklung der *Large-Scale-Assessments* relevant sind. Die Koordinierungsstelle am IQB übernimmt die Vorbereitung und Organisation von Treffen zwischen Mitarbeiterinnen und Mitarbeitern des ZIB und des IQB, die darauf abzielen, sich über aktuelle Tätigkeitsfelder, Abstimmungsbedarfe und Kooperationsmöglichkeiten auszutauschen. Daneben ist die Koordinierungsstelle an der inhaltlichen Ausgestaltung und Moderation von Tagungen und Workshops beteiligt, die von Mitarbeiterinnen und Mitarbeitern des IQB und des ZIB ausgerichtet werden.

Aktivitäten der Koordinierungsstelle

Die Koordinierungsstelle unterstützt unter anderem den Austausch und die Zusammenarbeit von Wissenschaftlerinnen und Wissenschaftlern, die auf der Basis von *Large-Scale-Assessments* über die Lernbedingungen und Lernerträge von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf forschen. Am NELSEN-Netzwerk (Network of large-scale studies on students with special educational needs) beteiligen sich seit 2012 Wissenschaftlerinnen und Wissenschaftler des ZIB, des IQB, des DIPF, des Nationalen Bildungspanels (NEPS) sowie Forscherinnen und Forscher in den Projekten TIMSS und BeLief (Bielefelder Längsschnittstudie zum Lernen in inklusiven und exklusiven Förderarrangements). Ziel dieses Netzwerkes ist es, die Daten- und Befundlage zur Situation von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf zu verbessern. Die Koordinierungsstelle sorgt unter anderem für die Abstimmung zwischen den Mitgliedern in inhaltlichen Fragen, stellt Informationen zu Studien und zur Fördersituation in den einzelnen Bundesländern zusammen, organisiert die Treffen des Netzwerkes und betreibt ein gemeinsames Internetportal, das den Informationsaustausch unterstützt und Ergebnisse dokumentiert.

Die Heterogenität von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf und die Vielfalt der Fördermaßnahmen für diese Schülergruppe stellt die Forschung vor zahlreiche Herausforderungen. Hierzu zählt zum Beispiel, dass die Diagnose von sonderpädagogischem Förderbedarf sowie die bereitgestellten Fördermaßnahmen zwischen den

Ländern erheblich variieren und die betroffenen Schülerinnen und Schüler in wissenschaftlichen Studien daher oft nicht eindeutig identifiziert und zugeordnet werden können. Eine Zielsetzung des NELSEN-Netzwerkes ist es daher, für Forschungszwecke Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf, die bestehenden Förderschwerpunkte sowie verschiedene Fördermaßnahmen in einheitlicher Weise zu kategorisieren. Hierfür hat das NELSEN-Netzwerk Entwürfe von Items entwickelt, die in *Large-Scale-Assessments* eingesetzt werden könnten.

Die Koordinierungsstelle ist darüber hinaus an Analysen zu Effekten sonderpädagogischer Förderung beteiligt. Im Zuge der bundesweiten Einführung von integrativer/inklusive Beschulung als Umsetzung der UN-Behindertenrechtskonvention hat die Frage an Bedeutung gewonnen, wie sich dieser Ansatz auf die Entwicklung schulischer Kompetenzen von Kindern und Jugendlichen mit sonderpädagogischem Förderbedarf auswirkt.

Im November 2012 war die Koordinierungsstelle an der inhaltlichen Ausgestaltung der Tagung der *Association for Educational Assessment Europe* (AEA-E) 2012 zum Thema „Assessment Across the Lifespan“ in Berlin beteiligt. Diese Tagung wurde vom IQB in Kooperation mit dem ZIB ausgerichtet. Auf der Tagung kam ein multinationales und multiprofessionelles Publikum zum Austausch über *Educational Assessment* aus inhaltlicher, methodischer und bildungspolitischer Perspektive zusammen.

Im Rahmen des Schwerpunkts „Förderung von Forschungsvorhaben in Anknüpfung an *Large-Scale-Assessments*“ finanziert das BMBF in größerem Umfang Untersuchungen, die mit Hilfe von Large-Scale-Daten inhaltlich vertiefte und methodisch anspruchsvolle Fragestellungen bearbeiten. Die Koordinierungsstelle organisiert den wissenschaftlichen Austausch der beteiligten Projekte.

Insgesamt tragen die Koordinierungsstelle und das FDZ am IQB dazu bei, die Qualität der *Large-Scale-Assessments* im Bildungsbereich und der damit verbundenen Forschung weiterzuentwickeln und deren Sichtbarkeit in der internationalen Bildungsforschung zu verbessern.



Stand Dezember 2013

07 // Publikationen

Als wissenschaftliche Einrichtung erfüllt das IQB nicht nur Serviceaufgaben, sondern führt auch zahlreiche Forschungsprojekte im Bereich der empirischen Bildungsforschung durch. Die Ergebnisse dieser Forschung werden regelmäßig in nationalen und internationalen Zeitschriften, Herausgeberwerken und Berichtsbänden dokumentiert. Nachfolgend finden sich die im Berichtszeitraum entstandenen Publikationen, an denen Mitarbeiterinnen und Mitarbeiter des IQB beteiligt waren.

Allemann-Ghionda, C., Stanat, P., Göbel, K. & Röhner, C. (2010). Migration, Identität, Sprache und Bildungserfolg. 55. *Beiheft der Zeitschrift für Pädagogik*, 7–16.

Artelt, C. & Stanat, P. (2010). Leistungen von Schülerinnen und Schülern im internationalen Vergleich – Die PISA Studie. In C. Spiel (Hrsg.), *Bildungspsychologie* (S. 352–356). Göttingen: Hogrefe.

Asbrand, B. (2007a). Grundlegende Kompetenzen religiöser Bildung: Ein Kommentar aus der Perspektive der Bildungsforschung. In D. Fischer & V. Elsenbast (Hrsg.), *Stellungnahmen und Kommentare zu „Grundlegende Kompetenzen religiöser Bildung“* (S. 40–50). Münster: Comenius-Institute.

Asbrand, B. (2007b). Orientierungen in der Weltgesellschaft – Globales Lernen als Gegenstand rekonstruktiver Sozialforschung und systemtheoretischer Analyse. In G. Büttner, A. Scheunpflug & V. Elsenbast (Hrsg.), *Zwischen Erziehung und Religion. Religionspädagogische Perspektiven nach Niklas Luhmann* (S. 74–94). Münster: LIT Verlag.

Asbrand, B. (2007c). Partnerschaft – eine Lerngelegenheit? *Zeitschrift für Internationale Bildungsforschung und Entwicklungspädagogik*, 30(3), 8–14.

Asbrand, B. (2007d). Wie erwerben Jugendliche Wissen und Handlungsorientierung in der Weltgesellschaft? In VENRO (Hrsg.), *Jahrbuch Globales Lernen 2007/2008. Standortbestimmungen – Praxisbeispiele – Perspektiven* (S. 22–33). Bonn.

Asbrand, B. & Lang-Wojtasik, G. (2007a). Globales Lernen in Forschung und Lehre. *Zeitschrift für Internationale Bildungsforschung und Entwicklungspädagogik*, 30, 2–6.

Asbrand, B. & Lang-Wojtasik, G. (2007b). Vorwärts nach weit? Anmerkungen zum Orientierungsrahmen für den Lernbereich Globale Entwicklung im Rahmen einer Bildung für nachhaltige Entwicklung. *Zeitschrift für Internationale Bildungsforschung und Entwicklungspädagogik*, 30, 33–36.

Back, M. D., Penke, L., Schmukle, S. C., Sachse, K., Borkenau, P. & Asendorpf, J. B. (2011). Why mate choices are not as reciprocal as we assume: The role of personality, flirting, and physical attractiveness. *European Journal of Personality*, 25, 120–132.

Behrens, U., Böhme, K. & Krelle, M. (2009). Zuhören – Operationalisierung und fachdidaktische Implikation. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 357–375). Weinheim: Beltz.

Benzing, M., Nikolova, R. & Hunger, S. (2010). Standardisierte Schulleistungstests. Individuelle Diagnosemöglichkeit und Unterstützung zur Unterrichtsentwicklung. Hamburg macht Schule. *Zeitschrift für Hamburger Lehrkräfte und Elternräte*, 3, 22.

Berendes, K., Dragon, N., Weinert, S., Heppt, B. & Stanat, P. (2013). Hürde Bildungssprache? Eine Annäherung an das Konzept „Bildungssprache“ unter Einbezug aktueller empirischer Forschungsergebnisse. In A. Redder & S. Weinert (Hrsg.), *Sprachförderung und Sprachdiagnostik. Interdisziplinäre Perspektiven* (S. 17–41). Münster: Waxmann.

Bergann, S. & Kroth, A. J. (2013). Geschlechts- und migrationsbezogene Disparitäten im Hochschulzugang. In P. Bornkessel, S. U. Kuhn & J. Asdonk (Hrsg.), *Von der Schule zur Hochschule: Analysen, Konzeptionen und Gestaltungsperspektiven des Übergangs* (S. 279–291). Münster: Waxmann.

Bergann, S., Stanat, P. & Ittel, A. (2010). Die Bremer Ostercamps – Analysen zur Wirksamkeit einer Maßnahme zur Förderung versetzungsgefährdeter Schülerinnen und Schüler. *Empirische Pädagogik*, 24(3), 201–223.

Blatter, K., Faust, V., Jäger, D., Schöppe, D., Artelt, C., Schneider, W. & Stanat, P. (2013). Vorschulische Förderung der phonologischen Bewusstheit und der Buchstaben-Laut-Zuordnung: Profitieren auch Kinder nichtdeutscher Herkunftssprache? In A. Redder & S. Weinert (Hrsg.), *Sprachförderung und Sprachdiagnostik. Interdisziplinäre Perspektiven* (S. 218–238). Münster: Waxmann.

Blum, W., Roppelt, A. & Müller, M. (2013). Kompetenzstufenmodelle für das Fach Mathematik. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 61–73). Münster: Waxmann.

Böhme, K. (2008). Wie gelingt es besser? SEIS: Datengestützte Selbstevaluation an Schulen. *Grundschule*, 40, 24–26.

Böhme, K. (2012). Prüfungsaufgaben im schriftlichen Deutschabitur. In H. Feilke, J. Köster & M. Steinmetz (Hrsg.), *Textkompetenzen in der Sekundarstufe II* (S. 307–330). Freiburg: Fillibach.

Böhme, K. & Bremerich-Vos, A. (2009). Diagnostik der Rechtschreibkompetenz in der Grundschule – Konstruktprüfung mittels Fehler- und Dimensionsanalysen. In D. Granzer, O. Köller, A. Bremerich Vos, M. van den Heuvel Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 330–356). Weinheim: Beltz.

Böhme, K. & Bremerich-Vos, A. (2010). Haben wir Tomaten auf den Augen? Eine Replik auf den Beitrag von Carl Ludwig Naumann und Karl-Ludwig Herné. *Didaktik Deutsch*, 1, 22–30.

Böhme, K. & Bremerich-Vos, A. (2012). Beschreibung der im Fach Deutsch untersuchten Kompetenzen. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 19–33). Münster: Waxmann.

Böhme, K., Bremerich-Vos, A. & Robitzsch, A. (2009). Aspekte der Kodierung von Schreibaufgaben: Vergleich holistischer und analytischer Kodierungen unter besonderer Berücksichtigung der Interraterreliabilität. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 290–329). Weinheim: Beltz.

Böhme, K., Felbrich, A., Weirich, S. & Stanat, P. (2013). Sprachliche Kompetenzen von Schülerinnen und Schülern mit Zuwanderungshintergrund am Ende der 4. Jahrgangsstufe – Ergebnisse des IQB-Ländervergleichs 2011. *Die Deutsche Schule*, 105, 128–143.

Böhme, K. & Haag, N. (2012). Zusammenfassung zentraler Ergebnisse des Ländervergleichs in den Fächern Deutsch und Mathematik. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 128–130). Münster: Waxmann.

Böhme, K. & Köller, O. (2010). Kompetenzstufenmodelle für den mittleren Schulabschluss – Einleitung. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 33–34). Münster: Waxmann.

Böhme, K., Leucht, M., Schipolowski, S., Porsch, R., Knigge, M. & Köller, O. (2010). Anlage und Durchführung des Ländervergleichs. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 65–85). Münster: Waxmann.

Böhme, K., Neumann, D. & Schipolowski, S. (2010). Beschreibung der im Ländervergleich im Fach Deutsch untersuchten Kompetenzen. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 19–24). Münster: Waxmann.

Böhme, K. & Richter, D. (2013). Der IQB-Ländervergleich 2011 – Inwieweit werden im Primarbereich die Bildungsstandards erreicht? *Schulmanagement*, 44(3), 30–34.

Böhme, K., Richter, D., Stanat, P., Pant, H. A. & Köller, O. (2012). Die länderübergreifenden Bildungsstandards in Deutschland. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 11–18). Münster: Waxmann.

Böhme, K. & Robitzsch, A. (2009). Methodische Aspekte der Lesekompetenz. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 250–289). Weinheim: Beltz.

Böhme, K., Robitzsch, A. & Busè, A.-K. (2010). Zur Abgrenzung des Hörverstehens gegenüber dem Leseverstehen mit Hilfe schwierigkeitsbestimmender Merkmale bei der Entwicklung von Textaufgaben. In V. Bernius & M. Imhof (Hrsg.), *Zuhörkompetenz in Unterricht und Schule* (S. 81–104). Münster: Waxmann.

Böhme, K. & Roppelt, A. (2012). Geschlechtsbezogene Disparitäten. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 173–190). Münster: Waxmann.

Böhme, K., Tiffin-Richards, S. P., Schipolowski, S. & Leucht, M. (2010). Migrationsbedingte Disparitäten bei sprachlichen Kompetenzen. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 203–224). Münster: Waxmann.

Böhme, K. & Weirich, S. (2012). Der Ländervergleich im Fach Deutsch. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 103–116). Münster: Waxmann.

Brandt, S. & Reimers, H. (2011). Lassen sich starke und schwache Schulen eindeutig unterscheiden, wenn man eine Rangordnung gemäß von Inspektionsdaten bildet? In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektion in Deutschland. Eine Zwischenbilanz aus empirischer Sicht* (S. 79–95). Münster: Waxmann.

Bremerich-Vos, A., Behrens, U., Böhme, K., Krelle, M., Neumann, D., Robitzsch, A., Schiplowski, S. & Köller, O. (2010). Kompetenzstufenmodelle für das Fach Deutsch. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich. Überprüfung der Bildungsstandards in den Fächern Deutsch und erste Fremdsprache in der neunten Jahrgangsstufe* (S. 37–49). Münster: Waxmann.

Bremerich-Vos, A. & Böhme, K. (2009a). Lesekompetenzdiagnostik – die Entwicklung eines Kompetenzstufenmodells für den Bereich Lesen. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 219–249). Weinheim: Beltz.

Bremerich-Vos, A. & Böhme, K. (2009b). Kompetenzdiagnostik im Bereich „Sprache und Sprachgebrauch untersuchen“. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 376–392). Weinheim: Beltz.

Bremerich-Vos, A., Böhme, K., Krelle, M., Weirich, S. & Köller, O. (2012). Kompetenzstufenmodelle im Fach Deutsch. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 56–71). Münster: Waxmann.

Bremerich-Vos, A., Böhme, K. & Robitzsch, A. (2009). Sprachliche Kompetenzen im Fach Deutsch – Strukturanalysen und Validierungsbefunde. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 198–218). Weinheim: Beltz.

Bremerich-Vos, A., Granzer, D., Behrens, U. & Köller, O. (2009). *Bildungsstandards für die Grundschule: Deutsch konkret: Aufgabenbeispiele – Unterrichtsanregungen – Fortbildungs-ideen*. Berlin: Cornelsen Scriptor.

Bremerich-Vos, A., Granzer, D. & Köller, O. (Hrsg.). (2008). *Lernstandsbestimmung im Fach Deutsch. Gute Aufgaben für den Unterricht*. Weinheim: Beltz.

Brunner, M., Stanat, P. & Pant, H. A. (im Druck). Diagnostik und Evaluation. In T. Seidel & A. Krapp (Hrsg.), *Pädagogische Psychologie*. Weinheim: Beltz.

Cortina, K. S. & Köller, O. (2008). Kontext: Schule. In R. Silber-eisen & M. Hasselhorn (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich C: Theorie und Forschung; Serie V: Entwicklungspsychologie, Bd. 5, Entwicklungspsychologie des Jugendalters* (S. 229–254). Göttingen: Hogrefe.

Cortina, K. S. & Pant, H. A. (2009). Metaanalyse mit Hilfe Hierarchischer Linearer Modelle. In H. Holling (Hrsg.), *Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden. Serie IV: Evaluation, Bd. 1, Evaluationsforschung: Modelle und Methoden*. Göttingen: Hogrefe.

Darsow, A., Paetsch, J. & Felbrich, A. (2012). Konzeption und Umsetzung der fachbezogenen Sprachförderung im BeFo-Projekt. In S. Jeuk & J. Schäfer (Hrsg.), *Deutsch als Zweitsprache in Kindertageseinrichtungen und Schulen. Aneignung, Förderung, Unterricht* (S. 215–234). Freiburg: Fillibach.

Darsow, A., Paetsch, J., Stanat, P. & Felbrich, A. (2012). Ansätze der Zweitsprachförderung: Eine Systematisierung. *Unterrichtswissenschaft, 40*, 64–82.

Edele, A., Dziobek, I. & Keller, M. (2013). Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy and justice sensitivity. *Learning and Individual Differences, 24*, 96–102.

Edele, A., Schotte, K., Hecht, M. & Stanat, P. (2012). *Listening comprehension tests of immigrant students' first languages (L1) Russian and Turkish in grade 9: Scaling procedure and results* (NEPS Working Paper No. 13). Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Edele, A. & Stanat, P. (2011). PISA's potential for analyses of immigrant students' educational success: The German case. In M. A. Pereyra, H. G. Kotthoff & R. Cowen (Eds.), *PISA under examination: Changing knowledge, changing tests, and changing schools* (pp. 185–206). Rotterdam: SensePublishers.

Edele, A., Stanat, P., Radmann, S. & Segeritz, M. (2013). Kulturelle Identität und Lesekompetenz von Jugendlichen aus eingewanderten Familien. 59. *Beiheft der Zeitschrift für Pädagogik*, 84–110.

Ehlert, A., Schroeders, U. & Fritz-Stratmann, U. (2012). Kritik am Diskrepanzkriterium in der Diagnostik von Legasthenie und Dyskalkulie. *Lernen und Lernstörungen, 1*(3), 169–184.

Ehmke, T., Klieme, E. & Stanat, P. (2013). Veränderungen der Lesekompetenz von PISA 2000 nach PISA 2009. Die Rolle von Unterschieden in den Bildungswegen und der Zusammensetzung der Schülerschaft. 59. *Beiheft der Zeitschrift für Pädagogik*, 132–150.

Eksner, J. & Stanat, P. (2011). Effects of ethnically segregated learning settings in the United States and Germany. In K. S. Gallagher, R. Goodyear, D. J. Brewer & R. Rueda (Eds.), *Introduction to Urban Education* (pp. 311–327). New York: Routledge.

Emmrich, R., Kuhl, P. & Vogt, A. (2011). Vergleichsarbeiten. In P. Lohauß, K. Rehkämper, U. Rockmann & W. Wendt (Hrsg.), *Bildung in Berlin und Brandenburg 2010. Ein indikatoren-gestützter Bericht zur Bildung im Lebenslauf* (S. 225–236). Berlin, Potsdam: Institut für Schulqualität der Länder Berlin und Brandenburg und Amt für Statistik Berlin-Brandenburg.

Felbrich, A., Darsow, A., Paetsch, J. & Stanat, P. (2012). Die BeFo-Interventionsstudie – Sprachsystematische und fachbezogene Sprachförderung in der Grundschule. In F. H. Hellmich, F. Hoya & S. Förster (Hrsg.), *Bedingungen des Lehrens und Lernens in der Grundschule. Bilanz und Perspektiven (Jahrbuch Grundschulforschung)* (S. 211–215). Wiesbaden: VS Verlag für Sozialwissenschaften.

Felbrich, A., Kaiser, G. & Schmotz, C. (2012). The cultural dimension of beliefs: An investigation of future primary teachers' epistemological beliefs concerning the nature of mathematics in 15 countries. *ZDM – The International Journal on Mathematics Education*, 44, 355–366.

Felbrich, A. & Stanat, P. (2011). Kinder mit Migrationshintergrund. In W. Einsiedler, M. Götz, A. Hartinger, F. Heinzl, J. Kahler & U. Sandfuchs (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (S. 191–194). Bad Heilbrunn: Klinkhardt.

Felbrich, A., Stanat, P., Paetsch, J. & Darsow, A. (2012). Das Erkenntnispotenzial experimenteller Studien zur Untersuchung der Wirksamkeit von Sprachfördermaßnahmen. In B. Ahrenholz (Hrsg.), *Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren* (S. 145–172). Berlin: De Gruyter Mouton.

Ferne, T. & Rupp, A. A. (2007). A synthesis of 15 years of research on DIF in language testing: Methodological advances, challenges, and recommendations. *Language Assessment Quarterly*, 4(2), 113–148.

Formazin, M., Schroeders, U., Köller, O., Wilhelm, O. & Westmeyer, H. (2011). Studierendenauswahl im Fach Psychologie. Testentwicklung und Validitätsbefunde. *Psychologische Rundschau*, 62(4), 221–236.

Formazin, M., Wilhelm, O., Schroeders, U., Kunina, O., Hildebrandt, A. & Köller, O. (2008). Validitäts- und Nützlichkeitsüberlegungen zur Studierendenauswahl im Allgemeinen mit Präzisierungen für das Fach Psychologie im Besonderen. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S. 204–214). Göttingen: Hogrefe.

Freund, A. M., Kourilova, S. & Kuhl, P. (2011). Stronger evidence for own-age effects in memory for older as compared to younger adults. *Memory*, 19(5), 429–448.

Gärtner, H. & Pant, H. A. (2011). Validierungsstrategien für Verfahren und Ergebnisse von Schulinspektion. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektion in Deutschland. Eine Zwischenbilanz aus empirischer Sicht* (S. 9–32). Münster: Waxmann.

Gärtner, H., Wurster, S. & Pant, H. A. (2013, nur online). The effect of school inspections on school improvement. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*.

Granzer, D. (2007a). Computer im Unterricht. In E. Jürgens (Hrsg.), *Taschenbuch Grundschule*. Hohengehren: Schneider.

Granzer, D. (2007b). Stichworte Computer, Internet, Vergleichsarbeiten. In D. H. Heckt & U. Sandfuchs (Hrsg.), *Grundschule von A bis Z*. Braunschweig: Westermann.

Granzer, D. (2009). Von den Bildungsstandards zu ihrer Überprüfung: Grundlagen der Item- und Testentwicklung. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 21–30). Weinheim: Beltz.

Granzer, D., Köller, O., Bremerich-Vos, A., van den Heuvel-Panhuizen, M., Reiss, K. & Walther, G. (Hrsg.). (2009). *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule*. Weinheim: Beltz.

Granzer, D. & Wendt, P. (2007). Selbstevaluation. *Grundschule*, 2, 6–7.

Granzer, D., Wendt, P. & Berger, R. (Hrsg.). (2008). *Selbstevaluation in der Schule*. Weinheim: Beltz.

Granzer, D., Winkelmann, H., Robitzsch, A. & Böhme, K. (2007). Bildungsstandards und Evaluation. Ein Weg zur Professionalisierung? *Perspektiven zur pädagogischen Professionalisierung*, 72, 9–15.

Groeneveld, I., Knigge, M. & Robitzsch, A. (2011). Soziale Disparitäten und Schutzfaktoren bei Schülerinnen und Schülern in der Primarstufe. *Psychologie in Erziehung und Unterricht*, 58(4), 268–279.

Grotjahn, R. & Tesch, B. (2010a). Messung der Hörverstehenskompetenz im Fach Französisch. In R. Porsch, B. Tesch & O. Köller (Hrsg.), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I* (S. 125–150). Münster: Waxmann.

Grotjahn, R. & Tesch, B. (2010b). Messung der Leseverstehenskompetenz im Fach Französisch. In R. Porsch, B. Tesch & O. Köller (Hrsg.), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I* (S. 91–124). Münster: Waxmann.

- Haag, N., Böhme, K. & Stanat, P. (2012).** Zuwanderungsbezogene Disparitäten. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 209–236). Münster: Waxmann.
- Haag, N., Heppt, B., Stanat, P., Kuhl, P. & Pant, H. A. (2013).** Second language learners' performance in mathematics: Disentangling the effects of academic language features. *Learning and Instruction*, 28, 24–34.
- Haag, N. & Roppelt, A. (2012).** Der Ländervergleich im Fach Mathematik. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 117–127). Münster: Waxmann.
- Haase, C., Heckhausen, J. & Köller, O. (2008).** Goal engagement during the school-to-work-transition: Beneficial for all, essential for girls. *Journal of Research on Adolescence*, 18(4), 671–698.
- Hachfeld, A., Anders, Y., Schroeder, S., Stanat, P. & Kunter, M. (2010).** Does immigration background matter? How teachers' predictions of students' performance relate to student background. *International Journal of Educational Research*, 49(2), 78–91.
- Hannighofer, J., van den Heuvel-Panhuizen, M., Weirich, S. & Robitzsch, A. (2011).** Revealing German primary school students' achievement in measurement. *ZDM: The International Journal for Mathematics Education*, 43(5), 651–665.
- Hannover, B. & Pöhlmann, C. (2012).** Selbstkonzept. In K. H. Horn, H. Kemnitz, W. Marotzki & U. Sandfuchs (Hrsg.), *Klinkhardt Lexikon Erziehungswissenschaft* (S. 182). Bad Heilbrunn: Klinkhardt.
- Harsch, C. & Hartig, J. (2011).** Modellbasierte Definition von fremdsprachlichen Kompetenzniveaus am Beispiel der Bildungsstandards Englisch. *Zeitschrift für Interkulturelle Fremdsprachenforschung*, 16, 6–17.
- Harsch, C. & Nöth, D. (2007).** Was können die fremdsprachlichen Bildungsstandards der KMK leisten? *Praxis fsu*, 06/07, 2–6.
- Harsch, C., Pant, H. A. & Köller, O. (Eds.). (2010).** *Calibrating standards-based assessment tasks for English as a first foreign language. Standard-setting procedures in Germany* (Vol. 2). Münster: Waxmann.
- Harsch, C. & Tiffin-Richards, S. P. (2010).** Setting standards in line with the Common European Framework of Reference. In C. Harsch, H. A. Pant & O. Köller (Eds.), *Calibrating Standards-based Assessment Tasks for English as a First Foreign Language. Standard-setting Procedures in Germany* (Vol. 2, pp. 9–38). Münster: Waxmann.
- Hecht, M., Roppelt, A. & Siegle, T. (2013).** Testdesign und Auswertung des Ländervergleichs. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 391–402). Münster: Waxmann.
- Heitmann, P. (2012).** *Bewertungskompetenz im Rahmen naturwissenschaftlicher Problemlöseprozesse. Modellierung und Diagnose der Kompetenzen Bewertung und analytisches Problemlösen für das Fach Chemie*. Berlin: Logos.
- Heitmann, P., Schmidt, F. & Tiemann, R. (2011).** Kriterien im Bewertungsprozess – Eine qualitative Analyse. In D. Höttecke (Hrsg.), *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie* (S. 297–1299). Münster: LIT Verlag.
- Heitmann, P. & Tiemann, R. (2011a).** Aspekte von Bewertungskompetenz im naturwissenschaftlichen Unterricht: Die Qual der Wahl zwischen Käse und Analogkäse. *CHEMKON*, 18(3), 129–133.
- Heitmann, P. & Tiemann, R. (2011b).** Bewertungskompetenz von Zehntklässlern im Fach Chemie. Betrachtung von Kriterien und Argumenten am Beispiel Pestizideinsatz. *Der mathematische und naturwissenschaftliche Unterricht (MNU)*, 64(4), 238–243.
- Heitmann, P. & Tiemann, R. (2012).** Diagnose von Bewertungskompetenz im Rahmen von Problemlöseprozessen. In S. Bernholt (Hrsg.), *Konzepte fachdidaktischer Strukturierung für den Unterricht* (S. 328–330). Münster: LIT Verlag.
- Heitmann, P. & Tiemann, R. (2013).** Erdbeeren im Winter? Bewertungskompetenz empirisch erfassen und beschreiben. In D. Höttecke, J. Menthe, I. Eilks & C. Höhle (Hrsg.), *Handeln in Zeiten des Klimawandels – Bewerten Lernen als Bildungsaufgabe* (S. 183–202). Münster: Waxmann.
- Helbig, M., Baier, T. & Kroth, A. J. (2012).** Die Auswirkung von Studiengebühren auf die Studierneigung in Deutschland. Evidenz aus einem natürlichen Experiment auf Basis der HIS–Studienberechtigtenbefragung. *Zeitschrift für Soziologie*, 41(3), 227–246.

- Henschel, S., Roick, T., Brunner, M. & Stanat, P. (2013).** Leseselbstkonzept und Textart: Lassen sich literarisches und faktuales Leseselbstkonzept trennen? *Zeitschrift für Pädagogische Psychologie*, 27(3), 181–191.
- Heppt, B., Dragon, N., Berendes, K., Stanat, P. & Weinert, S. (2012).** Beherrschung von Bildungssprache bei Kindern im Grundschulalter. *Diskurs Kindheits- und Jugendforschung*, 3, 349–356.
- Herzmann, G., Danthiir, V., Schacht, A., Wilhelm, O. & Sommer, W. (2008).** Towards a Comprehensive Test Battery for Face Cognition: Assessment of the Tasks. *Behavior Research Methods*, 40(3), 840–857.
- Herzmann, G., Danthiir, V., Wilhelm, O., Sommer, W. & Schacht, A. (2007).** Face memory: a cognitive and psychophysiological approach to the assessment of antecedents of emotional intelligence. In G. Matthews, M. Zeidner & R. D. Roberts (Eds.), *Emotional intelligence: Knowns and unknowns* (pp. 305–338). Oxford: Oxford University Press.
- Herzmann, G., Kunina, O., Sommer, W. & Wilhelm, O. (2009).** Individual differences in face cognition: Brain-behavior relationships. *Journal of Cognitive Neuroscience*, 22(3), 571–589.
- Herzmann, G., Wilhelm, O., Kunina, O. & Sommer, W. (2008).** Towards a Comprehensive Test Battery for Face Processing: Assessment of the Tasks. *Behavior Research Methods, Instruments, & Computers*, 40, 840–857.
- Hildebrandt, A., Herzmann, G., Sommer, W. & Wilhelm, O. (2010).** Structural invariance and age-related performance differences in face cognition. *Psychology and Aging*, 25(4), 794–810.
- Hildebrandt, A., Wilhelm, O. & Robitzsch, A. (2009).** Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology*, 16(2), 87–102.
- Hoffmann, L. & Kuhl, P. (2013).** Rezension zu: F. Schott & S. A. Ghanbari (Hrsg.): *Bildungsstandards, Kompetenzdiagnostik und kompetenzorientierter Unterricht zur Qualitätssicherung des Bildungswesens. Eine problemorientierte Einführung in die theoretischen Grundlagen*. *Zeitschrift für Pädagogik*, 59(1), 156–160.
- Hoffmann, L. & Sturzbecher, D. (2012).** Soziale Schulqualität, Schulverweigerung und Schülerverkehr. In D. Sturzbecher, A. Kleeberg-Niepage & L. Hoffmann (Hrsg.), *Aufschwung Ost? Lebenssituation und Wertorientierung ostdeutscher Jugendlicher*. (S. 189–214). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Holling, H., Preckel, F., Vock, M., Roßbach, H.-G., Baudson, T. & Kuger, S. (2009).** *Begabte Kinder finden und fördern. Ein Ratgeber für Eltern, Erzieherinnen und Erzieher, Lehrerinnen und Lehrer*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Hülür, G., Wilhelm, O. & Robitzsch, A. (2011a).** Multivariate Veränderungsmodelle für Schulnoten und Schülerleistungen in Deutsch und Mathematik. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43(4), 173–185.
- Hülür, G., Wilhelm, O. & Robitzsch, A. (2011b).** Intelligence dedifferentiation in early childhood. *Journal of Individual Differences*, 32, 170–179.
- Hülür, G., Wilhelm, O. & Schipolowski, S. (2011).** Prediction of self-reported knowledge with over-claiming, fluid and crystallized intelligence and typical intellectual engagement. *Learning and Individual Differences*, 21(6), 742–746.
- Jansen, M., Schroeders, U. & Stanat, P. (2013).** Motivationale Schülermerkmale in Mathematik und den Naturwissenschaften. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 347–366). Münster: Waxmann.
- Jude, N., Hartig, J., Schipolowski, S., Böhme, K. & Stanat, P. (2013).** Definition und Messung von Lesekompetenz. PISA und die Bildungsstandards. 59. Beiheft der *Zeitschrift für Pädagogik*, 200–228.
- Kampa, N., Kunter, M., Maaz, K. & Baumert, J. (2011).** Die soziale Herkunft von Lehrkräften in Deutschland: Eine empirische Studie zum Zusammenhang mit Berufsausübung und berufsbezogenen Überzeugungen bei Sekundarstufenlehrkräften. *Zeitschrift für Pädagogik*, 57(1), 70–92.
- Katzenbach, M., Blum, W., Drüke-Noe, C., Keller, K., Köller, O., Leiß, D., Müller, M. & Roppelt, A. (2009).** *Bildungsstandards: Kompetenzen überprüfen. Mathematik Sekundarstufe I*. Berlin: Cornelsen Scriptor.
- Kauertz, A. & Fischer, H. E. (2013).** Die Bildungsstandards im Fach Physik für den Mittleren Schulabschluss. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 48–49). Münster: Waxmann.

- Kauertz, A., Fischer, H. E. & Jansen, M. (2013).** Kompetenzstufenmodelle für das Fach Physik. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 92–99). Münster: Waxmann.
- Keye, D., Wilhelm, O. & Oberauer, K. (2009).** Structure and correlates of the German version of the Brief UPPS impulsive behavior scales. *European Journal of Psychological Assessment*, 25(3), 175–185.
- Keye, D., Wilhelm, O., Oberauer, K. & Ravenzwaaij, D. (2009).** Individual differences in conflict-monitoring: Testing means and covariance hypothesis about the Simon and the Eriksen Flanker Task. *Psychological Research*, 73(6), 762–776.
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S. & Baumert, J. (2013).** Teacher's content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education*, 64, 90–106.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W. & Stanat, P. (Hrsg.). (2010).** *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Knigge, M., Groeneveld, I., Stanat, P. & Pant, H. A. (2011).** PISA, PIRLS, and more – the Research Data Centre at the Institute for Educational Progress (FDZ-IQB). *Schmollers Jahrbuch*, 131(3), 521–528.
- Knigge, M. & Köller, O. (2010).** Effekte der sozialen Zusammensetzung der Schülerschaft. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 227–244). Münster: Waxmann.
- Knigge, M. & Leucht, M. (2010).** Soziale Disparitäten im Spracherwerb. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 185–202). Münster: Waxmann.
- Köller, O. (2007a).** Bildungsmonitoring und Vergleichsarbeiten in Deutschland. In P. Labudde (Hrsg.), *Bildungsstandards am Gymnasium* (S. 65–73). Bern: h.e.p.
- Köller, O. (2007b).** Bildungsstandards, einheitliche Prüfungsanforderungen und Qualitätssicherung in der Sekundarstufe II. In D. Benner (Hrsg.), *Bildungsstandards. Chancen und Grenzen, Beispiele und Perspektiven* (S. 13–28). Paderborn: Schöningh.
- Köller, O. (2007c).** Die deutsche Schule im Lichte internationaler Schulleistungsuntersuchungen (TIMSS, PISA, IGLU, DESI). In H. J. Apel & W. Sacher (Hrsg.), *Studienbuch Schulpädagogik* (Bd. 3, S. 138–154). Bad Heilbrunn: Klinkhardt.
- Köller, O. (2007d).** Evaluation. In H. E. Tenorth & R. Tippelt (Hrsg.), *Fachlexikon Pädagogik* (S. 214–217). Weinheim: Beltz.
- Köller, O. (2007e).** Standards und Qualitätssicherung zur Output-steuerung im System und in der Einzelinstitution. In J. v. Buer & C. Wagner (Hrsg.), *Qualität von Schule: Ein kritisches Handbuch* (S. 93–102). Frankfurt a. M.: Peter Lang.
- Köller, O. (2008a).** Bildungsstandards – Verfahren und Kriterien bei der Entwicklung von Messinstrumenten. *Zeitschrift für Pädagogik*, 54(2), 163–173.
- Köller, O. (2008b).** Bildungsstandards in Deutschland: Implikationen für die Qualitätssicherung und Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 10, 47–59.
- Köller, O. (2008c).** Bildungsstandards in einem Gesamtsystem der Qualitätssicherung im allgemeinbildenden Schulsystem Deutschlands. 53. Beiheft der *Zeitschrift für Pädagogik*, 59–75.
- Köller, O. (2008d).** Forschungsansätze in der Pädagogischen Psychologie. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch Pädagogische Psychologie* (S. 697–711). Göttingen: Hogrefe.
- Köller, O. (2008e).** Lehr-Lern-Forschung. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch Pädagogische Psychologie* (S. 210–222). Göttingen: Hogrefe.
- Köller, O. (2008f).** Operationalisierung und Überprüfung von Bildungsstandards: Ist Bildung messbar? In F. Hofmann, C. Schreiner & J. Thonhauser (Hrsg.), *Qualitative und quantitative Aspekte: Zu ihrer Komplementarität in der erziehungswissenschaftlichen Forschung* (S. 281–298). Münster: Waxmann.
- Köller, O. (2008g).** Qualitätssicherung in der schulischen Bildung. In G. Weißeno (Hrsg.), *Politikkompetenz: Was Unterricht zu leisten hat* (S. 22–32). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Köller, O. (2009a).** Bildungsstandards. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 529–548). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Köller, O. (2009b).** Evaluation pädagogisch-psychologischer Maßnahmen. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 333–351). Berlin: Springer.

- Köller, O. (2009c).** Philosophie von Bildungsstandards. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 12–20). Weinheim: Beltz.
- Köller, O. (2009d).** Quantitative Schulforschung. In S. Blömeke, T. Bohl, L. Haag, G. L. Wojtasik & W. Sacher (Hrsg.), *Handbuch Schule* (S. 126–134). Bad Heilbrunn: Klinkhardt.
- Köller, O. (2010).** Politische und inhaltliche Rahmenbedingungen bei der Setzung von Kompetenzstufen. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 35–36). Münster: Waxmann.
- Köller, O. & Baumert, J. (2008).** Entwicklung von Schulleistungen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (6. Aufl., S. 735–768). Weinheim: Beltz.
- Köller, O., Knigge, M., Schipolowski, S., Leucht, M., Tiffin-Richards, S. P., Porsch, R., Tesch, B., Vock, M., Böhme, K., Frenzel, J. & Andrade, J. (2010).** Der Blick in die Länder. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 107–176). Münster: Waxmann.
- Köller, O., Knigge, M. & Tesch, B. (Hrsg.). (2010).** *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.
- Köller, O. & Köller, M. (2007).** Bedeutung von Schulleiterinnen und Schulleitern für die Qualität von Unterricht. In T. R. Baulecke (Hrsg.), *SchulleitungPlus: Grundsätze und Verfahren wirksamer Führung* (S. 76–83). München: Oldenbourg.
- Köller, O., Leucht, M. & Pant, H. A. (2012).** Effekte bilingualer Unterrichtsangebote auf die Englischleistungen in der Sekundarstufe I. *Unterrichtswissenschaft*, 40(4), 334–350.
- Köller, O., Zeinz, H. & Trautwein, U. (2008).** Class – average achievement, marks, and academic self-concept in German primary schools. In H. Marsh, R. H. Craven & D. M. McInerney (Eds.), *Self-processes, Learning, and Enabling Human Potential: Dynamic New Approaches. Advances in SelfResearch* (Vol. 3, pp. 331–352). Charlotte, NC: Information Age Publishing Inc.
- König, J., Kaiser, G. & Felbrich, A. (2012).** Zum Zusammenhang von Wissen und Überzeugungen am Ende der Lehrerbildung: Spiegelt sich pädagogisches Wissen in den Kompetenzselbsteinschätzungen angehender Lehrkräfte? *Zeitschrift für Pädagogik*, 58(4), 476–491.
- Kristen, C., Edele, A., Kalter, F., Kogan, I., Schulz, B., Stanat, P. & Will, G. (2011).** The education of migrants and their children across the life course. *Zeitschrift für Erziehungswissenschaft, Sonderheft*, 14, 121–137.
- Kröner, S., Lüdtke, O., Maaz, K., Trautwein, U. & Köller, O. (2008).** Wer geht ins Theater? Künstlerisches Interesse und Offenheit für Erfahrung als Prädiktoren für Veränderungen kultureller Partizipation in der Emerging Adulthood. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 40(2), 100–110.
- Kroth, A. J. (2013).** The effects of college costs and financial aid in Germany. Why are students sensitive to college costs in a low-cost/high-aid system? In H. D. Meyer, E. P. St. John, M. Chankseiani & L. Uribe (Eds.), *Globalization and social justice: Comparative studies of access to higher education* (pp. 153–171). New York: AMS Press.
- Kroth, A. J. & Daniel, H.-D. (2008).** Internationale Hochschulrankings. Ein kritischer Methodenvergleich. *Zeitschrift für Erziehungswissenschaft*, 11(4), 542–558.
- Kühnen, U., Hannover, B., Pöhlmann, C. & Roeder, U.-R. (2013).** How self-construal affects dispositionalism in attributions. *Situated Social Cognition*, 31, 238–260.
- Kuhl, P., Felbrich, A., Richter, D., Stanat, P. & Pant, H. A. (2013).** Die Jahrgangsmischung auf dem Prüfstand: Effekte jahrgangsübergreifenden Lernens auf Kompetenzen und sozioemotionalen Wohlbefinden von Grundschülerinnen und Grundschulern. In A. Schulze & R. Becker (Hrsg.), *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen* (S. 299–324). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kuhl, P. & Hannover, B. (2012).** Differentielle Benotungen von Mädchen und Jungen? Der Einfluss der von der Lehrkraft eingeschätzten Kompetenz zum selbstgesteuerten Lernen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44(3), 153–162.
- Kuhl, P., Lenkeit, J., Wendt, W. & Pant, H. A. (2011).** Die Kontextuierung von Leistungswerten bei Vergleichs- und Prüfungsarbeiten: Verschiedene Wege, die Zusammensetzung der Schülerschaft in den Rückmeldungen an Schulen und die Schulinspektion zu berücksichtigen. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektionen in Deutschland – eine erste empirische Zwischenbilanz* (S. 237–259). Münster: Waxmann.

- Kuhl, P., Siegle, T. & Lenski, A. E. (2013).** Soziale Disparitäten. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 275–296). Münster: Waxmann.
- Kunina, O., Wilhelm, O., Formazin, M., Jonkmann, K. & Schroeders, U. (2007).** Extended criteria and predictors in college admission: Exploring the structure of study success and investigating the validity of domain knowledge. *Psychological Science*, 49(2), 88–114.
- Kunina-Habenicht, O., Rupp, A. A. & Wilhelm, O. (2009).** A practical illustration of multidimensional diagnostic skills profiling: Comparing results from confirmatory factor analysis and diagnostic classification models. *Studies in Educational Evaluation*, 35, 64–70.
- Kunina-Habenicht, O., Matthes, F., Wilhelm, O. & Rupp, A. A. (2010).** Kognitive Diagnosemodelle: Theoretisches Potential und methodische Probleme. *Zeitschrift für Pädagogik*, 56, 75–85.
- Kunter, M., Baumert, J. & Köller, O. (2007).** Effective classroom management and the development of subject-related interest. *Learning and Instruction*, 17(5), 494–509.
- Kunter, M., Kleickmann, T., Klusmann, U. & Richter, D. (2011).** Die Entwicklung professioneller Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse aus dem Forschungsprogramm COACTIV* (S. 55–68). Münster: Waxmann.
- Kunter, M., Kleickmann, T., Klusmann, U. & Richter, D. (2013).** The development of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 63–77). New York: Springer.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. & Hachfeld, A. (2013).** Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820.
- Lehmann, R. H., Hunger, S., Ivanov, S. & Gänsfuß, R. (2012).** LAU 21. Aspekt der Lernausgangslage und der Lernentwicklung – Klassenstufe 11. Ergebnisse einer längsschnittlichen Untersuchung in Hamburg. In B. f. S. u. Berufsbildung (Hrsg.), *LAU – Aspekte der Lernausgangslage und der Lernentwicklung. Klassenstufen 11 und 13* (S. 9–150). Münster: Waxmann.
- Lenski, A. E. (2011).** Rezension zu: Spiel, C., Schober, B., Wagner, P., Reimann, R. (Hrsg.): *Bildungspsychologie. Zeitschrift für Pädagogik*, 3, 441–443.
- Lesaux, N. K., Rupp, A. A. & Siegel, L. S. (2007).** Growth in reading skills of children from diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology*, 99, 821–834.
- Leucht, M., Frenzel, J. & Pöhlmann, C. (2010).** Der Ländervergleich im Fach Englisch. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 97–104). Münster: Waxmann.
- Leucht, M., Harsch, C. & Köller, O. (2012).** Steuerung zukünftiger Aufgabenentwicklung durch Vorhersage der Schwierigkeiten eines Tests für die erste Fremdsprache Englisch durch Dutch Grid Merkmale. *Diagnostica*, 58, 31–44.
- Leucht, M., Prieß-Buchheit, J., Pant, H. A. & Köller, O. (2013).** Sometimes less is more: Educational effectiveness of English as a foreign language instruction in German classrooms. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 24(4), 435–451.
- Leucht, M., Retelsdorf, J., Möller, J. & Köller, O. (2010).** Zur Dimensionalität rezeptiver Kompetenzen im Fach Englisch. *Zeitschrift für Pädagogische Psychologie*, 24, 123–138.
- Leucht, M. & Schipolowski, S. (2010).** Die sprachlichen Kompetenzen in den Ländern: Zusammenfassung. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 105). Münster: Waxmann.
- Leucht, M., Tiffin-Richards, S. P., Vock, M., Pant, H. A. & Köller, O. (2012).** Diagnostische Kompetenz von Englischlehrkräften bei der Bewertung von Schülerleistungen mit Hilfe des Gemeinsamen Europäischen Referenzrahmens für Sprachen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44, 163–177.
- Limbird, C. K., Maluch, J. T., Rjosk, C., Stanat, P. & Merkens, H. (2013).** Differential growth patterns in emerging reading skills of Turkish-German bilingual and German monolingual primary school students. *Reading and Writing*, published online 22 September 2013.
- Lintorf, K., McElvany, N., Rjosk, C., Schroeder, S., Baumert, J., Schnotz, W., Horz, H. & Ullrich, M. (2011).** Zuverlässigkeit von Lehrerurteilen – Reliabilität verschiedener Urteilsmaße bei Einschätzung von Aufgabenschwierigkeiten. *Unterrichtswissenschaft*, 39(2), 102–120.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T. & Muthén, B. O. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13.

Lüdtke, O. & Robitzsch, A. (2009). Assessing within-group agreement: A critical examination of a random-group resampling approach. *Organizational Research Methods*, 12(3), 461–487.

Lüdtke, O., Robitzsch, A., Köller, O. & Winkelmann, H. (2010). Kausale Effekte in der Empirischen Bildungsforschung. Ein Vergleich verschiedener Ansätze zur Schätzung des Effekts des Einschulungsalters. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (S. 257–284). Münster: Waxmann.

Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau*, 58(2), 103–117.

Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2008). Steht Transparenz einer adäquaten Datenauswertung im Wege? Eine Antwort auf Wuttke (2008). *Psychologische Rundschau*, 58(3), 180–181.

Lüdtke, O., Robitzsch, A., Trautwein, U., Kreuter, F. & Ihme, J. M. (2007). Are there test administrator effects in large-scale educational assessments? Using cross-classified multilevel analysis to probe for effects on math achievement and sample attrition. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 3(4), 149–159.

Lüdtke, O., Robitzsch, A., Trautwein, U. & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings in multilevel modelling. *Contemporary Educational Psychology*, 34(2), 120–131.

Maaz, K., Watermann, R. & Köller, O. (2009). Die Gewährung von Bildungschancen durch institutionelle Öffnung. *Pädagogische Rundschau*, 63, 159–177.

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. O. & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, 44(6), 764–802.

Marsh, H. W., Muthén, B. O., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S. & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Applications to students' evaluations of university teaching. *Structural Equation Modeling*, 16(3), 439–476.

Marsh, H. W., Trautwein, U., Lüdtke, O., Baumert, J. & Köller, O. (2007). The big fish little pond effect: Persistent negative effects of selective high schools on selfconcept after graduation. *American Educational Research Journal*, 44(3), 631–669.

Marsh, H. W., Trautwein, U., Lüdtke, O. & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, 100(3), 510–524.

Martens, M. & Asbrand, B. (2013). Kompetenzorientierter Unterricht. *Schulmagazin* 5–10(5), 7–10.

Marx, A. & Roick, T. (2012). Prädiktoren des Hörverstehens bei Jugendlichen deutscher und Jugendlichen nichtdeutscher Herkunftssprache. *Zeitschrift für Pädagogische Psychologie*, 26(2), 121–134.

Marx, A., Stanat, P., Roick, T., Segerer, R., Schneider, W. & Marx, P. (in Vorbereitung). *Determinants of reading in first-language and second-language students.*

Marx, A. & Sauer, M. (2011). Lerneffekte von Gedenkstättenbesuchen im Kontext des Geschichtsunterrichts. Eine quantitative Studie am Beispiel der KZ-Gedenkstätten Buchenwald und Moringen. In B. Pampel (Hrsg.), *Erschrecken – Mitgefühl – Distanz. Empirische Befunde über Schülerinnen und Schüler in Gedenkstätten und zeitgenössischen Ausstellungen* (S. 115–146). Leipzig: Leipziger Universitäts-Verlag.

Marx, A. & Stanat, P. (2012). Reading comprehension of immigrant students in Germany: Research evidence on determinants and target points for intervention. *Reading and Writing: An Interdisciplinary Journal*, 25(8), 1929–1945.

Mayer, J. & Wellnitz, N. (2013). Die Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 42–45). Münster: Waxmann.

- Mayer, J., Wellnitz, N., Klebba, N. & Kampa, N. (2013).** Kompetenzstufenmodelle für das Fach Biologie. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 74–82). Münster: Waxmann.
- McElvany, N. & Rjosk, C. (2013).** Wann kann Kompetenzdiagnostik negative Auswirkungen haben? *Zeitschrift für Erziehungswissenschaft, 16*, 65–70.
- Mechtel, M. (2008).** IT-Unterstützung der Testaufgabenentwicklung am IQB. In A. Breiter, A. Lange & E. Stauke (Eds.), *School Information Systems and Data-based Decision-making* (pp. 61–70). Frankfurt a. M.: Peter Lang.
- Meißner, F.-J. & Tesch, B. (2008).** Bildungsstandards für Französisch als erste Fremdsprache. In B. Tesch, E. Leupold & O. Köller (Hrsg.), *Bildungsstandards Französisch konkret Sekundarstufe I: Aufgabenbeispiele, Unterrichtsideen und Fortbildungsmöglichkeiten* (S. 44–50). Berlin: Cornelsen Scriptor.
- Meißner, F.-J. & Tesch, B. (2010).** Kompetenzorientierter Spanischunterricht. In F. J. Meißner & B. Tesch (Hrsg.), *Spanisch kompetenzorientiert unterrichten. Didaktische Grundlagen für die Aufgabenkonstruktion* (S. 14–27). Stuttgart: Klett.
- Meißner, F.-J. & Tesch, B. (Hrsg.). (2010).** *Spanisch kompetenzorientiert unterrichten. Didaktische Grundlagen für die Aufgabenkonstruktion*. Stuttgart: Klett.
- Meißner, F.-J., Tesch, B. & Vazquez, G. (2011).** Interkomprehension und Kompetenzförderung mit Blick auf die Konstruktion von Lehrwerken (nicht nur) für den Spanischunterricht. In F. J. Meißner & U. Krämer (Hrsg.), *Spanischunterricht gestalten. Wege zu Mehrsprachigkeit und Mehrkulturalität* (S. 81–122). Seelze: Friedrich Verlag.
- Metzeld, D., Isaac, K., Groß Ophoff, J., Speck-Hamdan, A. & Böhme, K. (2009).** Warum sich der Aufwand lohnt – Der Beitrag von Vergleichsarbeiten zum kompetenzorientierten Unterrichten. *Grundschulunterricht, 56*(2), 4–8.
- Meyer, H.-D., St. John, E. P., Jalava, M., Kroth, A. J. & Somers, P. (2013).** Fairness in access to higher education: Towards a global public debate? In H. D. Meyer, E. P. St. John, M. Chankeliani & L. Uribe (Eds.), *Globalization and social justice: Comparative studies of access to higher education* (pp. 277–289). New York: AMS Press.
- Möller, J., Pohlmann, B., Köller, O. & Marsh, H. W. (2009).** A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*(3), 1129–1176.
- Möller, J., Zaunbauer, A. & Leucht, M. (2010).** Fremdsprachenlernen. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 219–225). Weinheim: Beltz.
- Molz, G., Schulze, R., Schroeders, U. & Wilhelm, O. (2010).** TBS-TK Rezension: „Wechsler Intelligenztest für Erwachsene WIE. Deutschsprachige Bearbeitung und Adaptation des WAIS-III von David Wechsler“. *Psychologische Rundschau, 61*(4), 229–230.
- Müller-Hartmann, A., Schocker-v. Ditfurth, M. & Pant, H. A. (Hrsg.). (2013).** *Englischkompetenzen entwickeln: Praxiserprobte Lernaufgaben für die Sekundarstufe I*. Braunschweig: Diesterweg.
- Musiat, P., Hoffmann, L. & Schmidt, U. (2012).** Personalised computerised feedback in e-mental health. *Journal of Mental Health, 21*(4), 346–354.
- Naumann, J., Artelt, C., Schneider, W. & Stanat, P. (2010).** Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23–65). Münster: Waxmann.
- Neugebauer, B., Grotjahn, R. & Tesch, B. (2012).** Zeitbegrenzung in Lesetests. Auswirkungen auf das Testkonstrukt, testmethodische Konsequenzen und didaktisches Potential am Beispiel der VERA-8-Leseaufgaben im Fach Französisch. *Zeitschrift für Fremdsprachenforschung, 23*(2), 195–241.
- Neugebauer, B., Kampa, N. & Köller, O. (2011).** How significant are test conditions? *Cadmo, 19*, 99–112.
- Oberauer, K., Süß, H.-M., Wilhelm, O. & Sander, R. (2007).** Individual differences in working memory capacity and reasoning ability. In A. Conway, C. Jarrold, M. Kane, A. Miyake & J. Towse (Eds.), *Variation in working memory* (pp. 49–75). Oxford: Oxford University Press.
- Oberauer, K., Süß, H.-M., Wittmann, W. W. & Wilhelm, O. (2008).** Which working memory functions predict intelligence? *Intelligence, 36*(6), 641–652.

Paetsch, J., Wolf, K. M. & Stanat, P. (2010). *Förderung von Kindern und Jugendlichen aus Zuwandererfamilien*. Unveröffentlichte Expertise für den Expertenrat „Herkunft und Bildungserfolg“ im Auftrag des Ministeriums für Kultus, Jugend und Sport des Landes Baden-Württemberg. Berlin: AB Empirische Bildungsforschung, Freie Universität.

Paetsch, J., Wolf, K. M., Stanat, P. & Darsow, A. (im Druck). Sprachförderung von Kindern und Jugendlichen aus Zuwandererfamilien. *Zeitschrift für Erziehungswissenschaft*.

Pant, H. A. (im Druck). Aufbereitung von Evidenz für bildungspolitische und pädagogische Entscheidungen: Metaanalysen in der Bildungsforschung. In BMBF (Hrsg.), *Bildungsforschung 2020 – Herausforderungen und Perspektiven. Dokumentation der Tagung des Bundesministeriums für Bildung und Forschung vom 29.–30. März 2012*. Berlin: BMBF.

Pant, H. A. (2013). Wer hat einen Nutzen von Kompetenzmodellen? *Zeitschrift für Erziehungswissenschaft*, 16 (Sonderheft 1: Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen), 71–79.

Pant, H. A., Böhme, K. & Köller, O. (2012). Das Kompetenzkonzept der Bildungsstandards und die Entwicklung von Kompetenzstufenmodellen. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 49–55). Münster: Waxmann.

Pant, H. A., Böhme, K. & Köller, O. (2013). Das Kompetenzkonzept der Bildungsstandards und die Entwicklung von Kompetenzstufenmodellen. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 53–60). Münster: Waxmann.

Pant, H. A., Emmrich, R., Kuhl, P. & Harych, P. (2011). Leistungsüberprüfung durch Schulleistungsstudien und Vergleichsarbeiten. In H. U. Grunder, K. Kansteiner Schänzlin & H. Moser (Hrsg.), *Professionswissen für Lehrerinnen und Lehrer*. (Bd. 4, S. 123–141). Baltmannsweiler: Schneider.

Pant, H. A., Rupp, A. A., Tiffin-Richards, S. P. & Köller, O. (2009). Validity issues in standard-setting studies. *Studies in Educational Evaluation*, 35(2–3), 95–101.

Pant, H. A. & Stanat, P. (2013). Qualitätssicherung durch kompetenzorientierte Bildungsstandards im allgemeinbildenden Schulsystem. *BWP*, 42(2), 6–10.

Pant, H. A., Stanat, P., Pöhlmann, C. & Böhme, K. (2013). Die Bildungsstandards im allgemeinbildenden Schulsystem. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 13–22). Münster: Waxmann.

Pant, H. A., Stanat, P., Pöhlmann, C., Hecht, M., Jansen, M., Kampa, N., Lenski, A. E., Penk, C., Radmann, S., Roppelt, A., Schroeders, U., Siegle, T. & Ziemke, A. (2013). Der Blick in die Länder. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 159–248). Münster: Waxmann.

Pant, H. A., Stanat, P., Pöhlmann, C., Roppelt, A., Schroeders, U. & Siegle, T. (2013). Der IQB-Ländervergleich 2012: Zusammenfassung und Einordnung der Befunde. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 403–414). Münster: Waxmann.

Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.

Pant, H. A., Tiffin-Richards, S. P. & Köller, O. (2010). Standard-Setting für Kompetenztests im Large-Scale-Assessment. 56. Beiheft der *Zeitschrift für Pädagogik*, 175–188.

Pant, H. A., Vock, M., Pöhlmann, C. & Köller, O. (2008a). Eine modellbasierte Erfassung der Auseinandersetzung von Lehrkräften mit den länderübergreifenden Bildungsstandards. In E.-M. Lankes (Hrsg.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 245–260). Münster: Waxmann.

Pant, H. A., Vock, M., Pöhlmann, C. & Köller, O. (2008b). Offenheit für Innovationen: Befunde aus einer Studie zur Rezeption der Bildungsstandards bei Lehrkräften und Zusammenhänge mit Schülerleistungen. *Zeitschrift für Pädagogik*, 54(6), 827–845.

Pietsch, M., Böhme, K., Robitzsch, A. & Stubbe, T. (2009). Das Stufenmodell zur Lesekompetenz der länderübergreifenden Bildungsstandards im Vergleich zu IGLU 2006. In D. Granzer, O. Köller, A. Bremerich Vos, M. van den Heuvel Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 393–416). Weinheim: Beltz.

Pöhlmann, C. (2008). *Das independente und interdependente Selbst in Entscheidungssituationen. Befunde zu Motivation und Konsistenz.* Saarbrücken: VDM.

Pöhlmann, C., Carranza, E., Hannover, B. & Iyengar, S. S. (2007). Repercussions of self-construal for self-relevant and other-relevant choice. *Social Cognition*, 25(2), 284–305.

Pöhlmann, C., Haag, N. & Stanat, P. (2013). Zuwanderungsbezogene Disparitäten. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 297–330). Münster: Waxmann.

Pöhlmann, C., Neumann, D., Tesch, B. & Köller, O. (2010). Die Bildungsstandards im allgemeinbildenden Schulsystem. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 13–18). Münster: Waxmann.

Pöhlmann, C., Pant, H. A., Frenzel, J., Roppelt, A. & Köller, O. (im Druck). Auswirkungen einer Intervention auf die Auseinandersetzung und Arbeit mit Bildungsstandards bei Mathematik-Lehrkräften. *Zeitschrift für Erziehungswissenschaft*.

Pöhlmann, C., Pant, H. A. & Köller, O. (2010). Bildungsstandards. In T. Hascher & B. Schmitz (Hrsg.), *Handbuch Pädagogische Interventionsforschung* (S. 225–234). Weinheim: Juventa.

Popescu, I. & Neugebauer, B. (2012). Testung und Evaluierung des Leseverstehens unter besonderer Berücksichtigung methodischer Kompetenzen. *Fremdsprachen Lehren und Lernen*, 41(1), 87–101.

Porsch, R. (2010a). Die Erprobung eines Kodierschemas zur Messung der Schreibkompetenz im Fach Französisch. In R. Porsch, B. Tesch & O. Köller (Hrsg.), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I* (S. 267–286). Münster: Waxmann.

Porsch, R. (2010b). Schreibkompetenzvermittlung im Englischunterricht – Eine bundesweite Umfrage unter Lehrkräften und Schülern der Sekundarstufe I. *Zeitschrift für Fremdsprachenforschung*, 21(1), 53–77.

Porsch, R. (2010c). *Schreibkompetenzvermittlung im Englischunterricht in der Sekundarstufe I. Empirische Analysen zu Leistungen, Einstellungen, Unterrichtsmethoden und Zusammenhängen von Leistungen in der Mutter- und Fremdsprache.* Band 25. Reihe „Empirische Erziehungswissenschaft“. Münster: Waxmann.

Porsch, R. (2011). Texte verfassen im Französischunterricht am Gymnasium in der Sekundarstufe I: Eine Auswertung von Daten aus standardisierten Fragebögen. *Themenheft „Empirische Methoden und Ergebnisse in der Fremdsprachenforschung“ der Zeitschrift für Interkulturellen Fremdsprachenunterricht* 16(2), 50–65.

Porsch, R., Grotjahn, R. & Tesch, B. (2010). Hörverstehen und Hör-Sehverstehen in der Fremdsprache – unterschiedliche Konstrukte? *Zeitschrift für Fremdsprachenforschung*, 21(2), 143–190.

Porsch, R. & Köller, O. (2010a). Erste empirische Befunde der Pilotierungsstudie im Fach Französisch (Sekundarstufe I). In R. Porsch, B. Tesch & O. Köller (Hrsg.), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I* (S. 209–243). Münster: Waxmann.

Porsch, R. & Köller, O. (2010b). Standardbasiertes Testen von Schreibkompetenzen im Fach Englisch. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (S. 85–104). Münster: Waxmann.

Porsch, R. & Robitzsch, A. (2010). Umgang mit nicht bearbeiteten Textproduktionsaufgaben – Konsequenzen bei der Anwendung von Multi-Facetten-Raschmodellen. In B. Schwarz, P. Nenninger & R. S. Jäger (Hrsg.), *Erziehungswissenschaftliche Forschung – nachhaltige Bildung. Beiträge zur 5. DGfE-Sektionstagung „Empirische Bildungsforschung“.* *Erziehungswissenschaft* (Bd. 28, S. 305–312). Landau: Empirische Pädagogik.

Porsch, R. & Tesch, B. (2010a). Beschreibung der in der ersten Fremdsprache (Englisch/Französisch) untersuchten Kompetenzen. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 25–28). Münster: Waxmann.

Porsch, R. & Tesch, B. (2010b). Messung der Schreibkompetenz im Fach Französisch. In R. Porsch, B. Tesch & O. Köller (Hrsg.), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I* (S. 151–176). Münster: Waxmann.

Porsch, R., Tesch, B. & Köller, O. (2010a). Die Entwicklung von Kompetenzstufenmodellen zum Lese- und Hörverstehen im Fach Französisch. In R. Porsch, B. Tesch & O. Köller (Hrsg.), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I* (S. 244–267). Münster: Waxmann.

Porsch, R., Tesch, B. & Köller, O. (Hrsg.). (2010b). *Standard-basierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I*. Münster: Waxmann.

Preckel, F. & Vock, M. (2007). Diagnostik von Intelligenz und bereichsspezifischem Underachievement bei Hochbegabten mit dem Berliner Intelligenzstrukturtest für Jugendliche: Begabungs- und Hochbegabungsdagnostik (BIS-HB). In özbF (Hrsg.), *Versteckt, verkannt, verborgen. Erkennen und Fördern hochbegabter Underachiever*. Salzburg: özbF.

Preckel, F. & Vock, M. (2010). Intelligenzdiagnostik. In E. Walther, F. Preckel & S. Mecklenbräuker (Hrsg.), *Befragung von Kindern und Jugendlichen zwischen 5 und 15 Jahren* (S. 99–132). Göttingen: Hogrefe.

Reiss, K., Roppelt, A., Haag, N., Pant, H. A. & Köller, O. (2012). Kompetenzstufenmodelle im Fach Mathematik. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 72–84). Münster: Waxmann.

Reiss, K. & Winkelmann, H. (2009). Kompetenzstufenmodelle für das Fach Mathematik im Primarbereich. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 120–141). Weinheim: Beltz.

Richter, D. (2011). Lernen im Beruf. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse aus dem Forschungsprogramm COACTIV* (S. 55–68). Münster: Waxmann.

Richter, D. (2013). Professional development across the teaching career. In M. Kunter, J. Baumert, W. Blum, U. Klusmann & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 333–342). New York: Springer.

Richter, D., Engelbert, M., Böhme, K., Haag, N., Hannighofer, J., Reimers, H., Roppelt, A., Weirich, S., Pant, H. A. & Stanat, P. (2012). Anlage und Durchführung des Ländervergleichs. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 85–102). Münster: Waxmann.

Richter, D., Engelbert, M., Weirich, S. & Pant, H. A. (2013). Differentielle Teilnahme an Lehrerfortbildung und deren Zusammenhang zu professionsbezogenen Merkmalen von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 27(3), 193–207.

Richter, D. & Klein, R. (2013). Lehrerfortbildung im Ländervergleich: Ergebnisse einer Lehrkräftebefragung. *Schulverwaltung Baden-Württemberg*, 22(1), 2–4.

Richter, D., Kuhl, P., Haag, N. & Pant, H. A. (2013). Aspekte der Aus- und Fortbildung von Mathematik- und Naturwissenschaftslehrkräften im Ländervergleich. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 367–390). Münster: Waxmann.

Richter, D., Kuhl, P. & Pant, H. A. (2012). Soziale Disparitäten. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 191–207). Münster: Waxmann.

Richter, D., Kuhl, P., Reimers, H. & Pant, H. A. (2012). Aspekte der Aus- und Fortbildung von Lehrkräften in der Primarstufe. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 237–250). Münster: Waxmann.

Richter, D., Kunter, M., Anders, Y., Klusmann, U., Lüdtke, O. & Baumert, J. (2010). Inhalte und Prädiktoren beruflicher Fortbildung von Mathematiklehrkräften. *Empirische Pädagogik*, 24(2), 151–168.

Richter, D., Kunter, M., Klusmann, U., Lüdtke, O. & Baumert, J. (2011a). Professional development across the teaching career: Teachers' uptake of formal and informal learning opportunities. *Teaching and Teacher Education*, 27(1), 116–126.

Richter, D., Kunter, M., Klusmann, U., Lüdtke, O. & Baumert, J. (2011b). Soziale Unterstützung beim Berufseinstieg im Lehramt: Eine empirische Untersuchung zur Bedeutung von Mentoren und Mitreferendaren. *Zeitschrift für Erziehungswissenschaft*, 14(1), 35–59.

Richter, D., Kunter, M., Lüdtke, O., Klusmann, U., Anders, Y. & Baumert, J. (2013). How different mentoring approaches affect beginning teachers' development in the first years of practice. *Teaching and Teacher Education*, 36, 166–177.

Rjosk, C., McElvany, N., Anders, Y. & Becker, M. (2011). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung der basalen Lesefähigkeit ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 58(2), 92–105.

Robitzsch, A. (2009). Methodische Herausforderungen bei der Kalibrierung von Leistungstests. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 42–106). Weinheim: Beltz.

Robitzsch, A. & Rupp, A. A. (2009). The impact of missing data on the detection of differential item functioning. *Educational Psychological Measurement*, 69(1), 18–34.

Robitzsch, A. & Wilhelm, O. (2007). Review zu Borsboom, D. (2005) Measuring the mind. *Zeitschrift für Pädagogische Psychologie*, 21(2), 183–184.

Roick, T., Stanat, P., Dickhäuser, O., Frederking, V., Meier, C. & Steinhauer, L. (2010). Strukturelle und kriteriale Validität der literarästhetischen Urteilskompetenz. Projekt literarästhetische Urteilskompetenz. 56. Beiheft der Zeitschrift für Pädagogik, 165–174.

Rolff, H. G., Leucht, M. & Rösner, E. (2008). Sozialer und familiärer Hintergrund. In D. Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (S. 283–300). Weinheim: Beltz.

Roppelt, A. (2009a). Alles vergessen nach dem Abitur? Ein Vergleich der mathematischen Grundkompetenzen von Studierenden und Schülern. In M. Neubrand (Hrsg.), *Beiträge zum Mathematikunterricht 2009* (S. 235–244). Münster: WTM Verlag.

Roppelt, A. (2009b). Mathematische Grundkompetenzen von Studierenden. In A. Heinze & M. Grüßing (Hrsg.), *Mathematiklernen vom Kindergarten bis zum Studium – Kontinuität und Kohärenz als Herausforderung für den Mathematikunterricht* (S. 235–244). Münster: Waxmann.

Roppelt, A., Blum, W. & Pöhlmann, C. (2013). Beschreibung der untersuchten mathematischen Kompetenzen. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 23–37). Münster: Waxmann.

Roppelt, A., Keller, K., Leiß, D., Drüke-Noe, C. & Blum, W. (2007). *Kommentierung der schriftlichen Prüfungsarbeit zum mittleren Schulabschluss 2007 im Fach Mathematik (Berlin) bzw. der Prüfungen am Ende der Jahrgangsstufe 10 Mathematik (Brandenburg) hinsichtlich ihrer Orientierung an den länderübergreifenden Bildungsstandards*. Berlin: Institut für Schulqualität der Länder Berlin und Brandenburg.

Roppelt, A., Penk, C., Pöhlmann, C. & Pietsch, E. (2013). Der Ländervergleich im Fach Mathematik. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 123–140). Münster: Waxmann.

Roppelt, A. & Reiss, K. (2012). Beschreibung der im Fach Mathematik untersuchten Kompetenzen. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 34–48). Münster: Waxmann.

Rösch, H., Rotter, D. & Darsow, A. (2012). FoF und FoM: Konzeption der sprachsystematischen und fachbezogenen Zweitsprachförderung im BeFo-Projekt. In B. Ahrenholz (Hrsg.), *Sprachstand erheben – Spracherwerb erforschen* (S. 173–186). Stuttgart: Fillibach bei Klett.

Rösch, H. & Stanat, P. (2011). Bedeutung und Form (BeFo): Formfokussierte und bedeutungsfokussierte Förderung in Deutsch als Zweitsprache. In N. Hahn & T. Roelcke (Hrsg.), *Grenzen überwinden mit Deutsch. Beiträge der 37. Jahrestagung DaF an der PH Freiburg (MatDaF Bd. 85)* (S. 149–161). Göttingen: Universitätsverlag.

Rupp, A. A. (2007). The answer is in the question: A guide for describing and investigating the conceptual foundations and statistical properties of cognitive psychometric models. *International Journal of Testing*, 7(2), 95–125.

Rupp, A. A. & Böhme, K. (2008). Handbook of test development [Book review]. *International Journal of Testing*, 8, 111–125.

Rupp, A. A. & Mislevy, R. J. (2007). Cognitive foundations of structured item response theory models. In J. Leighton & M. Gierl (Eds.), *Cognitive diagnostic assessment in education: Theory and applications* (pp. 205–241). Cambridge: Cambridge University Press.

Rupp, A. A. & Porsch, R. (2010). Standard-setting item pool. In C. Harsch, H. A. Pant & O. Köller (Eds.), *Calibrating standards-based assessment tasks for English as a first foreign language. Standard-setting procedures in Germany* (Vol. 2, pp. 37–54). Münster: Waxmann.

Rupp, A. A. & Vock, M. (2007). National educational standards in Germany: Methodological challenges for developing and calibrating standards-based tests. In D. Waddington, P. Nentwig & S. Schanze (Eds.), *Making it comparable: Standards in science education* (pp. 173–198). Münster: Waxmann.

Rupp, A. A., Vock, M., Harsch, C. & Köller, O. (2008). *Developing standards-based assessment items for English as a first foreign language – Context, processes, and outcomes in Germany* (Vol. 1). Münster: Waxmann.

Schauber, S. K., Hecht, M., Nouns, Z. M. & Dettmer, S. (2013). On the role of biomedical knowledge in the acquisition of clinical knowledge. *Medical Education*, 47, 1223–1235.

Schipolowski, S. & Böhme, K. (2010). Der Ländervergleich im Fach Deutsch. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 87–96). Münster: Waxmann.

Schipolowski, S., Wilhelm, O. & Schroeders, U. (im Druck). BEFKI GC-K. Berliner Test zur Erfassung fluider und kristalliner Intelligenz – GC-Kurzskala. In C. J. Kemper, M. Zenger & E. Brähler (Hrsg.), *Psychologische und sozialwissenschaftliche Kurzskalen – Standardisierte Erhebungsinstrumente für Wissenschaft und Praxis*. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.

Schipolowski, S., Wilhelm, O., Schroeders, U., Kovaleva, A., Kemper, C. J. & Rammstedt, B. (2013). BEFKI GC-K: Eine Kurzskala zur Messung kristalliner Intelligenz. *Methoden, Daten, Analysen*, 7, 153–181.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O. & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 1089–1096.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M. & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, 136(3), 414–429.

Schöppe, D., Blatter, K., Faust, V., Jäger, D., Artelt, C., Schneider, W. & Stanat, P. (2012). Phonologische Sprachförderung bei Vorschulkindern deutscher und nichtdeutscher Herkunftssprache. In J. Heide, T. Fritzsche, C. B. Meyer & S. Ott (Hrsg.), *Spektrum Patholinguistik* (Bd. 5, S. 123–128). Potsdam: Universitätsverlag Potsdam.

Schöppe, D., Blatter, K., Faust, V., Jäger, D., Stanat, P., Artelt, C. & Schneider, W. (2013). Effekte eines Trainings der phonologischen Bewusstheit bei Vorschulkindern mit unterschiedlichem Sprachhintergrund. *Zeitschrift für Pädagogische Psychologie*, 27, 241–254.

Schroeder, L.-F. & Tesch, B. (2010a). Kompetenzaufgaben zum Leseverstehen. In F. J. Meißner & B. Tesch (Hrsg.), *Spanisch kompetenzorientiert unterrichten. Didaktische Grundlagen für die Aufgabenkonstruktion* (S. 184–193). Stuttgart: Klett.

Schroeder, L.-F. & Tesch, B. (2010b). Descubrir Avila. In F.-J. Meißner & B. Tesch (Hrsg.), *Spanisch kompetenzorientiert unterrichten. Didaktische Grundlagen für die Aufgabenkonstruktion*. Stuttgart: Klett.

Schroeder, L.-F. & Tesch, B. (2010c). Vacaciones en la Horcajada. In F.-J. Meißner & B. Tesch (Hrsg.), *Spanisch kompetenzorientiert unterrichten. Didaktische Grundlagen für die Aufgabenkonstruktion*. Stuttgart: Klett.

Schroeders, U. (2009a). Computergestützte Diagnostik und Förderung rechenschwacher Kinder am Beispiel von TeDDy-PC. In V. D. Schweiz (Hrsg.), *Dyskalkulie – Ansätze zu Diagnostik und Förderung in einer integrativen Schule* (S. 53–60). Zürich: Verband Dyslexie Schweiz.

Schroeders, U. (2009b). Testing for equivalence of test data across media. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment. Lessons learned from the PISA 2006 computer-based assessment of science (CBAS) and implications for large scale testing. JRC Scientific and Technical Report* (pp. 164–170). Luxembourg: Office for Official Publications of the European Communities.

Schroeders, U., Bucholtz, N., Formazin, M. & Wilhelm, O. (2013). Modality specificity of comprehension abilities in sciences. *European Journal of Psychological Assessment*, 29(1), 3–11.

Schroeders, U., Hecht, M., Heitmann, P., Jansen, M., Kampa, N., Klebba, N., Lenski, A. E. & Siegle, T. (2013). Der Ländervergleich in naturwissenschaftlichen Fächern. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 141–158). Münster: Waxmann.

Schroeders, U., Penk, C., Jansen, M. & Pant, H. A. (2013).

Geschlechtsbezogene Disparitäten. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 249–274). Münster: Waxmann.

Schroeders, U., Bucholtz, N., Formazin, M. & Wilhelm, O. (2013).

Modality specificity of comprehension abilities in sciences. *European Journal of Psychological Assessment*, 29(1), 3–11.

Schroeders, U. & Schneider, W. (2008). *TeDDy-PC. Test zur*

Diagnose von Dyskalkulie. Göttingen: Hogrefe.

Schroeders, U., Siegle, T., Weirich, S. & Pant, H. A. (2013).

Der Einfluss von Kontext- und Schülermerkmalen auf die naturwissenschaftlichen Kompetenzen. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 331–346). Münster: Waxmann.

Schroeders, U. & Wilhelm, O. (2010). Testing reasoning ability with handheld computers, notebooks, and paper and pencil.

European Journal of Psychological Assessment, 26(4), 284–292.

Schroeders, U. & Wilhelm, O. (2011a). Computer usage question-

naire: Structure, correlates, and gender differences. *Computers in Human Behavior*, 27(2), 899–904.

Schroeders, U. & Wilhelm, O. (2011b). Equivalence of reading and

listening comprehension across test media. *Educational and Psychological Measurement*, 71(5), 849–869.

Schroeders, U., Wilhelm, O. & Bucholtz, N. (2010). Reading,

listening, and viewing comprehension in English as a foreign language: One or more constructs? *Intelligence*, 38(6), 562–573.

Schroeders, U., Wilhelm, O. & Schipolowski, S. (2010). Internet-

based ability testing: Problems and opportunities. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced Methods for Conducting Online Behavioral Research* (pp. 131–148). Washington, DC: American Psychological Association.

Schubert, S., Ortwein, H., Dumitsch, A., Schwantes, U., Wilhelm, O. & Kiessling, C. (2008). A situational judgement test of

professional behaviour: development and validation. *Medical Teacher*, 30, 528–533.

Schulze, R., Wilhelm, O. & Kyllonen, P. C. (2007). New

approaches to assessing emotional intelligence. In G. Matthews, M. Zeidner & R. D. Roberts (Eds.), *Emotional intelligence: Knowns and unknowns* (pp. 305–338). Oxford: Oxford University Press.

Seeber, S., Jähnig, C. & Knigge, M. (2011). Validierung eines

Indikators zur Prozessqualität von Schule und Unterricht. In BMBF (Hrsg.), *Bildungsforschung Band 35. Vertiefende Studien zu ausgewählten Aspekten der Indikatorenentwicklung für den nationalen Bildungsbericht* (S. 17–56). Berlin und Bonn: Bundesministerium für Bildung und Forschung.

Segerer, R., Marx, A. & Marx, P. (2012). Unlösbare Items im KFT

4–12+R. *Diagnostica*, 58(1), 45–50.

Segerer, R., Marx, A., Stanat, P., Schneider, W., Roick, T. & Marx, P. (2013). Determinanten der Lesekompetenz bei Jugendlichen

nichtdeutscher Herkunftssprache: Zur Bedeutung der Spracherwerbsreihenfolge im Falle von Mehrsprachigkeit. 59. Beiheft der Zeitschrift für Pädagogik, 111–131.

Segeritz, M. & Pant, H. A. (2013). Do they feel the same way

about math? Testing measurement invariance of the PISA “Students’ Approaches to Learning” instrument across immigrant groups within Germany. *Educational and Psychological Measurement*, 73, 601–630.

Segeritz, M., Stanat, P. & Walter, O. (2010). Muster des schuli-

schen Erfolgs von Mädchen und Jungen mit Migrationshintergrund. 55. Beiheft der Zeitschrift für Pädagogik, 165–186.

Segeritz, M., Walter, O. & Stanat, P. (2010). Muster des schuli-

schen Erfolgs von jugendlichen Migranten in Deutschland: Evidenz für segmentierte Assimilation? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62(1), 113–138.

Siegle, T., Schroeders, U. & Roppelt, A. (2013). Anlage und

Durchführung des Ländervergleichs. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 101–122). Münster: Waxmann.

Stanat, P., Becker, M., Baumert, J., Lüdtke, O. & Eckhardt, A. G. (2012). Improving second language skills of immigrant students:

A field trial study evaluating the effects of a summer learning program. *Learning and Instruction*, 22(3), 159–170.

Stanat, P. & Edele, A. (2011). Soziale Ungleichheit im Bildungs-

wesen – Migration. In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Lehrbuch Empirische Bildungsforschung* (S. 181–192). Opladen: VA Verlag.

- Stanat, P. & Felbrich, A. (2013).** Sprachförderung als Voraussetzung für die Sicherung von Mindeststandards im Bildungssystem: Ansatzpunkte und Herausforderungen. In D. Deißner & M. Speich (Hrsg.), *Chancen bilden. Wege zu einer gerechteren Bildung – ein internationaler Erfahrungsaustausch* (S. 79–100). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stanat, P. & Lüdtke, O. (2013).** International large-scale assessment studies of student achievement. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 481–483). New York: Routledge.
- Stanat, P. & Pant, H. A. (2010).** Externe Leistungsfeststellungen in deutschen Schulen. Potenziale und Risiken der testgestützten Qualitätssicherung. *Pädagogik*, 62(11), 40–45.
- Stanat, P., Pant, H. A., Böhme, K. & Richter, D. (Hrsg.). (2012).** *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011*. Münster: Waxmann.
- Stanat, P., Pant, H. A. & Richter, D. (2011).** Educational research continues coming to LIFE: The Institute for Educational Quality Improvement (IQB) as a new LIFE member. *LIFE newsletter*, 5, 9–12.
- Stanat, P., Pant, H. A., Richter, D. & Böhme, K. (2012).** IQB-Ländervergleich 2011: Zusammenfassung und Einordnung der Befunde. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 291–300). Münster: Waxmann.
- Stanat, P., Pant, H. A., Richter, D., Böhme, K., Engelbert, M., Haag, N., Hannighofer, J., Prengel, J., Reimers, H., Roppelt, A. & Weirich, S. (2012).** Der Blick in die Länder. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 131–172). Münster: Waxmann.
- Stanat, P., Pant, H. A., Richter, D., Pöhlmann, C. & Kuhl, P. (2013).** Was kann das IQB leisten? In S. L. Klitzing, D. S. Di Fuccia & G. M. Frerich (Hrsg.), *Zur Vermessung von Schule* (S. 125–152). Bad Heilbrunn: Klinkhardt.
- Stanat, P., Rauch, D. & Segeritz, M. (2010).** Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 200–230). Münster: Waxmann.
- Stanat, P., Schwippert, K. & Gröhlich, C. (2010).** Der Einfluss des Migrantenanteils in Schulklassen auf den Kompetenzerwerb: Längsschnittliche Überprüfung eines umstrittenen Effekts. *55. Beiheft der Zeitschrift für Pädagogik*, 147–164.
- Stanat, P., Segeritz, M. & Christensen, G. (2010).** Schulbezogene Motivation und Aspiration von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (S. 31–57). Münster: Waxmann.
- Stanat, P., Weirich, S. & Radmann, S. (2012).** Sprach- und Leseförderung. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 251–276). Münster: Waxmann.
- Sturzbacher, D., Kleeberg-Niepage, A. & Hoffmann, L. (Hrsg.). (2012).** *Aufschwung Ost? Lebenssituation und Wertorientierung ostdeutscher Jugendlicher*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Sumfleth, E., Klebba, N., Kauertz, A., Mayer, J., Fischer, H. E., Walpuski, Wellnitz, N. (2013).** Das Kompetenzstrukturmodell in den naturwissenschaftlichen Fächern. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 38–41). Münster: Waxmann.
- Tarasow, T., Borghs, M. & Laouris, Y. (2013).** SMS or Apps? A comparison using SWOT analysis. In W. Kinuthia & S. Marshall (Eds.), *On the move. Mobile learning for development* (pp. 43–68). Charlotte: Information Age Publishing.
- Tesch, B. (2007).** Bildungsstandards: die Kastanien im Feuer. Bedingungen gelingender Implementation. *Der Fremdsprachliche Unterricht*, 88, 9–13.
- Tesch, B. (2009a).** Die Tests des IQB – Sinn und Nutzen. *Praxis Fremdsprachenunterricht*, 1, 41–45.
- Tesch, B. (2009b).** „Tout le monde ist doch nicht die ganze Welt, oder?“ Rekonstruktive Studien im Dienste der Unterrichts- und Lernaufgabenforschung. In A. Berndt & K. Kleppin (Hrsg.), *Sprachlehrforschung: Theorie und Empirie. Festschrift für Rüdiger Grotjahn* (S. 239–248). Frankfurt a. M.: Peter Lang.

Tesch, B. (2010a). Ablauf und Organisation der Testentwicklung am IQB im Projekt Französisch. In R. Porsch, B. Tesch & O. Köller (Hrsg.), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I* (S. 75–81). Münster: Waxmann.

Tesch, B. (2010b). Aufgabenorientierung: Übungsaufgaben, Testaufgaben und Lernaufgaben. In F. J. Meißner & B. Tesch (Hrsg.), *Spanisch kompetenzorientiert unterrichten. Didaktische Grundlagen für die Aufgabenkonstruktion* (S. 57–71). Stuttgart: Klett.

Tesch, B. (2010c). *Kompetenzorientierte Lernaufgaben im Fremdsprachenunterricht. Konzeptionelle Grundlagen und eine rekonstruktive Fallstudie zur Unterrichtspraxis (Französisch)*. Frankfurt a. M.: Peter Lang.

Tesch, B. (2010d). Leseverstehen. In F. J. Meißner & B. Tesch (Hrsg.), *Spanisch kompetenzorientiert unterrichten. Didaktische Grundlagen für die Aufgabenkonstruktion* (S. 87–95). Stuttgart: Klett.

Tesch, B. (2010e). Sprechen. In F. J. Meißner & B. Tesch (Hrsg.), *Spanisch kompetenzorientiert unterrichten. Didaktische Grundlagen für die Aufgabenkonstruktion*. Stuttgart: Klett (S. 96–112). Stuttgart: Klett.

Tesch, B. (2010f). Tests für Französisch auf Grundlage des Gemeinsamen europäischen Referenzrahmens für Sprachen (GER). In R. Porsch, B. Tesch & O. Köller (Hrsg.), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I* (S. 69–74). Münster: Waxmann.

Tesch, B. (2010g). Testung der funktionalen kommunikativen Kompetenzen. In R. Porsch, B. Tesch & O. Köller (Hrsg.), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I* (S. 81–90). Münster: Waxmann.

Tesch, B. (2010h). „Wir reden zu wenig“. Ergebnisse einer empirischen Studie als Anregung für das eigene Unterrichten. *Praxis Fremdsprachenunterricht*, 7, 7–10.

Tesch, B. (2011). Sind Fehler für die Bewertung out? Pro-Position. *Fremdsprachen lehren und lernen*, 40, 134.

Tesch, B. (2012). Sprachendidaktische Innovations- und Transferforschung. Ein Forschungsbereich mit Zukunftspotenzial. *Zeitschrift für Fremdsprachenforschung*, 23, 77–98.

Tesch, B. & Grotjahn, R. (2010). Messung der fremdsprachlichen Sprechkompetenz im Fach Französisch. In R. Porsch, B. Tesch & O. Köller (Hrsg.), *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I* (S. 177–205). Münster: Waxmann.

Tesch, B., Leupold, E. & Köller, O. (Hrsg.). (2008). *Bildungsstandards Französisch konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsideen und Fortbildungsmöglichkeiten*. Berlin: Cornelsen Scriptor.

Tesch, B., Porsch, R., Leupold, E., Grotjahn, R., Kleppin, K., Frenzel, J., Harsch, C., Leucht, M., Pant, H. A., Rupp, A. A., Tiffin-Richards, S. P. & Köller, O. (2010). Kompetenzstufenmodelle für Lese- und Hörverstehen in der ersten Fremdsprache. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 50–64). Münster: Waxmann.

Thonke, F. & Hunger, S. (2010). Lernstandserhebungen. Hamburg macht Schule. *Zeitschrift für Hamburger Lehrkräfte und Elternräte*, 3, 41–42.

Tiffin-Richards, S. P. (2010). The bookmark standard-setting method. In C. Harsch, H. A. Pant & O. Köller (Eds.), *Calibrating standards-based assessment tasks for english as a first foreign language. Standard-setting procedures in Germany* (Vol. 2, pp. 57–64). Münster: Waxmann.

Tiffin-Richards, S. P. & Köller, O. (2010). Comparison and synthesis of multiple standard-setting methods and panels. In C. Harsch, H. A. Pant & O. Köller (Eds.), *Calibrating standards-based assessment tasks for english as a first foreign language – Standard-setting procedures in Germany* (Vol. 2, pp. 107–110). Münster: Waxmann.

Tiffin-Richards, S. P., Pant, H. A. & Köller, O. (2013). Setting standards for english foreign language assessment: Methodology, validation and a degree of arbitrariness. *Educational Measurement: Issues and Practice*, 32(2), 15–25.

Trautwein, U., Köller, O., Lehmann, R. H. & Lüdtke, O. (Hrsg.). (2007). *Schulleistungen von Abiturienten. Regionale, schulformbezogene und soziale Disparitäten*. Münster: Waxmann.

van den Heuvel-Panhuizen, M., Robitzsch, A., Treffers, A. & Köller, O. (2009). Large-scale assessment of change in student achievement: Dutch primary school students' results on written division in 1997 and 2004 as an example. *Psychometrika*, 74(2), 351–365.

- van Horen, F., Pöhlmann, C., Koeppen, K. & Hannover, B. (2008).** Importance of personal goals in people with independent versus interdependent selves. *Social Psychology*, 39(4), 222–230.
- Vock, M. (2008a).** Berufsinteressen hochbegabter Jugendlicher [Buchrezension]. *Zeitschrift für Pädagogische Psychologie*, 21(3-4), 315–317.
- Vock, M. (2008b).** Effekte schulischer Förderprogramme für leistungsstarke und besonders begabte Schülerinnen und Schüler. In H. Ullrich & S. Strunck (Hrsg.), *Begabtenförderung an Gymnasien. Entwicklungen, Befunde, Perspektiven* (S. 78–99). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Vock, M. (2008c).** SON-R 2-7. Non-verbaler Intelligenztest. *Diagnostica*, 54(2), 112–115.
- Vock, M., Gauck, L. & Vogl, K. (2010).** Diagnostik von Schulleistungen und Underachievement. In F. Preckel, W. Schneider & H. Holling (Hrsg.), *Jahrbuch der Pädagogischen Diagnostik: Tests & Trends, Band Hochbegabung* (S. 1–17). Göttingen: Hogrefe.
- Vock, M. & Hasselhorn, M. (2010).** Diagnostik der Funktionstüchtigkeit des Arbeitsgedächtnisses bei Hochbegabten. In F. Preckel, W. Schneider & H. Holling (Hrsg.), *Jahrbuch der Pädagogischen Diagnostik: Tests & Trends, Band Hochbegabung* (S. 119–140). Göttingen: Hogrefe.
- Vock, M. & Holling, H. (2007).** Begabung und Berufserfolg. In K. A. Heller & A. Ziegler (Hrsg.), *Begabt sein in Deutschland* (S. 233–264). Münster: Lit-Verlag.
- Vock, M. & Holling, H. (2008).** The measurement of visuo-spatial and verbal-numerical Working Memory: Development of IRT based scales. *Intelligence*, 36(2), 161–182.
- Vock, M., Hupbach, A. & Mecklenbräuker, S. (2010).** Gedächtnis. In E. Walther, F. Preckel & S. Mecklenbräuker (Hrsg.), *Befragung von Kindern und Jugendlichen zwischen 5 und 15 Jahren* (S. 201–226). Göttingen: Hogrefe.
- Vock, M., Köller, O. & Nagy, G. (2013).** Vocational interests of intellectually gifted and high achieving young adults. *British Journal of Educational Psychology*, 83, 305–328.
- Vock, M. & Pant, H. A. (2008).** Standardbasierte Aufgaben, Leistungsvergleiche und Schulinspektionen. Potentiale der Ländermaßnahmen für die schulische Qualitätssicherung. In G. Weißeno (Hrsg.), *Politikkompetenz: Was Unterricht zu leisten hat* (S. 310–330). Wiesbaden: VS Verlag.
- Vock, M., Penk, C. & Köller, O. (im Druck).** Wer überspringt eine Schulklasse? Befunde zum Klassenüberspringen in Deutschland. *Psychologie in Erziehung und Unterricht*.
- Vock, M., Preckel, F. & Holling, H. (2007).** Förderung Hochbegabter in der Schule. *Evaluationsbefunde und Wirksamkeit von Maßnahmen*. Göttingen: Hogrefe.
- Vock, M., Preckel, F. & Holling, H. (2011).** Mental abilities and school achievement: A test of a mediation hypothesis. *Intelligence*, 39, 357–369.
- Wäckerle, M., Tesch, B. & Rjosk, C. (2012).** Kompetenzorientierte Lernaufgaben für die gymnasiale Oberstufe. Konzeption, Innovation, Erprobung. In M. Bär, A. Bonnet, H. Decke-Cornill, A. Grünewald & A. Hu (Hrsg.), *Globalisierung – Migration – Fremdsprachenunterricht. Dokumentation zum 24 Kongress für Fremdsprachendidaktik der Deutschen Gesellschaft für Fremdsprachenforschung (DGFF) Hamburg, 28. September – 1. Oktober 2011* (S. 369–381). Baltmannsweiler: Schneider-Verlag Hohengehren.
- Walpuski, M., Kampa, N., Kauertz, A. & Wellnitz, N. (2008).** Evaluation der Bildungsstandards in den Naturwissenschaften. *MNU*, 61(6), 323–326.
- Walpuski, M., Kauertz, A., Kampa, N., Fischer, H. E., Mayer, J., Sumfleth, E. & Wellnitz, N. (2010).** ESNaS – Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In A. Gehrmann, U. Hericks & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 171–184). Bad Heilbrunn: Klinkhardt.
- Walpuski, M. & Sumfleth, E. (2013).** Die Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 46–47). Münster: Waxmann.
- Walpuski, M., Sumfleth, E. & Pant, H. A. (2013).** Kompetenzstufenmodelle für das Fach Chemie. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 83–91). Münster: Waxmann.
- Walther, G. & Granzer, D. (2009).** Kompetenzmodell Mathematik. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 108–119). Weinheim: Beltz.

Walther, G., van den Heuvel-Panhuizen, M., Granzer, D. & Köller, O. (Hrsg.). (2007). *Bildungsstandards für die Grundschule: Mathematik konkret*. Berlin: Cornelsen Scriptor.

Weirich, S., Haag, N. & Roppelt, A. (2012). Testdesign und Auswertung des Ländervergleichs: Technische Grundlagen. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 277–290). Münster: Waxmann.

Weirich, S., Hecht, M. & Böhme, K. (im Druck). Modeling Item Position Effects Using Generalized Linear Mixed Models. *Applied Psychological Measurement*.

Weirich, S., Hoffman, F., Meißner, L., Heinz, A. & Bengner, T. (2011). Sex influence on face recognition memory moderated by presentation duration and re-encoding. *Neuropsychology*, 25(6), 806–813.

Wellnitz, N., Fischer, H. E., Kauertz, A., Mayer, J., Neumann, I., Pant, H. A., Sumfleth, E. & Walpuski, M. (2012). Evaluation der Bildungsstandards – eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung. *Zeitschrift für Didaktik der Naturwissenschaften*, 18(2), 261–291.

Wellnitz, N., Hartmann, S. & Mayer, J. (2010). Developing a paper-and-pencil-test to assess students' skills in scientific inquiry. In G. Çakmakci & M. F. Taşar (Eds.), *Contemporary science education research: learning and assessment* (pp. 289–294). Ankara: Pegem Akademi.

Wichert, J. M. & Wilhelm, O. (2007). What is the national g-factor? *European Journal of Personality*, 21(5), 763–765.

Wilhelm, O. (2009). Issues in computerized ability measurement: Getting out of the jingle and jangle jungle. In F. Scheuermann & A. G. Pereira (Eds.), *The transition to computer-based assessment. Lessons learned from the PISA 2006 computer-based assessment of science (CBAS) and implications for large scale testing. JRC Scientific and Technical Report* (pp. 145–150). Luxembourg: Office for Official Publications of the European Communities.

Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A. & Sommer, W. (2010). Individual differences in perceiving and recognizing faces-one element of social cognition. *Journal of Personality and Social Psychology*, 99(3), 530–548.

Wilhelm, O., Hildebrandt, A., Herzmann, G. & Sommer, W. (2010). Computergestütztes Testverfahren zur Erfassung gesichterspezifischer Denkleistungen – Berlin Face Test (BeFaT). *Diagnostica*, 56(2), 119–122.

Wilhelm, O., Hülür, G., Köller, O. & Radalewski, M. (2010). Empirische Grundlagen zum Orientierungskurs- und Einwanderungstest. In G. Weißeno (Hrsg.), *Bürgerrolle in der Einwanderungsgesellschaft* (S. 49–64). Bonn: Bundeszentrale für politische Bildung.

Wilhelm, O. & Kunina, O. (2009). Pädagogisch-psychologische Diagnostik. In E. Wild & J. Möller (Hrsg.), *Einführung in die Pädagogische Psychologie* (S. 307–332). Heidelberg: Springer

Wilhelm, O. & Robitzsch, A. (2009). Have cognitive diagnostic models delivered their goods? Some substantial and methodological concerns. *Measurement*, 7(1), 53–57.

Wilhelm, O. & Schipolowski, S. (2010). Intelligenzdiagnostik in der Pädagogischen Psychologie. In G. L. Huber (Hrsg.), *Enzyklopädie Erziehungswissenschaft Online. Fachgebiet Pädagogische Psychologie*. Weinheim: Juventa.

Wilhelm, O. & Schroeders, U. (2008). Computerized ability measurement: Some substantive dos and don'ts. In F. Scheuermann & A. G. Pereira (Eds.), *Towards a research agenda in computer-based assessment. Challenges and needs for European Educational Measurement. JRC Scientific and Technical Report* (pp. 76–84). Luxembourg: Office for Official Publications of the European Communities.

Wilhelm, O. & Schroeders, U. (2010). Bildung in Zeiten neuer Medien aus kognitionspsychologischer Sicht. In J.-D. Gauer & J. Kraus (Hrsg.), *Bildung und Unterricht in Zeiten von Google und Wikipedia* (S. 27–45). Sankt Augustin: Konrad-Adenauer-Stiftung.

Wilhelm, O., Schroeders, U. & Schipolowski, S. (im Druck). BEFKI 8–10. *Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 8. bis 10. Jahrgangsstufe*. Göttingen: Hogrefe.

Wilhelm, O., Witthöft, M. & Schipolowski, S. (2010). Self-reported cognitive failures: Competing measurement models and self-report correlates. *Journal of Individual Differences*, 31(1), 1–14.

Winkelmann, H. & Groeneveld, I. (2010). Geschlechterdisparitäten. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 177–184). Münster: Waxmann.

Winkelmann, H. & Robitzsch, A. (2009a). Kompetent ... kompetenter. Eine Erhebung der Klassenstufenunterschiede mathematischer Kompetenzen. *Grundschule*, 41(6), 24–28.

Winkelmann, H. & Robitzsch, A. (2009b). Modelle mathematischer Kompetenzen: Empirische Befunde zur Dimensionalität. In D. Granzer, O. Köller, A. Bremerich Vos, M. van den Heuvel Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 169–196). Weinheim: Beltz.

Winkelmann, H., Robitzsch, A., Stanat, P. & Köller, O. (2012). Mathematische Kompetenzen in der Grundschule. Struktur, Validierung und Zusammenspiel mit allgemeinen kognitiven Fähigkeiten. *Diagnostica*, 58(1), 15–30.

Winkelmann, H. & van den Heuvel-Panhuizen, M. (2009a). Die Erfassung mathematischen Argumentierens. In D. Granzer, O. Köller, A. Bremerich Vos, M. van den Heuvel Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 157–168). Weinheim: Beltz.

Winkelmann, H. & van den Heuvel-Panhuizen, M. (2009b). Geschlechtsspezifische mathematische Kompetenzen. Leistungsmessung in der Grundschule. In D. Granzer, O. Köller, A. Bremerich Vos, M. van den Heuvel Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 142–156). Weinheim: Beltz.

Winkelmann, H., van den Heuvel-Panhuizen, M. & Robitzsch, A. (2008). Gender differences in the mathematics achievements of German primary school students: results from a German large-scale study. *ZDM – The International Journal on Mathematics Education*, 40, 601–616.

Wolf, K. M., Felbrich, A., Stanat, P. & Wendt, W. (2011). Evaluation der kompensatorischen Sprachförderung in Brandenburger Kindertagesstätten. *Empirische Pädagogik*, 25(4), 423–438.

Wurster, S. & Gärtner, H. (2013a). Erfassung von Bildungsprozessen im Rahmen von Schulinspektion und deren potenzieller Nutzen für die empirische Bildungsforschung. *Unterrichtswissenschaft*, 41(3), 216–234.

Wurster, S. & Gärtner, H. (2013b). Schulen im Umgang mit Schulinspektion und deren Ergebnissen. *Zeitschrift für Pädagogik*, 59(3), 425–445.

Wurster, S., Richter, D., Schliesing, A. & Pant, H. A. (2013). Nutzung unterschiedlicher Evaluationsdaten an Berliner und Brandenburger Schulen. Rezeption und Nutzung von Ergebnissen aus Schulinspektion, Vergleichsarbeiten und interner Evaluation im Vergleich. *Die Deutsche Schule*, 12. Beiheft, 19–50.

Zeitler, S., Asbrand, B. & Pöhlmann, C. (2009). Unterrichtsentwicklung durch unterstützte Kooperation in Fachgruppen? Ein Projekt zur Implementation der Bildungsstandards in Berliner und Brandenburger Schulen. In K. Maag Merki (Hrsg.), *Kooperation und Netzwerkbildung. Strategien zur Qualitätsentwicklung in Schulen* (S. 14–25). Seelze: Klett Kallmeyer.

Zeitler, S., Heller, N. & Asbrand, B. (2009). Evaluation im Kontext von Unterrichtsentwicklung. In U. Klinger (Hrsg.), *Mit Kompetenz Unterricht entwickeln; Fortbildungskonzepte und -materialien* (S. 55–67). Troisdorf: Bildungsverlag Eins.

Zeitler, S., Köller, O. & Tesch, B. (2010). Bildungsstandards und ihre Implikationen für Qualitätssicherung und Qualitätsentwicklung. In A. Gehrmann, U. Hericks & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 23–35). Bad Heilbrunn: Klinkhardt.

Abgeschlossene Dissertationen

Böhme, K. (2011). Methodische und didaktische Überlegungen sowie empirische Befunde zur Erfassung sprachlicher Kompetenzen im Deutschen. Dissertation, Humboldt-Universität zu Berlin.

Groeneveld, I. (2011). Moderation sozialer Disparitäten in der Primarstufe. Dissertation, Humboldt-Universität zu Berlin.

Heitmann, P. (2012). Bewertungskompetenz im Rahmen naturwissenschaftlicher Problemlöseprozesse. Modellierung und Diagnose der Kompetenzen Bewertung und analytisches Problemlösen für das Fach Chemie. Berlin: Logos.

Henschel, S. (2013). Effekte motivationaler und affektiver Merkmale auf das Verstehen von literarischen und faktualen Texten. Dissertation, Freie Universität Berlin.

Kampa, N. N. (2012). Aspekte der Validierung eines Tests zur Kompetenz in Biologie – eine Studie zur Kompetenz in Biologie und ihren Teilkompetenzen Konzept- und Prozesswissen. Dissertation, Humboldt-Universität zu Berlin.

Leucht, M. (2010). Diagnostik rezeptiver Kompetenzen in der ersten Fremdsprache Englisch auf Basis des Gemeinsamen Europäischen Referenzrahmens für Sprachen. Dissertation, Freie Universität Berlin.

Neumann, D. (2012). Schwierigkeitsbeeinflussende Merkmale bei Aufgaben zum Hörverstehen im Fach Deutsch in der Sekundarstufe I. Dissertation, Humboldt-Universität zu Berlin.

Porsch, R. (2010). Schreibkompetenzvermittlung im Englischunterricht in der Sekundarstufe I. Dissertation, Humboldt-Universität zu Berlin.

Tiffin-Richards, S. P. (2011). Setting standards for the assessment of English as a foreign language: Establishing validity evidence for criterion-referenced interpretations of test-scores. Dissertation, Freie Universität zu Berlin.

Winkelmann, H. (2009). Validierung der länderübergreifenden Bildungsstandards für mathematische Kompetenzen im Primarbereich. Dissertation, Humboldt-Universität zu Berlin.

Zeitler, S. (2010): Bildungsstandards in der Schule. Eine rekonstruktive Studie zur Implementation der Bildungsstandards. Dissertation, Humboldt-Universität zu Berlin.



08 // Anhang

Ehemalige Mitarbeiterinnen und Mitarbeiter am IQB // Aktuelle Dissertationsvorhaben von
IQB-Mitarbeiterinnen und IQB-Mitarbeitern // Wissenschaftliche Fachtagungen (Auswahl)
mit Beiträgen von IQB-Mitarbeiterinnen und Mitarbeitern // Gastvorträge am IQB (Auswahl) //
Ausgerichtete Tagungen // Lehrveranstaltungen

Ehemalige Mitarbeiterinnen und Mitarbeiter am IQB

Direktion

Prof. Dr. Olaf Köller, Geschäftsführender Direktor des Leibniz-Instituts für die Pädagogik der Naturwissenschaften und Mathematik (IPN) und Professor für Empirische Bildungsforschung, Christian-Albrechts-Universität zu Kiel

Verwaltung

Anne Dressler, Mitarbeiterin der Fakultätsverwaltung der Philosophischen Fakultät I, Humboldt-Universität zu Berlin

Martina Rosin, im Ruhestand

Fränze Schröder, Sozialpädagogin bei Kindeswohl-Berlin e. V., Berlin

Catherin Thäter, Sekretariat Abteilung Erziehungswissenschaft und Pädagogisch-Psychologische Methodenlehre, Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN), Christian-Albrechts-Universität zu Kiel

EDV

Hannes Schurig, IT-Administrator, Barketing IMS GmbH, Berlin

Fachkoordination

Gabriele Gippner, Schulleiterin der Senefelder-Schule, Treuchtlingen

Dr. Dietlinde Granzer, Wissenschaftliche Leiterin des Instituts für Angewandtes Schulmanagement (ifas), Stuttgart

Dr. Anke Greve, Ministerium für Bildung, Jugend und Sport, Brandenburg (Koordination, Steuerung der Schulaufsicht, Qualitätsentwicklung und -sicherung), Potsdam

Dr. Claudia Harsch, Associate Professor am Centre for Applied Linguistics, University of Warwick, UK

Henriette Hartmann, Buchhaltung und Controlling im Kulturzentrum Tollhaus e. V., Karlsruhe

Ralph Hartung, Schulleiter der Goetheschule, Neu-Isenburg

Nicola Klebba, Niedersächsisches Kultusministerium, Referat 31, Eigenverantwortung und Unterstützung der Schulen; Qualitätsentwicklung, Hannover

Dr. Walther Kösters, Qualitätsmanager Campus Lichtenberg an der Hochschule für Wirtschaft und Recht, Berlin

Ralf Machnik, Abteilungsleiter Katholische Schulen im Dezernat Schule und Bildung im Bischöflichen Ordinariat Limburg und Pädagogischer Leiter der St. Hildegard-Schulgesellschaft, Limburg

Dr. Daniela Neumann, Leitung des Sachgebiets Zentrale Lernstandserhebungen, Landesschulamt und Lehrkräfteakademie Wiesbaden

Elke Pietsch, Lehrkraft für besondere Aufgaben, Universität Rostock

Petra Schattschneider, Fachleiterin Spanisch, Zentrum für schulpraktische Lehrerbildung, Bonn

Prof. Dr. Bernd Tesch, Institut für Romanistik, Didaktik des Französischen und Spanischen, Universität Kassel

Frank Weigand, Schulleiter der Deutschen Schule in Santa Cruz de la Sierra, Bolivien

Wissenschaftliche Mitarbeiterinnen und Mitarbeiter

Prof. Dr. Barbara Asbrand, Professorin für Erziehungswissenschaft, Goethe-Universität Frankfurt a. M.

Dr. Tobias Bothe-Hutschenreuter, Projektmanager am Kienbaum Institut für Management Diagnostik, Düsseldorf

Dr. Maren Formazin, Mitarbeiterin an der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Fachgruppe Statistische Methoden in Epidemiologie und Biometrie, Berlin

Jenny Frenzel, Studienrätin an einer Bremer Schule und freie Mitarbeiterin in der Forschungsgruppe Berufsbildungsforschung (i:BB) an der Universität Bremen

Fidan Gasimova, Wissenschaftliche Mitarbeiterin am Institut für Differentielle Psychologie und Psychologische Diagnostik, Universität Ulm

Dany Grimm, Psychotherapeutin in Ausbildung am Zentrum für Psychotherapie am Institut für Psychologie (ZPHU), Humboldt-Universität zu Berlin

Dr. Imke Groeneveld, Wissenschaftliche Mitarbeiterin am Fachbereich Informationswissenschaften, FH Potsdam

Jasmin Hannighofer, Wissenschaftliche Mitarbeiterin am Institut für Psychologie, Technische Universität Braunschweig

Prof. Dr. Andrea Hildebrandt, Juniorprofessorin für Psychologische Diagnostik und Persönlichkeitspsychologie, Ernst-Moritz-Arndt Universität Greifswald

Dr. Gizem Hülür, Wissenschaftliche Mitarbeiterin am Institut für Psychologie, Entwicklungs- und Pädagogische Psychologie, Humboldt-Universität zu Berlin

Dr. Nele Kampa, Wissenschaftliche Mitarbeiterin am Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN), Christian-Albrechts-Universität zu Kiel

Dr. Doris Keye, Mitarbeiterin am German Aerospace Center (DLR), Hamburg

Prof. Dr. Michel Knigge, Professor für Pädagogische Psychologie in der Erziehungswissenschaft, Martin-Luther-Universität Halle-Wittenberg

Dr. Olga Kunina-Habenicht, Wissenschaftliche Mitarbeiterin im Projekt „Bildungswissenschaftliches Wissen und der Erwerb professioneller Kompetenz in der Lehramtsausbildung“, Goethe-Universität Frankfurt a. M.

Daniela Lange, Koordinatorin im Regionalmanagement Berlin Schöneeweide, WISTA-Management GmbH, Berlin

Dr. Michael Leucht, Wissenschaftlicher Mitarbeiter am Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN), Christian-Albrechts Universität zu Kiel

Dr. Raphaela Porsch, Institut für Erziehungswissenschaft, Abteilung Schulpädagogik / Schul- und Unterrichtsforschung, Westfälische Wilhelms-Universität Münster

Andrea Prater, Referentin in der Agentur für Qualitätssicherung durch Akkreditierung von Studiengängen, Köln

Josefine Prengel, Sonderpädagogin an der Grundschule Neues Tor, Berlin

Melanie Radalewski, Psychologin am Medizinisch & pädagogisch therapeutischen Kompetenzzentrum, Oensingen, Schweiz

Claudia Rienäcker, Wissenschaftliche Mitarbeiterin am Institut für Psychologie, Klinische Gerontopsychologie, Humboldt-Universität zu Berlin

Alexander Robitzsch, Wissenschaftlicher Mitarbeiter am Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens (BIFIE), Zentrum für Bildungsmonitoring und Bildungsstandards, Salzburg

Prof. Dr. André A. Rupp, Research Director for the Innovations in the Development and Evaluation of Automated Scoring (IDEAS), Group at the Educational Testing Service (ETS), University of Maryland, USA

Dr. Simon Tiffin-Richards, Wissenschaftlicher Mitarbeiter in der Max-Planck-Forschungsgruppe Schriftspracherwerb und Lesentwicklung (REaD), Berlin

Prof. Dr. Marja van den Heuvel-Panhuizen, Professor of Mathematics Education, Freudenthal Institute for Science and Mathematics Education, Faculty of Science & Faculty of Social and Behavioural Sciences, Utrecht University

Prof. Dr. Miriam Vock, Professorin für Empirische Unterrichts- und Interventionsforschung, Universität Potsdam

Prof. Dr. Oliver Wilhelm, Professor für Differentielle Psychologie und Psychologische Diagnostik, Universität Ulm

Dr. Henrik Winkelmann, Psychotherapeut in Ausbildung, Wiesbaden

Dr. Sigrid Zeitler, Wissenschaftliche Mitarbeiterin am Lehrstuhl für Schulpädagogik, Friedrich-Alexander-Universität Erlangen-Nürnberg

Aktuelle Dissertationsvorhaben von IQB-Mitarbeiterinnen und IQB-Mitarbeitern

Name	Vorname	Thema der Dissertation (Arbeitstitel)	Arbeitsbereich am IQB
Maluch	Jessica	Die Rolle von Bilingualismus für den weiteren Fremdspracherwerb	Sekundarstufe I Englisch
Canz	Thomas	Validitätsaspekte bei der Messung von Schreibkompetenzen	Sekundarstufe I Deutsch
Haag	Nicole	Effekte sprachlicher Komplexität von Mathematikaufgaben auf die Messung mathematischer Kompetenzen	Primarstufe Mathematik
Hecht	Martin	Optimization of measurement instruments in large-scale assessments	Sekundarstufe I Naturwissenschaften
Hoffmann	Lars	Untersuchungen zur diagnostischen Kompetenz von Lehrkräften	Sekundarstufe II Deutsch
Jansen	Malte	Das akademische Selbstkonzept in den naturwissenschaftlichen Fächern: Strukturmodelle und Korrelate	Sekundarstufe I Naturwissenschaften
Kocaj	Aleksander	Schulische Entwicklung von Kindern mit sonderpädagogischem Förderbedarf im Vergleich zwischen Regel- und Förderschulen	Forschungsdatenzentrum (FDZ)
Kroth	Anna	Tuition Fees and Their Effect on Social and Gender Disparities in College Enrollment in Germany – Results from a Natural Experiment	Koordinierungsstelle ZIB, IQB
Lenski	Anna	Implementation der Bildungsstandards	Sekundarstufe I Naturwissenschaften
Neugebauer	Bettina	Der Einfluss von Eltern auf Schülerleistungen	Sekundarstufe I Französisch
Penk	Christiane	Die Rolle von Testteilnahmemotivation in Large-Scale-Assessments	Sekundarstufe I Mathematik
Rjosk	Camilla	Klassenkomposition, Unterricht und Bildungsergebnisse: Die Rolle der Lernumwelt für die Leistungs- und psychosoziale Entwicklung von Schülerinnen und Schülern mit und ohne Migrationshintergrund	Sekundarstufe II Englisch, Sekundarstufe I Französisch
Roppelt	Alexander	Nachhaltigkeit schulischer Kompetenzen: Mathematikkompetenz fünf Jahre nach dem Abitur	Sekundarstufe I Mathematik
Sachse	Karoline	Schätzung der Veränderung von Schülerleistungen in Large-Scale-Assessments	Sekundarstufe I Englisch
Schipolowski	Stefan	Crystallized intelligence as domain-specific knowledge: Structure, validity, and development	Sekundarstufe I Deutsch
Siegle	Thilo	Anwendung latenter Regressionsmodelle im Large-Scale-Assessment	Sekundarstufe I Naturwissenschaften
Wäckerle	Maike	Konstruktionen kultureller Differenz und Gemeinsamkeit. Eine rekonstruktive Studie zu Orientierungen von Fremdsprachenlernern im Fach Französisch der gymnasialen Oberstufe	Sekundarstufe II Französisch
Weirich	Sebastian	Situative Effekte in Large-Scale-Assessments	Primarstufe Deutsch

Aktuelle Dissertationsvorhaben von IQB-Mitarbeiterinnen und IQB-Mitarbeitern im Rahmen von Drittmittelprojekten

Name	Vorname	Thema der Dissertation (Arbeitstitel)	Arbeitsbereich am IQB
Darsow	Annkathrin	Der Gebrauch sprachlicher Mittel des Vergleichens von Kindern nichtdeutscher Herkunftssprache in der Primarstufe	BMBF-Projekt: Bedeutung und Form (BeFo)
Edele	Aileen	Die Rolle herkunftssprachlicher Kompetenz und kultureller Identität für den Bildungserfolg von Heranwachsenden aus zugewanderten Familien	DFG-Schwerpunktprogramm "Education as a Lifelong Process"
Gentrup	Sarah	Die Qualität der Lehrkraft-Kind-Beziehung, unterrichtliches Handeln und die Entwicklung schulrelevanter Kompetenzen in der ersten Klasse	BMBF-Projekt: Kompetenzerwerb und Lernvoraussetzungen (KuL)
Heppt	Birgit	Verständnis von Bildungssprache bei Kindern deutscher und nichtdeutscher Herkunftssprache	BMBF-Projekt: Bildungssprachliche Kompetenzen (BiSpra) – Anforderungen, Sprachverarbeitung und Diagnostik
Marx	Alexandra	Hörverstehen als Determinante von Lesekompetenz bei Jugendlichen deutscher und nichtdeutscher Herkunftssprache	DELKO-Projekt
Paetsch	Jennifer	Der Einfluss sprachlicher Kompetenzen auf das Lernen von Mathematik bei Kindern nichtdeutscher Herkunftssprache	BMBF-Projekt: Bedeutung und Form (BeFo)
Schliesing	Anna	Entwicklung und Evaluation eines elektronischen Rückmeldeformats für Daten aus Vergleichsarbeiten	BMBF-Projekt: Schulen als Steuerungsakteure im Bildungssystem (StaBil)
Schöppe	Doreen	Early Literacy und Determinanten der Lesekompetenz bei mehrsprachigen Kindern im Elementar- und Primarbereich	BMBF-Projekt (abgeschlossen): Kurz- und langfristige Effekte eines Trainings zur phonologischen Bewusstheit
Schotte	Kristin	Wahrgenommene Diskriminierung und Integration von Heranwachsenden mit Zuwanderungshintergrund	Nationales Bildungspanel (NEPS)
Taraszow	Tatjana	Sprachliche Kompetenzen bei Kindern und Jugendlichen mit Zuwanderungshintergrund	Nationales Bildungspanel (NEPS)
Wolf	Katrin	Sprachförderung und Diagnostik von Sprachförderbedarf im Elementarbereich	EkoS-Projekt Brandenburg: Evaluation kompensatorischer Sprachförderung (Projekt abgeschlossen)
Wurster	Sebastian	Effekte des evaluativen Potenzials von Schulen auf Schülerleistungen	BMBF-Projekt: Schulen als Steuerungsakteure im Bildungssystem (StaBil)

Wissenschaftliche Fachtagungen (Auswahl) mit Beiträgen von IQB-Mitarbeiterinnen und IQB-Mitarbeitern

Nationale Fachtagungen
Arbeitsgruppe für Empirische Pädagogische Forschung (AEPF) der DGfE, 2007, 2008, 2009, 2010, 2011, 2012, 2013
Bund-Länder-Konferenz „Der Deutsche Qualifikationsrahmen für Lebenslanges Lernen – Erwartungen und Herausforderungen“, 2008
Deutsche Gesellschaft für Erziehungswissenschaft (DGfE), 2008, 2009, 2010, 2012
Deutsche Gesellschaft für Lesen und Schreiben (DGLS), 2013
Deutsche Gesellschaft für Psychologie (DGPs), 2010, 2012
Deutsche Gesellschaft für Sprach- und Stimmheilkunde (DGSS), 2011
DFG Summer Academy 2008 on Educational Measurement, 2008
Deutsche Gesellschaft für Fremdsprachenforschung (DGFF), 2007, 2009, 2011, 2012
Fachgruppentagung des Arbeitskreises Vergleichsarbeiten der GDM, 2007, 2008
Fachgruppentagung Didaktik der Biologie (FDdB), 2009
Fachgruppentagung Differentielle Psychologie, Persönlichkeitspsychologie u. Psychologische Diagnostik der DGPs, 2007, 2009, 2011, 2013
Fachgruppentagung Methoden und Evaluation der DGPs, 2007, 2009, 2011, 2013
Fachgruppentagung Pädagogische Psychologie der DGPs, 2007, 2009, 2011, 2013
Fachtagung Begabungsförderung in Gymnasium und Hochschule, 2007
Gesellschaft für Didaktik der Chemie und Physik (GDGP), 2010, 2011, 2013
Gesellschaft für Didaktik der Mathematik (GDM), 2009
Gesellschaft für Empirische Bildungsforschung (GEBF), 2013
Gemeinsame Jahrestagung der Deutschen Mathematiker-Vereinigung (DMV) und der GDM, 2010
Zentrum für empirische Unterrichts- und Schulforschung (ZeUS), 2009
Kommission für Bildungsorganisation, Bildungsplanung, Bildungsrecht (KBBB) der DGfE, 2007, 2008
Kommission Grundschulforschung und Pädagogik der Primarstufe der DGfE, 2011
OECD/CERI-Regionalseminar der deutschsprachigen Länder, 2007
Tagung experimentell arbeitender Psychologen (TeaP), 2011
Tagung „Sprache im Fach – Sprachlichkeit und fachliches Lernen“, 2011

Wissenschaftliche Fachtagungen (Auswahl) mit Beiträgen von IQB-Mitarbeiterinnen und IQB-Mitarbeitern

Internationale Fachtagungen
American Educational Research Association (AERA), 2007, 2008, 2009, 2010, 2011, 2012, 2013
Association for Educational Assessment (AEA) Europe, 2010, 2011, 2012
Association for the Study of Higher Education (ASHE), 2007
Austrian Psychological Association, 2008
Conference on Information Systems (ICIS), 2012
European Association for Institutional Research (EAIR), 2008, 2010
European Association of Language Testing and Assessment (EALTA), 2007
European Association for Research on Learning and Instruction (EARLI), 2007, 2008, 2009, 2011, 2013
European Conference of Educational Research (ECER), 2009, 2010, 2011
European Conference of Psychological Assessment (EAPA), 2009
European Congress of Psychology (ECP), 2011
European Congress of Methodology of the European Association of Methodology (EAM), 2010
European Congress of Methodology (SMABS), 2009, 2010
International Amsterdam Conference on Multilevel Analysis, 2007
International Conference ALTE Project, 2008, 2009
International Conference on Motivation (ICM), 2012
International Conference on Multilingual Individuals and Multilingual Societies (MIMS), 2013
International Congress of Psychology (ICP), 2008
International Federation of Data Organizations for the Social Science (IFDO), 2009
International Max-Planck-Research School LIFE, 2007, 2008, 2011, 2012
International Meeting of the European Researchers in Didactics of Biology (ERIDOB), 2008, 2010
International Society for the Study of Individual Differences (ISSID), 2009, 2011
International Test Commission (ITC), 2012
Junior Researchers of EARLI Conference (JURE), 2010, 2011
Language and Reading Comprehension for Immigrant Children Conference (LARCIC), 2009
National Association for Research in Science Teaching (NARST), 2009, 2010
National Council on Measurement in Education (NCME), 2007, 2009
Social Science Information Service and Technology (IASSIST), 2009

Gastvorträge am IQB (Auswahl)

Dr. Patrick Kyllonen (Educational Testing Service, Princeton)
„Understanding Achievement“
02. 07. 2010

Dr. Richard Roberts (Educational Testing Service, Princeton)
„Opportunities and Challenges in Examining Background Variables in Large-Scale Survey Assessments“
02. 07. 2010

Prof. Dr. Hartmut Ditton (LMU München)
„Mehrebenenmodell im HLM“
21. 09. 2010

Hartmut Angermüller (HU Berlin)
„Prinzipien, Ratschläge und Ideen zur wirkungsvollen Gestaltung wissenschaftlicher Texte und Übungen“
25. 10. 2011

Prof. Dr. Frank Lipowsky (Universität Kassel)
„Befunde und Desiderate in der Forschung zur Wirksamkeit von Lehrerfortbildungen“
12. 07. 2012

Prof. Dr. Helen Timperley (The University of Auckland, New Zealand)
„Realizing the power of professional learning“
28. 09. 2012

Dr. Marc Kleinknecht (TU München)
„Aufgaben und Unterricht verbessern. Chancen und Grenzen einer allgemeindidaktischen Analyse von Lern- und Leistungsaufgaben“
29. 11. 2012

Prof. Dr. Sascha Schröder (MPIB)
„Cross the border – close the gap: Verbindungsmöglichkeiten von produkt- und prozessbasierter Lesekompetenzzerfassung“
17. 01. 2013

Lena Nusser und Markus Messingschlager (NEPS)
„Kompetenzmessung und Befragung von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf im Rahmen des NEPS“
31. 01. 2013

Prof. Dr. Denis Gerstorf (HU Berlin)
„Differentielle Entwicklung über die Lebensspanne: Potentiale und Herausforderungen“
12. 02. 2013

Prof. Dr. Martin Brunner (ISQ)
„Das Genetics Lab – Ein neues Computerinstrument für die Messung der Kompetenz zum komplexen Problemlösen“
30. 05. 2013

Prof. Dr. Timo von Oertzen (University of Virginia)
„Strukturgleichungsmodelle als Generalisierung aller normalen Modelle mit Onyx“
20. 06. 2013

Prof. Dr. Felicitas Thiel (FU Berlin)
„Expertise im Klassenmanagement. Ergebnisse einer Experten-Novizen Studie zu Übergängen“
04. 12. 2013

Prof. Dr. Yvonne Anders (FU Berlin)
„Empirische Befunde aus dem Modellprojekt KIDZ“
11. 12. 2013

Ausgerichtete Tagungen

6. Fachtagung

Netzwerk „Empiriegestützte Schulentwicklung“ (EMSE)

Berlin, 07./08. Juni 2007, ca. 80 TeilnehmerInnen

11. Tagung

Fachgruppe Pädagogische Psychologie der DGPs

Berlin, 17.–19. September 2007, ca. 230 TeilnehmerInnen

4. Fachtagung

Biennial EARLI/Northumbria Assessment Conference

Potsdam, 27.–29. August 2008, ca. 200 TeilnehmerInnen

Workshop

Working with Large Scale Assessment Data in Kooperation mit der International Organization for the Evaluation of Educational Achievement (IEA)

Berlin, 06.–09. Oktober 2008, ca. 20 TeilnehmerInnen

FDZ-Winterakademie

Einführung in die quantitativen Methoden der empirischen Bildungsforschung

Berlin, 08.–12. März 2010, ca. 40 TeilnehmerInnen

Tagung im Auftrag der Steuergruppe VERA der Länder

VERA 8 Mathematik: Weiterentwicklung des Konzepts – Forschung, Entwicklung und Begleitung

Berlin, 08. Juni 2010, ca. 55 TeilnehmerInnen

4. Fachtagung

European Congress of Methodology Gemeinsam mit der 27th Biennial Conference of the Society for Multivariate Analysis in the Behavioral Sciences

Potsdam, 21.–23. Juli 2010, ca. 170 TeilnehmerInnen

FDZ-Herbstakademie

Methoden der empirischen Bildungsforschung

Berlin, 20.–24. September 2010, ca. 130 TeilnehmerInnen

FDZ-Frühjahrsakademie

Methoden der empirischen Bildungsforschung

Berlin, 22.–25. März 2011, ca. 150 TeilnehmerInnen

FDZ-Herbstakademie

Methoden der empirischen Bildungsforschung

Berlin, 26.–29. September 2011, ca. 160 TeilnehmerInnen

Jahreskolloquium

DFG-Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“

Berlin, 06.–07. Oktober 2011, ca. 80 TeilnehmerInnen

FDZ-Frühjahrsakademie

Methoden der empirischen Bildungsforschung

Berlin, 27.–30. März 2012, ca. 160 TeilnehmerInnen

FDZ-Herbstakademie

Methoden der empirischen Bildungsforschung

Berlin, 04.–07. September 2012, ca. 110 TeilnehmerInnen

13. Fachtagung

Association for Educational Assessment (AEA)-Europe: Assessment Across the Lifespan

Berlin, 08.–10. November 2012, ca. 200 TeilnehmerInnen

FDZ-Frühjahrsakademie

Methoden der empirischen Bildungsforschung

Berlin, 05.–08. März 2013, ca. 160 TeilnehmerInnen

FDZ-Herbstakademie

Methoden der empirischen Bildungsforschung

Berlin, 10.–13. September 2013, ca. 140 TeilnehmerInnen

Aufgabenentwicklertagungen

Das IQB koordiniert außerdem in jedem Arbeitsbereich (Deutsch, Mathematik, Englisch, Französisch, Naturwissenschaften im Primarbereich bzw. in der Sekundarstufe I und/oder II) jährlich zwei bis acht zweitägige Aufgabenentwicklertagungen, an denen jeweils die abgeordneten Lehrkräfte der Länder und die fachdidaktischen Kooperationspartner des IQB teilnehmen.

Lehrveranstaltungen

Dr. Anja Felbrich (in Zusammenarbeit mit Jennifer Paetsch und Annkathrin Darsow)

Seminar: „Didaktische Konzepte für das Fach Deutsch/DaZ“

FU Berlin, SoSe 2010

Dr. Anja Felbrich

Hauptseminar: „Diagnostik, Rückmeldung und Evaluation“

FU Berlin, WiSe 2010/2011, WiSe 2011/2012

Dr. Anja Felbrich

Seminar: „Pädagogisches Handeln und Lernorte – Schule revisited“

HU Berlin, SoSe 2011, WiSe 2011/2012,

Dr. Anja Felbrich

Seminar: „Pädagogisches Handeln und Lernorte – Heterogenität“

HU Berlin, SoSe 2012, WiSe 2012/2013

Dr. Anja Felbrich

Kolloquium: „Einführung in die Pädagogische Diagnostik“

HU Berlin, SoSe 2013, WiSe 2013/14

Stefan Hartmann

Workshop: „Fachdidaktisches Forschungsseminar“

HU Berlin, WiSe 2011/2012, WiSe 2012/2013

Patricia Heitmann

Übung: „Fachbezogene Vermittlungskompetenzen/ Fachdidaktik“

HU Berlin, SoSe 2009, WiSe 2009/2010, SoSe 2010, WiSe 2010/2011

Patricia Heitmann

Seminar: „Innovative Konzepte der Fachdidaktik“

HU Berlin, SoSe 2011

Birgit Heppt

Hauptseminar: „Diagnostik, Rückmeldung und Evaluation“

FU Berlin, WiSe 2012/2013, SoSe 2013

Dr. Sebastian Kempert

Kolloquium: „Diagnostik, Rückmeldung und Evaluation“

HU Berlin, SoSe 2013, WiSe 2013/2014

Prof. Dr. Olaf Köller

Vorlesung: „Determinanten von Schulleistungen“

HU Berlin, SoSe 2007

Prof. Dr. Olaf Köller

Vorlesung: „Einführung in die international vergleichende empirische Schulleistungsforschung“

HU Berlin, WiSe 2006/2007, WiSe 2007/2008, WiSe 2008/2009

Prof. Dr. Olaf Köller

Vorlesung: „Einführung in die empirische Erziehungswissenschaft“

HU Berlin, SoSe 2008, SoSe 2009

Anna Kroth & Dr. Simon P. Tiffin-Richards

Seminar und Übung: „Statistik für Erziehungswissenschaften I“

HU Berlin, SoSe 2012

Dr. Poldi Kuhl

Seminar: „Schulmanagement und Qualitätsentwicklung“

Christian-Albrechts-Universität zu Kiel, WiSe 2011/2012, SoSe 2012, WiSe 2012/2013

Dr. Poldi Kuhl

Kolloquium: „Einführung in die Pädagogische Diagnostik“

HU Berlin, SoSe 2012

Dr. Poldi Kuhl

Seminar und Übung: „Präsentation und Veröffentlichung von Forschungsergebnissen“

Freie Universität, WiSe 2012/2013

Alexandra Marx

Kolloquium: „Einführung in die Pädagogische Diagnostik“

HU Berlin, WiSe 2010/2011, SoSe 2011, WiSe 2012/2013, SoSe 2013, WiSe 2013/2014

Raphaella Oehler

Seminar: „Diagnose und Förderung“

TU Berlin, WiSe 2008/2009, SoSe 2009

Prof. Dr. Hans Anand Pant

Forschungsseminar: „Testtheorie – Grundlagen empirischer Schulleistungsstudien“

HU Berlin, WiSe 2010/2011, WiSe 2011/2012, WiSe 2012/2013, WiSe 2013/2014

Prof. Dr. Hans Anand Pant

Forschungsseminar: „Lineare Strukturgleichungsmodelle“

HU Berlin, SoSe 2011

Prof. Dr. Hans Anand Pant

Forschungsseminar: „Kausale Modellierung in der Schulleistungsforschung“

HU Berlin, SoSe 2012

Dr. Dirk Richter

Vorlesung: „Forschungsmethodologie und -methoden“

FU, WiSe 2010/2011

Dr. Dirk Richter

Seminar: „Schulmanagement und Qualitätsentwicklung“

Christian-Albrechts-Universität zu Kiel, WiSe 2011/2012,

SoSe 2012, WiSe 2012/2013

Dr. Dirk Richter

Kolloquium: „Einführung in die Pädagogische Diagnostik“

HU Berlin, WiSe 2011/2012

Dr. Dirk Richter

Übung zur Vorlesung: „Methodenlehre II“

HU Berlin, SoSe 2012

Dr. Dirk Richter

Seminar und Übung: „Präsentation und Veröffentlichung von Forschungsergebnissen“

FU Berlin, WiSe 2012/2013

Alexander Robitzsch

Seminar: „Methoden der Mehrebenenanalyse“

HU Berlin, WiSe 2008/2009

Alexander Robitzsch

Seminar: „Einführung in grundlegende Konzepte der Bayes-Statistik mit R und WinBUGS“

FU Berlin, WiSe 2007/2008

Dr. Ulrich Schroeders

Kolloquium: „Einführung in die Pädagogische Diagnostik“

HU Berlin, SoSe 2009

Dr. Ulrich Schroeders

Übung: „Forschungsmethoden II – Analysen mit latenten Variablen“

HU Berlin, SoSe 2012

Dr. Ulrich Schroeders

Seminar: „Trends in der psychologischen Methodenlehre. Strukturgleichungsmodelle“

HU Berlin, SoSe 2013

Doreen Schöppe

Seminar: „Pädagogisches Handeln und Lernorte – Heterogenität und Sprachkompetenz“

HU Berlin, WiSe 2012/2013, SoSe 2013

Doreen Schöppe

Seminar: „Pädagogisches Handeln und Lernorte – Heterogenität im schulischen Kontext“

HU Berlin, WiSe 2013

Prof. Dr. Petra Stanat

Vorlesung: „Einführung in die Pädagogische Diagnostik“

HU Berlin, SoSe 2010, WiSe 2010/2011, SoSe 2011,

WiSe 2011/2012, SoSe 2012, WiSe 2012/2013

Prof. Dr. Petra Stanat

Vorlesung: „Diagnostik, Rückmeldung und Evaluation“

HU Berlin, SoSe 2013, WiSe 2013/2014

Dr. Bernd Tesch

Einführungskurs: „Grundlagen der Fachdidaktik“

HU Berlin, WiSe 2008/2009

Katrin Wolf

Kolloquium: „Einführung in die Pädagogische Diagnostik“

HU Berlin, SoSe 2011, WiSe 2011/2012, SoSe 2012,

WiSe 2012/2013



IMPRESSUM

HERAUSGEBER

Institut zur Qualitätsentwicklung im Bildungswesen
an der Humboldt-Universität zu Berlin

SITZ

Luisenstr. 56
10117 Berlin
Tel: +49 (30) 2093-46500
Fax: +49 (30) 2093-5336
E-Mail: iqboffice@iqb.hu-berlin.de
www.iqb.hu-berlin.de

POSTANSCHRIFT

Institut zur Qualitätsentwicklung im Bildungswesen
an der Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin

FOTOGRAFIE (FARBE) UND GESTALTUNG

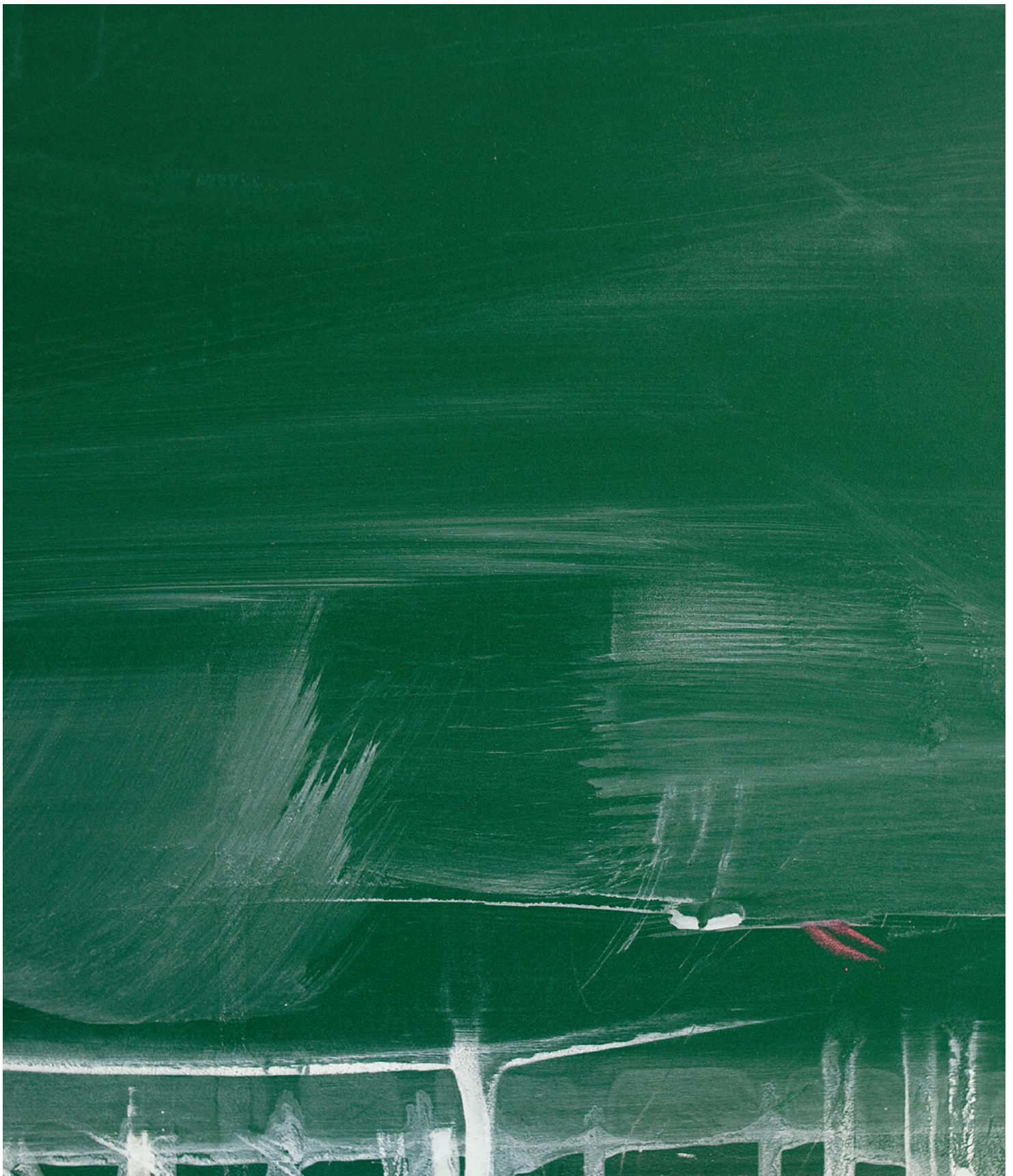
Michael Miethe
www.michaelmiethe.de

FOTOGRAFIE (SCHWARZ-WEISS)

Ralf Schultzik (IQB)

DRUCK

Trigger Medien, Berlin



HUMBOLDT-UNIVERSITÄT ZU BERLIN

