



Wilfried Bos
Heike Wendt
Olaf Köller
Christoph Selter
(Hrsg.)

TIMSS 2011

Mathematische und
naturwissenschaftliche
Kompetenzen von
Grundschulkindern
in Deutschland im
internationalen
Vergleich

WAXMANN

TIMSS 2011

Mathematische und naturwissenschaftliche
Kompetenzen von Grundschulkindern in Deutschland
im internationalen Vergleich

Wilfried Bos, Heike Wendt,
Olaf Köller, Christoph Selter (Hrsg.)

TIMSS 2011

Mathematische und naturwissenschaftliche
Kompetenzen von Grundschulkindern
in Deutschland im internationalen Vergleich



Waxmann 2012
Münster/New York/München/Berlin

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-8309-2814-0

© 2012, Waxmann Verlag GmbH,
Postfach 8603, D-48046 Münster

Waxmann Publishing Co.,
P. O. Box 1318, New York, NY 10028, USA.

www.waxmann.com
info@waxmann.com

Umschlaggestaltung: Christian Averbek, Münster
Titelfoto: © Szasz-Fabian Ilka Erika – Shutterstock
Die Motive vor den Kapiteln nach Grafiken von:
© aeroking – Fotolia.com

Satz: Stoddart Satz- und Layoutservice, Münster
Druck: Hubert & Co., Göttingen

Gedruckt auf alterungsbeständigem Papier,
säurefrei gemäß ISO 9706



Printed in Germany

Alle Rechte vorbehalten. Nachdruck, auch auszugsweise, verboten.
Kein Teil dieses Werkes darf ohne schriftliche Genehmigung des
Verlages in irgendeiner Form reproduziert oder unter Verwendung
elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Inhalt

Vorwort	11
----------------------	-----------

Kapitel I

TIMSS 2011: Wichtige Ergebnisse im Überblick	13
---	-----------

Heike Wendt, Wilfried Bos, Christoph Selter und Olaf Köller

Mathematische Kompetenzen im internationalen Vergleich	14
Naturwissenschaftliche Kompetenzen im internationalen Vergleich	17
Merkmale der Lehr- und Lernbedingungen im Primarbereich	20
Geschlechtsspezifische Unterschiede in mathematischen und naturwissenschaftlichen Kompetenzen	21
Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften	22
Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund	23
Leistungsprofile von Viertklässlerinnen und Viertklässlern in Deutschland	24
Literatur	26

Kapitel II

Ziele, Anlage und Durchführung der Trends in International Mathematics and Science Study (TIMSS 2011)	27
--	-----------

Heike Wendt, Irmela Tarelli, Wilfried Bos, Kristina Frey und Mario Vennemann

1 Zielsetzung	27
1.1 Erkenntnismöglichkeiten und zentrale Fragestellungen	28
2 TIMSS 2011 – ein kooperatives Unternehmen	30
3 Zur Teilnahme der Bundesrepublik Deutschland an international- vergleichenden Schulleistungsuntersuchungen in den Domänen Mathematik und Naturwissenschaften am Ende der vierten Jahrgangsstufe	32
4 Anlage und Durchführung von TIMSS 2011	33
4.1 Die Rahmenkonzeption der Studie	33
4.2 Zu den Teilnehmern der Studie	37
4.3 Verfahren und Kriterien der Stichprobenziehung	41
5 Entwicklung und Charakteristika der Instrumente	48
5.1 Leistungstests	48
5.2 Kontextfragebögen	50
6 Erhebung	53
6.1 Gemeinsame Erhebung mit IGLU/PIRLS 2011	53
6.2 Aufbau der Untersuchung	53
6.3 Durchführung der Erhebung	55
6.4 Qualitätssicherung	55
7 Stichprobe und Beteiligungsquoten in Deutschland	56
8 Aufbereitung und Analyse der Daten	56
8.1 Kodierung der Leistungstests	56
9 Skalierung der Leistungstests	57
9.1 Skalierungsmodelle	58
10 Gewichtung und Schätzung von Stichproben- und Messfehlern	59
11 Zur Darstellung und Interpretation der Ergebnisse	60
Literatur	66

Kapitel III

Mathematische Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse 69

Christoph Selter, Gerd Walther, Jan Wessel und Heike Wendt

1	Einleitung.....	69
2	Mathematische Grundbildung in Deutschland.....	75
3	TIMSS-Rahmenkonzeption zur Erfassung mathematischer Kompetenzen.....	78
3.1	Mathematische Inhaltsbereiche.....	78
3.2	Kognitive Anforderungsbereiche.....	80
3.3	Aufgabenbeispiele zu Inhaltsbereichen und kognitiven Anforderungsbereichen.....	80
4	Der TIMSS-Test zur Erfassung der mathematischen Kompetenzen.....	82
4.1	Testaufbau.....	82
4.2	Curriculare Validität.....	83
4.3	Kompetenzstufen in Mathematik.....	88
5	Ergebnisse.....	93
5.1	Kompetenzen im internationalen Vergleich.....	94
5.2	Kompetenzstufen.....	102
5.3	Inhaltsbereiche.....	105
5.4	Kognitive Anforderungen.....	109
5.5	Einstellungen und Selbstkonzept.....	112
6	Zusammenfassung.....	117
	Literatur.....	120

Kapitel IV

Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse 123

Thilo Kleickmann, Theresia Brehl, Steffani Saß, Manfred Prenzel
und Olaf Köller

1	Einleitung.....	123
2	Naturwissenschaftliche Grundbildung in Deutschland.....	126
3	TIMSS-Rahmenkonzeption zur Erfassung naturwissenschaftlicher Kompetenzen.....	128
3.1	Naturwissenschaftliche Inhaltsbereiche.....	128
3.2	Kognitive Anforderungsbereiche.....	129
3.3	Aufgabenbeispiele zu Inhaltsbereichen und kognitiven Anforderungsbereichen.....	130
4	Der TIMSS-Test zur Erfassung naturwissenschaftlicher Kompetenzen.....	132
4.1	Testaufbau.....	132
4.2	Curriculare Validität.....	134
4.3	Kompetenzstufen in den Naturwissenschaften.....	137
5	Ergebnisse.....	142
5.1	Kompetenzen im internationalen Vergleich.....	143
5.2	Kompetenzstufen.....	150
5.3	Inhaltsbereiche.....	154
5.4	Kognitive Anforderungsbereiche.....	157
5.5	Einstellungen und Selbstkonzept.....	160
6	Zusammenfassung.....	165
	Literatur.....	167

Kapitel V

Merkmale der Lehr- und Lernbedingungen im Primarbereich.....171

Kerstin Drossel, Heike Wendt, Silvia Schmitz und Birgit Eickelmann

1	Einleitung.....	171
1.1	Grundschule in Deutschland.....	172
2	Bildungspolitische Merkmale und äußere schulische Rahmenbedingungen.....	173
2.1	Bildungsausgaben.....	173
2.2	Klassengröße.....	173
2.3	Unterrichtszeit.....	174
2.4	Einschulungsalter.....	174
2.5	Ausbaustand und außerschulische Angebote der Ganztagschulen in Deutschland.....	174
3	Schulinterne Merkmale.....	178
3.1	Ausstattungsmerkmale.....	178
3.2	Probleme mit der Ausstattung.....	180
4	Unterrichtsmerkmale.....	183
4.1	Methodische Vorgehensweisen der Lehrkräfte im Unterricht.....	184
4.2	Computernutzung im Mathematik- und naturwissenschaftlichen Sachunterricht.....	185
4.3	Computernutzung von Lehrkräften.....	186
5	Klassenmerkmale.....	188
6	Personelle Merkmale: Lehrkräfte und Schulleitungen.....	193
6.1	Geschlechterverhältnis und Alter der Lehrkräfte.....	193
6.2	Ausbildung der Lehrkräfte.....	194
6.3	Fortbildung der Lehrkräfte und Schulleitungen.....	195
7	Zusammenfassung.....	198
	Literatur.....	200

Kapitel VI

Geschlechtsspezifische Unterschiede in mathematischen und naturwissenschaftlichen Kompetenzen203

Theresia Brehl, Heike Wendt und Wilfried Bos

1	Einleitung.....	203
2	Bisherige Forschungsergebnisse.....	203
3	Erklärungsansätze zu Geschlechterdisparitäten im fachspezifischen Kompetenzerwerb.....	205
4	Ergebnisse zu geschlechtsspezifischen Leistungsunterschieden in Mathematik.....	208
5	Ergebnisse zu geschlechtsspezifischen Leistungsunterschieden in den Naturwissenschaften.....	214
6	Ergebnisse zu geschlechtsspezifischen Unterschieden in Einstellungen und Selbstkonzept.....	220
7	Geschlechtsspezifische Leistungsunterschiede nach Kontrolle weiterer Variablen.....	224
8	Zusammenfassung.....	226
	Literatur.....	228

Kapitel VII

Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften231

Tobias C. Stubbe, Irmela Tarelli und Heike Wendt

1	Einleitung.....	231
2	Einkommensarmut.....	234
3	Der Zusammenhang zwischen sozialer Herkunft und Leistungen in Mathematik und Naturwissenschaften im internationalen Vergleich.....	235
4	Der Zusammenhang zwischen sozialer Herkunft und Mathematik-beziehungswise Naturwissenschaftskompetenz in Deutschland.....	240
5	Zusammenfassung.....	244
	Literatur.....	244

Kapitel VIII

Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund247

Irmela Tarelli, Knut Schwippert und Tobias C. Stubbe

1	Einführung.....	247
2	Mathematische und naturwissenschaftliche Kompetenzen der Schülerinnen und Schüler nach familiärem Sprachgebrauch im internationalen Vergleich.....	249
3	Deskriptive Befunde zur Lage von Schülerinnen und Schülern mit Migrationshintergrund in Deutschland.....	256
4	Kompetenzunterschiede in Deutschland.....	258
5	Zusammenhänge zwischen soziokulturellen Bedingungen und Kompetenzen.....	262
6	Schlussfolgerungen.....	265
	Literatur.....	266

Kapitel IX

Leistungsprofile von Viertklässlerinnen und Viertklässlern in Deutschland269

Wilfried Bos, Heike Wendt, Ali Ünlü, Renate Valtin, Benjamin Euen, Daniel Kasper und Irmela Tarelli

1	Einleitung.....	269
2	Domänenübergreifende Betrachtung der Verteilung auf die Kompetenzstufen.....	271
3	Multidimensionale Item-Response-Modellierung und latente Profilanalyse.....	276
3.1	Simultane Skalierung der Daten von IGLU 2011 und TIMSS 2011.....	276
3.2	Ergebnisse der latenten Profilanalyse.....	281
4	Beschreibung der Leistungstypen.....	284
4.1	Beschreibung der Leistungstypen nach der durchschnittlichen Leistung in den drei Kompetenzdomänen und ihren inhaltsbezogenen Subdomänen.....	285
4.2	Beschreibung der Leistungstypen nach relevanten Hintergrundmerkmalen.....	288
4.3	Beschreibung der Leistungstypen nach fachbezogenen Einstellungen und Selbstkonzepten.....	292
5	Zusammenfassung und Diskussion.....	293
	Literatur.....	299

Anhang A	303
-----------------------	------------

Anhang B

Beschreibung der internationalen Benchmarks für das Mathematik-Kompetenzmodell in TIMSS 2011	306
---	------------

Benchmark 1 – Kompetenzstufe II.....	307
Benchmark 2 – Kompetenzstufe III	308
Benchmark 3 – Kompetenzstufe IV	309
Benchmark 4 – Kompetenzstufe V	309

Anhang C

Beschreibung der internationalen Benchmarks für das Naturwissenschafts-Kompetenzmodell in TIMSS 2011	310
---	------------

Benchmark 1 – Kompetenzstufe II.....	307
Benchmark 2 – Kompetenzstufe III	308
Benchmark 3 – Kompetenzstufe IV	309
Benchmark 4 – Kompetenzstufe V	309

Abbildungsverzeichnis	316
------------------------------------	------------

Tabellenverzeichnis	320
----------------------------------	------------

Vorwort

Für eine nachhaltige Sicherung und Steigerung der Qualität von Unterricht und Schule bedarf es zuverlässiger empirischer Erkenntnisse, die zentrale Informationen und Ansatzpunkte für eine sachkompetente Unterstützung und Optimierung der schulischen Arbeit bieten. Im Prozess zunehmender Europäisierung und Globalisierung erweist sich eine auf den internationalen Vergleich gestützte nationale Standortbestimmung als – vielleicht zwingende – Notwendigkeit. Die regelmäßige Teilnahme an international-vergleichenden Schulleistungsuntersuchungen bietet durch die Erweiterung der nationalen Perspektive um einen globalen Referenzrahmen einzigartige Möglichkeiten des langfristigen Bildungsmonitorings. Hierzu zählt insbesondere die Schaffung von Diskussionsgrundlagen für die Generierung von wertvollem Steuerungswissen zur Bildungsplanung. Eine wiederholte Überprüfung der Leistungsfähigkeit des eigenen Bildungssystems kann somit wichtige Erkenntnisse darüber liefern, inwiefern sich eingeleitete Reformen und Veränderungen im Schulwesen als zielführend und wirksam erweisen.

Die *Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland* (KMK) hat die Bedeutung und Einzigartigkeit internationaler Vergleichsstudien erkannt. Folgerichtig bildet die Teilnahme an internationalen Schulleistungsstudien einen zentralen Eckpfeiler der im Jahr 2006 beschlossenen *Gesamtstrategie zum Bildungsmonitoring*. Es zeigt sich zunehmend, dass mit der Teilnahme der Bundesrepublik Deutschland an international-vergleichenden Schulleistungsuntersuchungen wertvolle Informationen über die Kompetenzen von Schülerinnen und Schülern in zentralen Leistungsbereichen an entscheidenden Schnittstellen des Schulsystems generiert werden können. Die so gewonnenen Erkenntnisse liefern nicht nur bedeutsame Beiträge zur empirischen Bildungs- und erziehungswissenschaftlichen Grundlagenforschung, sondern stellen auch entscheidende Diskursgrundlagen für Bildungsfragen und Reformen bereit: So wäre beispielsweise ohne die Grundlagenforschung, die in den international-vergleichenden Schulleistungsstudien geleistet wurde, die weiterführende Forschung zu Bereichen wie den Bildungsstandards, den Vergleichsarbeiten, zu neuen Kerncurricula et cetera wohl nicht möglich gewesen.

In diesem Band werden die Ergebnisse der *Trends in International Mathematics and Science Study* (TIMSS) vorgestellt. TIMSS ist eine international-vergleichende Schulleistungsstudie, die seit 1995 alle vier Jahre durchgeführt wird. Im Zentrum von TIMSS 2011 steht der internationale Vergleich von Schülerleistungen am Ende der vierten und achten Jahrgangsstufe sowie am Ende der Sekundarstufe II im mathematischen und naturwissenschaftlichen Bereich unter Berücksichtigung curricularer Vorgaben und anderer zentraler Rahmenbedingungen schulischer Lernumgebungen. Wie bereits im Jahr 2007 hat sich Deutschland auch im Jahr 2011 ausschließlich mit der vierten Jahrgangsstufe an der Studie beteiligt, weshalb in diesem Band entsprechend ausschließlich Ergebnisse der Grundschulstudie von TIMSS vorgestellt werden. Neben der differenzierten Betrachtung der Schülerleistungen im internationalen Vergleich und im Vergleich zur Erhebungsrunde 2007 werden zentrale Merkmale der Lehr- und Lerngelegenheiten, Geschlechterdisparitäten sowie Leistungsunterschiede von Schülerinnen und Schülern unterschiedlicher sozialer und kultureller Herkunft fo-

kussiert und wesentlich ausführlicher besprochen als im parallel erscheinenden Ergebnisbericht der internationalen Studienleitung.

Eine zentrale Besonderheit dieser Erhebungsrunde ist, dass es im Jahr 2011 möglich war, TIMSS gemeinsam mit der Schwesterstudie *Progress in International Reading Literacy Study* (PIRLS)/*Internationale Grundschul-Lese-Untersuchung* (IGLU) durchzuführen und eine gemeinsame Untersuchungsstichprobe zu realisieren. In Deutschland wurde diese Gelegenheit genutzt, um so einen umfassenden Einblick in die Kompetenzniveaus von Viertklässlerinnen und Viertklässlern in den Bereichen Lesen, Mathematik und Naturwissenschaften zu erhalten. Ersten Ergebnissen einer die Kompetenzdomänen übergreifenden Betrachtung von Schülerleistungen am Ende der Grundschulzeit ist das letzte Kapitel in diesem Band gewidmet.

Zentrale Ergebnisse der Schwesterstudie IGLU werden in einem zeitgleich erscheinenden, gesonderten Ergebnisbericht vorgestellt. Das Anliegen war es, zwei Bände zu schaffen, die sowohl parallel als auch unabhängig voneinander gelesen werden können. Auf Grund einer großen Überschneidung der an der Erstellung dieser Berichtsbände beteiligten Teams von Wissenschaftlerinnen und Wissenschaftlern bot es sich an – insbesondere um ein paralleles Lesen der Bände zu erleichtern – einige Kapitel für beide Bände gemeinsam zu erstellen. Entsprechend waren die Autorinnen und Autoren beider Bände um eine Vereinheitlichung von Abschnitten, die beide Studien betreffen, bemüht. Insofern sind einige Passagen dieses Buches identisch gehalten oder auf die Inhalte des IGLU-Bands abgestimmt. Selbstverständlich gibt es aber auch eine Reihe von spezifischen Teilen, die jeweils nur in einem der beiden Bände enthalten sind. Dies betrifft in erster Linie die Kompetenzbeschreibungen. Darüber hinaus wurde jeweils ein Schwerpunktkapitel vorbereitet: In diesem Band ist es das Kapitel 6 zu geschlechterspezifischen Unterschieden in mathematischen und naturwissenschaftlichen Kompetenzen. Im IGLU-Band ist dem Thema Übergang von der Primar- in die Sekundarstufe ein eigenes Kapitel gewidmet worden. Ein gemeinsames Kapitel ist entstanden, in dem Kompetenzprofile von Viertklässlerinnen und Viertklässlern in Deutschland für die drei Domänen Lesen, Mathematik und Naturwissenschaften untersucht werden.

Weltweit waren an TIMSS 2011 59 Staaten und Regionen mit insgesamt mehr als 291 000 Schülerinnen und Schülern, rund 160 000 Eltern, über 13 500 Lehrerinnen und Lehrern an ungefähr 10 000 Schulen beteiligt. In Deutschland liegen Daten für rund 4 000 Schülerinnen und Schüler, deren Eltern, 300 Lehrpersonen und für knapp 200 Schulleitungen vor. All diesen Personen gilt ein besonders herzliches Dankeschön für ihre hohe Bereitschaft, die Leistungstests zu bearbeiten beziehungsweise die Fragebögen auszufüllen, und auf diese Weise jene Informationen zur Verfügung zu stellen, über die hier berichtet wird. Ohne ihre Beteiligung wäre weder die Durchführung noch die Berichterstattung möglich gewesen.

Auch den zahlreichen weiteren Personen, die zum Gelingen von TIMSS 2011 und zur Realisierung der vorliegenden Publikation beigetragen haben, sei an dieser Stelle gedankt. Nicht zuletzt gebührt den Mitarbeiterinnen und Mitarbeitern in den Ministerien der Länder und des Bundes Dank für ihre Unterstützung.

Professor Dr. Wilfried Bos

Wissenschaftlicher Leiter von TIMSS 2011 und IGLU 2011

Kapitel I

TIMSS 2011: Wichtige Ergebnisse im Überblick

Heike Wendt, Wilfried Bos, Christoph Selter und Olaf Köller

Die *Trends in International Mathematics and Science Study* (TIMSS) ist eine international-vergleichende Schulleistungsuntersuchung, die von der *International Association for the Evaluation of Educational Achievement* (IEA) verantwortet wird. Die IEA ist ein unabhängiger, internationaler Zusammenschluss von Wissenschaftlerinnen und Wissenschaftlern, Forschungseinrichtungen und Regierungsstellen und führt seit 1959 international-vergleichende Schulleistungstudien durch. TIMSS gehört seit 1995 zu den Kernstudien der IEA und wird im Abstand von vier Jahren durchgeführt. Ein zentrales Anliegen ist es, langfristige Entwicklungen in den teilnehmenden Bildungssystemen zu dokumentieren.

Mit TIMSS werden unter Berücksichtigung curricularer Vorgaben und anderer zentraler Rahmenbedingungen schulischer Lernumgebungen im internationalen Vergleich Fachleistungen von Schülerinnen und Schülern in den Domänen Mathematik und Naturwissenschaften am Ende der vierten und achten Jahrgangsstufe sowie am Ende der Sekundarstufe II (*TIMSS Advanced*) untersucht.

Mit der Erhebung im Jahr 2011 beteiligte sich Deutschland zum zweiten Mal an der Grundschuluntersuchung von TIMSS. Die Teilnahme erfolgte auf Beschluss der *Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland* (KMK) und einer Vereinbarung zwischen der KMK und dem *Bundesministerium für Bildung und Forschung* (BMBF). Die erste Teilnahme an der Untersuchung mit Schülerinnen und Schülern der vierten Jahrgangsstufe erfolgte für den Studienzyklus 2007 (vgl. Bos et al., 2008).

Insgesamt haben sich an TIMSS 2011 weltweit 59 Bildungssysteme beteiligt, davon 50 Staaten und Regionen als reguläre Teilnehmer und weitere sieben Regionen mit gesonderten Teilnahmebedingungen als sogenannte Benchmark-Teilnehmer. Zudem nutzen weitere drei Staaten die Möglichkeit, sich auch (wie im Fall von Jemen) oder ausschließlich mit Schülerinnen und Schülern der sechsten Jahrgangsstufe auf der TIMSS-Skala international zu verorten. In diesem Berichtsband werden primär die Ergebnisse von zentralen Vergleichsstaaten berichtet.

TIMSS basiert auf einem umfassenden Rahmenkonzept, das Expertinnen und Experten aus vielen Staaten erarbeitet haben. Erhoben werden einerseits Schülerleistungen im mathematischen und naturwissenschaftlichen Bereich und andererseits Merkmale von Schülerinnen und Schülern, von Lehrkräften, Unterricht und Schulen, die den Aufbau mathematischer und naturwissenschaftlicher Kompetenzen potentiell beeinflussen können.

In seiner Konzeption folgt TIMSS einem Curriculum-Modell, das drei Ebenen umfasst: Das *intendierte Curriculum* repräsentiert Inhalte und Prozesse, welche die Schülerinnen und Schüler in einem Bildungssystem lernen sollen. Das intendierte Curriculum wird in der Regel in Lehrplänen und Prüfungsvorschriften festgelegt. Das *implementierte Curriculum* stellt den tatsächlich unterrichteten Lernstoff dar. Der TIMSS-Test erfasst das *erreichte Curriculum*, also das von den Schülerinnen und Schülern Gelernte sowie ihre Einstellungen zur Mathematik und zu den Naturwissenschaften.

Das Kompetenzmodell von TIMSS 2011 differenziert fachspezifisch je drei Inhaltsbereiche und drei kognitive Anforderungsbereiche. Inhaltlich wird für die Erfassung mathematischer Kompetenzen am Ende der vierten Jahrgangsstufe zwischen den Bereichen *Arithmetik*, *Geometrie/Messen* und dem *Umgang mit Daten* unterschieden. Für die Naturwissenschaften werden die Bereiche *Biologie*, *Physik* und *Geographie* differenziert, wobei der Inhaltsbereich *Physik* auch Themengebiete umfasst, die in Deutschland traditionell der *Chemie* zugeordnet werden. Die einzelnen Inhaltsbereiche werden zudem fachbezogen in verschiedene Themengebiete aufgegliedert. Um grundlegende Denkprozesse zu beschreiben, die für eine erfolgreiche Aufgabenbearbeitung zu bewältigen sind, werden die kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen* unterschieden, denen wiederum konkrete Verhaltensweisen zugeordnet sind.

Um die Leistungsstände von Grundschülerinnen und Grundschülern am Ende der vierten Jahrgangsstufe nicht isoliert zu betrachten, sondern darüber hinaus Faktoren identifizieren zu können, die Leistungen und Einstellungen beeinflussen können, werden die institutionellen und gesellschaftlichen Rahmenbedingungen, die die Bereitstellung und Nutzung schulischer Lerngelegenheiten beeinflussen, durch die standardisierte schriftliche Befragung der teilnehmenden Schülerinnen und Schüler, der unterrichtenden Lehrkräfte und der Schulleitungen erfasst. Ein Rückgriff auf diese Rahmendaten ermöglicht vertiefende Analysen zu den Zusammenhängen zwischen den Leistungsergebnissen und Hintergrundmerkmalen, die wertvolle Informationen und Erklärungsansätze insbesondere zu sozialen und kulturellen Disparitäten der Schülerleistungen liefern.

Mathematische Kompetenzen im internationalen Vergleich

Vorrangiges Ziel von TIMSS 2011 ist der internationale Vergleich von Fachkompetenzen der Schülerinnen und Schüler der vierten Jahrgangsstufe. Im internationalen Vergleich zwischen den Teilnehmerstaaten liegt die mittlere Mathematikleistung der Schülerinnen und Schüler am Ende der Grundschulzeit in Deutschland im oberen Drittel der Rangreihe.

Mit einem *Leistungsmittelwert* von 528 Punkten liegt Deutschland gleichauf mit Litauen (534), Portugal (532) und Irland (527) und signifikant über dem internationalen Mittelwert (491) sowie über den mittleren Leistungsmittelwerten aller teilnehmenden Mitgliedsstaaten der *Europäischen Union* (EU) (519) und al-

ler teilnehmenden Mitgliedsstaaten der *Organisation for Economic Co-operation and Development* (OECD) (521).

Trotz insgesamt zufriedenstellender Resultate besteht für Deutschland noch erhebliches Entwicklungspotential. Insgesamt erzielen die Schülerinnen und Schüler aus 13 Teilnehmerstaaten bessere Leistungsmittelwerte als die Viertklässlerinnen und Viertklässler in Deutschland. Im Vergleich zu den Staaten mit den höchsten mathematischen Kompetenzniveaus schneidet Deutschland deutlich schwächer ab. Ein erheblicher Abstand besteht hier insbesondere zu einigen Teilnehmerstaaten in Asien: Singapur (606), die Republik Korea (Südkorea) (605), Hongkong (602), Taiwan (591) und Japan (585). Aber auch die Leistungen von Schülerinnen und Schülern aus nicht wenigen westeuropäischen Staaten wie Nordirland (562), der Flämischen Gemeinschaft in Belgien (549), Finnland (545), England (542), den Niederlanden (540) und Dänemark (537) heben sich signifikant von den mathematischen Leistungen der Grundschul Kinder in Deutschland ab.

Im Vergleich zu TIMSS 2007 unterscheidet sich das mittlere Niveau der mathematischen Kompetenzen in Deutschland 2011 mit 528 Punkten nur unerheblich von dem Niveau im Jahr 2007, als 525 Punkte erreicht wurden. Insgesamt zeigt sich, dass von 23 Staaten, die an beiden Untersuchungen teilnahmen, 9 Staaten signifikant besser abschneiden und dass kein Staat signifikant schlechtere Leistungen erzielt. Deutschland gehört damit zu der Gruppe von 14 Teilnehmerstaaten, in denen das Leistungsniveau von 2007 gehalten wurde. Im Vergleich zu TIMSS 2007 zeigt sich jedoch auch, dass eine signifikante Leistungssteigerung durchaus erreichbar ist, wie die Beispiele Dänemark und USA zeigen. Während die Leistungsmittelwerte der Schülerinnen und Schüler aus diesen Staaten sich bei TIMSS 2007 nicht signifikant von dem der Viertklässlerinnen und Viertklässler in Deutschland unterscheiden, ist dies 2011 der Fall.

Erfreulicherweise zeigt sich für Deutschland im internationalen Vergleich eine geringe Streuung der mathematischen Kompetenzen. Die vergleichsweise hohe ‚Leistungshomogenität‘ lässt sich sowohl an der Größe der Standardabweichung (62) als auch an der Leistungsspanne zwischen den jeweils 5 Prozent der leistungsschwächsten und den 5 Prozent der leistungsstärksten Schülerinnen und Schüler ablesen. Zunächst ist es erfreulich, dass die Schere zwischen Kindern in Deutschland mit guter und schwacher Leistung vergleichsweise gering ist und sogar leicht unter dem Niveau von TIMSS 2007 liegt. Dennoch wäre es wünschenswert, das Leistungsniveau der Grundschul Kinder in Deutschland insgesamt auf ein höheres Niveau zu heben.

Bei einer kriterialen Betrachtung nach *Kompetenzstufen* lässt sich für TIMSS 2011 feststellen, dass 81 Prozent der Schülerinnen und Schüler mindestens Stufe III erreichen. Fast ein Fünftel der Schülerinnen und Schüler zeigt Leistungen auf dem Niveau der ersten und zweiten Kompetenzstufe und verfügt damit über elementares mathematisches Wissen und elementare mathematische Fertigkeiten und Fähigkeiten, um ihnen vertraute Aufgaben zu bearbeiten. Besonderer Aufmerksamkeit bedürfen die 3 Prozent der Schülerinnen und Schüler, die sich auf der untersten Kompetenzstufe befinden und nur über rudimentäres mathematisches Wissen verfügen.

Auch am oberen Ende des Leistungsspektrums ergeben sich noch Spielräume: Lediglich 5 Prozent der Schülerinnen und Schüler zeigen Leistungen auf der höchsten Kompetenzstufe V. Damit können nur sehr wenige Schülerinnen und Schüler in Deutschland ihre mathematischen Fertigkeiten und Fähigkeiten ver-

ständig beim Lösen verhältnismäßig komplexer Probleme anwenden und ihr Vorgehen erläutern. Im internationalen Vergleich zeigt sich, dass die Mittelwerte für die teilnehmenden EU- beziehungsweise OECD-Staaten mit jeweils 9 Prozent über dem Anteil der Viertklässlerinnen und Viertklässler in Deutschland liegen, die die Kompetenzstufe V erreichen. Anteile von mehr als 30 Prozent, wie sie in den Staaten an der Spitze der Leistungsskala wie in Singapur (43%), Korea (39%) oder Hongkong (38%) erreicht werden, liegen in weiter Ferne.

Im Vergleich der Untersuchungen TIMSS 2007 und 2011 zeigt sich, dass sich für Deutschland trotz geringfügiger nomineller Verschiebungen keine signifikanten Veränderungen der Verteilung der Viertklässlerinnen und Viertklässler auf den *Kompetenzstufen* ergeben haben.

Die TIMSS-Rahmenkonzeption zur Erfassung mathematischer Kompetenzen erlaubt die differenzierte Darstellung von Ergebnissen in den drei Inhaltsbereichen *Arithmetik*, *Geometrie/Messen* und *Umgang mit Daten* sowie in den drei kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen*. Bei einer differenzierten Betrachtung nach mathematischen Inhaltsbereichen zeigen die Viertklässlerinnen und Viertklässler in Deutschland relative Stärken in den Bereichen *Geometrie/Messen* (536) und *Umgang mit Daten* (546) und relative Schwächen – verglichen mit dem Gesamtmittelwert Mathematik für Deutschland (528) – im Bereich *Arithmetik* (520).

Betrachtet man die kognitiven Anforderungsbereiche, so zeigen sich bei den Viertklässlerinnen und Viertklässlern in Deutschland relative Stärken im *Problemlösen* (532) und durchschnittliche, das heißt auf dem Niveau des deutschen Gesamtmittelwerts liegende Leistungen im Bereich *Anwenden* (528). Weniger gute Leistungen sind im Bereich *Reproduzieren* (524) zu finden. Vergleicht man die Ergebnisse von TIMSS 2007 und 2011, so lässt sich Folgendes feststellen: Signifikante positive Veränderungen ergeben sich im Bereich *Reproduzieren* mit 524 Punkten (2011) gegenüber 515 Punkten (2007). Auch in den anderen Inhaltsbereichen *Geometrie/Messen* und *Umgang mit Daten* sind signifikante Verbesserungen von 9 beziehungsweise 14 Punkten zu verzeichnen. Keine signifikanten Veränderungen gibt es im Bereich *Anwenden*, in dem ein Minus von 2 Punkten zu berichten ist, beim *Problemlösen* (ein Plus von 2 Punkten) und im Bereich *Arithmetik* (ein Minus von 4 Punkten).

In TIMSS werden nicht nur Inhalts- und kognitive Anforderungsbereiche mathematischer Kompetenzen erfasst, sondern auch motivationale Aspekte des Lernens. Diese stellen neben den Kompetenzen einen wichtigen Zielbereich des Unterrichts dar. In Bezug auf die *Einstellungen zur Mathematik* ergeben sich im Vergleich zu TIMSS 2007 keine bedeutsamen Veränderungen: Die Schülerinnen und Schüler berichten wie auch 2007 eine positive Einstellung zum Fach Mathematik. Allerdings ist der Anteil der Kinder mit einem niedrigen Wert in der positiven Einstellung signifikant um 2,6 Prozentpunkte auf knapp 16,6 Prozent gestiegen. Entsprechend ist der Anteil der Schülerinnen und Schüler, die eine hohe positive Einstellung angeben, geringer geworden, doch mit gut 67 Prozent nach wie vor hoch: Zwei Drittel der Schülerinnen und Schüler der vierten Jahrgangsstufe in Deutschland geben an, dass sie eine positive Einstellung zur Mathematik haben. In Bezug auf das *mathematikbezogene Selbstkonzept* ist im Vergleich zu TIMSS 2007 eine leicht positive Veränderung zu beobachten. Die Schülerinnen und Schüler berichten wie 2007 ein recht positives mathematikbezogenes Selbstkonzept. Der Anteil der Schülerinnen und Schüler mit einem niedrigen mathematischen Selbstkonzept bleibt mit 10 Prozent nahezu stabil. Der Anteil der Kinder mit mittlerem positiven Selbstkonzept fällt leicht auf

19 Prozent und der Anteil der Schülerinnen und Schüler, die ein hohes mathematikbezogenes Selbstkonzept angeben, steigt im Gegenzug auf knapp 71 Prozent. Mehr als 7 von 10 Schülerinnen und Schülern der vierten Jahrgangsstufe in Deutschland geben also an, dass sie über ein positives mathematikbezogenes Selbstkonzept verfügen.

Zusammenfassend bleibt Folgendes festzuhalten: Die Mathematikleistungen der Schülerinnen und Schüler in Deutschland am Ende der vierten Jahrgangsstufe sind nach wie vor recht gut, liegen im oberen Mittelfeld, aber mit deutlichem Abstand zu Staaten an der Spitze der Leistungsskala wie Singapur, Hongkong, Taiwan oder Japan. Im Vergleich zu TIMSS 2007 ist keine wesentliche Veränderung zu beobachten. Zudem sind in Bezug auf die Förderung der leistungsschwachen (Kompetenzstufen I und II) und insbesondere auch der sehr leistungsstarken Schülerinnen und Schüler (Kompetenzstufe V) kaum Veränderungen zu verzeichnen.

In der Zusammenschau der verschiedenen Fragestellungen, die untersucht wurden, fallen eine Reihe von Staaten auf, bei denen im Vergleich von TIMSS 2007 und 2011 erfreuliche Entwicklungen zu konstatieren sind. Dass sich Teilnehmerstaaten in einzelnen Bereichen signifikant verschlechtern, ist nur selten zu beobachten. Es verbleiben einige Staaten, in denen sich im Vergleich von TIMSS 2007 und 2011 häufig keine oder nur wenige bedeutsame Unterschiede feststellen lassen. Zu dieser Gruppe gehört Deutschland. Das ist nicht zufriedenstellend und markiert die Notwendigkeit, die Umsetzung der Bildungsstandards im Mathematikunterricht nicht nur durch qualitätsmessende sondern auch durch qualitätsentwickelnde Maßnahmen zu unterstützen.

Naturwissenschaftliche Kompetenzen im internationalen Vergleich

Das Niveau naturwissenschaftlicher Kompetenzen von Schülerinnen und Schülern am Ende ihrer Grundschulzeit liegt bei 528 Punkten auf der TIMSS-Skala. Deutschland liegt damit auf Rangplatz 17 von insgesamt 50 Teilnehmerstaaten. Die naturwissenschaftliche Kompetenz der Schülerinnen und Schüler liegt in Deutschland deutlich über dem internationalen Mittelwert von 486 Punkten, während der Abstand zum Durchschnitt der Vergleichsgruppe EU (521 Punkte) zwar statistisch signifikant, aber eher gering ist. Der Unterschied zum Durchschnitt der teilnehmenden OECD-Staaten ist nicht signifikant. Im Vergleich zu den Staaten mit dem höchsten naturwissenschaftlichen Kompetenzniveau schneidet Deutschland deutlich schlechter ab. Der Abstand zur Republik Korea (Südkorea) und Singapur an der Spitze der Rangreihe ist mit 59 beziehungsweise 55 Punkten beträchtlich. Das in Deutschland erzielte Ergebnis ist vergleichbar mit dem Leistungsniveau in Staaten wie Schweden, den Niederlanden, England und Italien.

Bei TIMSS 2011 erzielt Deutschland genau denselben Mittelwert in den naturwissenschaftlichen Kompetenzen wie in 2007. Es lassen sich also keine Unterschiede im Niveau der naturwissenschaftlichen Kompetenzen zwischen TIMSS 2007 und 2011 feststellen. In einer Reihe weiterer Staaten, darunter die Russische Föderation, die USA und Taiwan, finden sich ebenfalls keine Veränderungen im Kompetenzniveau. Staaten, die hingegen eine stark positive Veränderung des Niveaus der naturwissenschaftlichen Kompetenzen zeigen, sind beispielsweise Georgien, die Tschechische Republik und Iran.

In Deutschland ist die Streuung der Leistungen in den Naturwissenschaften vergleichsweise gering ausgeprägt, das heißt, die Unterschiede zwischen den naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler fallen vergleichsweise niedrig aus. Dabei ist die Streuung der Leistungswerte der Schülerinnen und Schüler in Deutschland bei TIMSS 2011 ähnlich stark ausgeprägt wie bei TIMSS 2007. Der Unterschied zwischen den 5 Prozent leistungsschwächsten und den 5 Prozent leistungsstärksten Schülerinnen und Schülern ist im Vergleich zu anderen Staaten eher gering. Ähnliche Streuungen in der naturwissenschaftlichen Kompetenz lassen sich aber auch in anderen OECD- und EU-Staaten finden. Eine Ausnahme bilden die Niederlande und auch Belgien, in denen die Kompetenzen der Schülerinnen und Schüler noch homogener ausgeprägt sind.

Anhand von fünf Kompetenzstufen können die naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler inhaltlich illustriert werden. In Deutschland befinden sich 7 Prozent der Schülerinnen und Schüler auf der höchsten Kompetenzstufe (V). Diese Schülerinnen und Schüler verstehen und begründen bereits am Ende der Grundschulzeit naturwissenschaftliche Zusammenhänge, sie interpretieren einfache naturwissenschaftliche Experimente und ziehen Schlussfolgerungen daraus. Weitere 32 Prozent der Schülerinnen und Schüler befinden sich auf Kompetenzstufe IV. Diese Schülerinnen und Schüler verfügen über naturwissenschaftliche Konzepte, mit denen sie Alltagsphänomene und Sachverhalte beschreiben können. Sie besitzen ausreichendes Basiswissen, welches eine zentrale Voraussetzung für eine erfolgreiche Teilnahme am naturwissenschaftlichen Unterricht an weiterführenden Schulen ist. Demgegenüber sind die Voraussetzungen der Gruppe von Schülerinnen und Schülern in Deutschland, die sich auf den Kompetenzstufen I und II befinden, als ungünstig zu bezeichnen. Sie können zwar elementares Faktenwissen abrufen, allerdings fehlt es ihnen an einem grundlegenden naturwissenschaftlichen Verständnis. Kritisch ist der Anteil von 4 Prozent der Schülerinnen und Schüler einzuschätzen, die am Ende ihrer Grundschulzeit nicht in der Lage sind, einfache Aufgaben zu elementarem naturwissenschaftlichen Faktenwissen sicher zu lösen (Kompetenzstufe I). Die Anteile der Schülerinnen und Schüler auf den einzelnen Kompetenzstufen liegen in Deutschland aber in etwa in der gleichen Größenordnung wie im Durchschnitt der teilnehmenden EU- und OECD-Staaten. Im Vergleich zu TIMSS 2007 zeigt sich eine geringfügige Verschiebung der Verteilung in Richtung Mitte: Während der Anteil der Kinder auf Kompetenzstufe V geringfügig kleiner geworden ist, ist der Anteil der Schülerinnen und Schüler auf Kompetenzstufe III geringfügig größer geworden.

Die TIMSS-Rahmenkonzeption zur Erfassung naturwissenschaftlicher Kompetenzen erlaubt eine differenzierte Darstellung von Ergebnissen in den drei Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geographie* sowie in den drei kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen*. Auf diese Weise können relative Stärken und Schwächen in den einzelnen Bereichen identifiziert werden. Beim Vergleich der drei Inhaltsbereiche zeigen sich in Deutschland leichte relative Stärken im Bereich *Physik/Chemie* und eine leichte relative Schwäche im Bereich *Geographie*. Hinsichtlich der kognitiven Anforderungsbereiche zeigen sich im *Reproduzieren*, im *Anwenden* und auch im *Problemlösen* ähnlich hohe Leistungsmittelwerte der Schülerinnen und Schüler in Deutschland, das heißt, die Leistungen sind ausgewogen. Der Abstand zu den Staaten mit dem höchsten Niveau naturwissenschaftlicher Kompetenzen ist allerdings in allen drei Inhaltsbereichen und kognitiven Anforderungsbereichen groß.

Im Vergleich zu TIMSS 2007 zeigen sich weder in einem der Inhaltsbereiche noch in einem der kognitiven Anforderungsbereiche signifikante Unterschiede.

In TIMSS werden nicht nur Inhalts- und kognitive Anforderungsbereiche naturwissenschaftlicher Kompetenzen erfasst, sondern auch motivationale Orientierungen. Diese stellen neben den Kompetenzen einen wichtigen Zielbereich naturwissenschaftlichen Unterrichts dar. Da in Deutschland naturwissenschaftliche Inhalte in der Grundschule im Rahmen des integrativen Fachs Sachunterricht unterrichtet werden, haben Grundschul Kinder oft noch keine Vorstellung davon, was Naturwissenschaften beziehungsweise naturwissenschaftliche Inhalte sind. Aus diesem Grund werden bei TIMSS nicht naturwissenschaftsbezogene, sondern auf den Sachunterricht bezogene Einstellungen und Selbstkonzepte der Schülerinnen und Schüler erfasst. Es zeigt sich, dass die Schülerinnen und Schüler am Ende der Grundschulzeit über insgesamt sehr positive Einstellungen zum Sachunterricht verfügen. 80 Prozent der Kinder befinden sich in der Gruppe der Schülerinnen und Schüler mit hohen positiven Einstellungen. Beim sachunterrichtsbezogenen Selbstkonzept zeigt sich ein ähnliches Bild: 80 Prozent der Schülerinnen und Schüler verfügen über ein hoch positiv ausgeprägtes Selbstkonzept.

Sowohl die Einstellungen zum Sachunterricht als auch das sachunterrichtsbezogene Selbstkonzept korrelieren schwach, aber signifikant positiv mit der naturwissenschaftlichen Kompetenz. Kinder, die sich auf den oberen Kompetenzstufen befinden, haben also eher auch positive Einstellungen und ein positives Selbstkonzept. Umgekehrt finden sich auf den unteren Kompetenzstufen auch eher Kinder mit weniger positiven Einstellungen und Selbstkonzepten. Bei dieser Gruppe von Kindern zeichnen sich möglicherweise sowohl leistungsbezogene als auch motivationale Probleme hinsichtlich der weiteren Auseinandersetzung mit Naturwissenschaften ab. Vergleicht man die Werte für Einstellungen und Selbstkonzept von 2011 mit denen von 2007, so zeigen sich bei den Einstellungen keine Unterschiede und beim Selbstkonzept geringfügig höhere Werte im Jahr 2011. Insgesamt sind die Einstellungen zum Sachunterricht und die sachunterrichtsbezogenen Selbstkonzepte der Schülerinnen und Schüler in Deutschland sehr positiv ausgeprägt, was erwarten lässt, dass die Bereitschaft, sich aktiv mit naturwissenschaftlichen Fragestellungen auch nach der Grundschulzeit auseinanderzusetzen, hoch ist.

Zusammenfassend bleibt festzuhalten, dass im Bereich der Naturwissenschaften die Kompetenzen der Schülerinnen und Schüler in Deutschland bei TIMSS 2011 auf demselben Niveau wie 2007 liegen. Deutschland befindet sich damit 2011 auf dem Niveau des Durchschnitts der übrigen teilnehmenden OECD-Staaten. Ob dieses Niveau in Deutschland weiterhin gehalten werden kann, ist unter anderem vor dem Hintergrund demographischer Entwicklungen fraglich. Zweifel sind auch deshalb angebracht, weil das Programm *SINUS an Grundschulen* zur Steigerung der Unterrichtsqualität im mathematischen und naturwissenschaftlichen Unterricht im Jahre 2013 ausläuft, ohne dass ernsthafte Bestrebungen erkennbar sind, dieses Programm fortzuführen beziehungsweise in die Fläche zu tragen. Möglichst vielen Schülerinnen und Schülern eine profunde naturwissenschaftliche Grundbildung zu vermitteln wäre aber die Grundlage dafür, auch den Herausforderungen des Nachwuchsmangels in vielen naturwissenschaftlichen und technischen Berufsfeldern zu begegnen. Im Zuge der Fokussierung bildungspolitischer Anstrengungen auf dem Bereich der Sprachförderung und der Tatsache, dass für das Fach Sachunterricht keine Bildungsstandards existieren, kommt die Sorge auf, dass der naturwissen-

schaftliche Lernbereich in der Grundschule zum Stiefkind bildungspolitischer Bemühungen wird.

Merkmale der Lehr- und Lernbedingungen im Primarbereich

Für TIMSS 2011 wurden in Deutschland erstmals ausgewählte Rahmenbedingungen, in denen das Lehren und Lernen an Grundschulen stattfindet, in den Blick genommen. Die thematischen Schwerpunkte wurden sowohl auf die Entwicklung der Ganztagsgrundschulen als auch auf die Ausstattung und Nutzung digitaler Medien im Mathematik- und Sachunterricht gelegt. Im Kapitel „Lehr- und Lernbedingungen an Grundschulen im internationalen Vergleich“ im parallel erscheinenden Ergebnisbericht zu IGLU 2011 (Bos, Tarelli, Bremerich-Vos & Schwippert, 2012) liegt der Fokus hingegen auf den äußeren schulischen Rahmenbedingungen wie Klassengröße und Unterrichtszeit, dem Schulleitungshandeln sowie spezifischen Aspekten des Leseunterrichts (Tarelli, Lankes, Drossel & Gegenfurtner, 2012). Als Vergleichsgruppe wurden die Teilnehmerstaaten der EU ausgewählt.

In Bezug auf die Entwicklung der Ganztagsgrundschulen in Deutschland zeigt sich zunächst, dass 2011 fast jedes zweite Grundschulkind in Deutschland eine Schule besucht, die einen Ganztagsbetrieb anbietet. Damit lässt sich auch anhand der TIMSS-Daten ein deutlicher Ausbau in diesem Bereich feststellen. Die Daten weisen in diesem Zusammenhang darauf hin, dass sich die teilweise gebundene Form nicht in größerer Zahl etabliert hat. Trotz der Ausbaumühnungen zeigt sich jedoch, dass auch im Jahr 2011 lediglich 13 Prozent aller Ganztagsgrundschulkind eine Schule besuchen, an der die Teilnahme am Ganztagsunterricht verpflichtend ist (voll gebundene Form). Mehr als zwei Drittel aller Ganztagsgrundschulkind sind hingegen an Schulen angemeldet, in der die Teilnahme am Ganztagsprogramm optional nach dem Wunsch der Eltern in Anspruch genommen werden kann (offene Form). Im Vergleich der Angebotsbereitstellung von Ganztagsgrundschulen und Halbtagsgrundschulen, die erweiterte Angebote außerunterrichtlich und damit ohne Teilnahmeverpflichtung anbieten, lässt sich in Bezug auf eine Verfügbarkeit von erweiterten Lern- und Förderangeboten kein deutlicher Unterschied feststellen. Im Vergleich zu TIMSS 2007 zeigt sich, dass sowohl Ganztagsgrundschulen als auch Halbtagsgrundschulen ihr Angebot an außerunterrichtlichen Förderangeboten ausgeweitet haben.

In Bezug auf die für den Mathematik- und Sachunterricht relevanten Ausstattungsmerkmale zeigt sich im europäischen Vergleich für die Grundschulen in Deutschland eine verhältnismäßig günstige Situation. Nachbesserungsbedarf könnte jedoch in Bezug auf eine fachspezifische Ausstattung von Schul- und Klassenbibliotheken mit Büchern für den Mathematikunterricht, Labore für den Sachunterricht sowie eine zeitgemäße und ausreichende Verfügbarkeit von digitalen Medien bestehen. Mehr Computer zur Verfügung zu stellen reicht allerdings nicht aus, um eine Nutzung im Unterricht auszuweiten. Entsprechend müsste über Maßnahmen nachgedacht werden, die Lehrkräfte besser beim Einsatz digitaler Medien im Unterricht unterstützen. Auf der einen Seite könnten gezielte Fortbildungsangebote dazu beitragen, dass sich mehr Lehrkräfte einen souveränen Einsatz zutrauen, auf der anderen Seite könnten mehr Unterstützungsangebote an der Einzelschule, beispielsweise durch technischen Support, die Computernutzung im Unterricht erleichtern.

Geschlechtsspezifische Unterschiede in mathematischen und naturwissenschaftlichen Kompetenzen

Wie bereits in TIMSS 2007 lassen sich auch in TIMSS 2011 signifikante geschlechtsspezifische Unterschiede in den mathematischen und naturwissenschaftlichen Leistungen von Viertklässlerinnen und Viertklässlern in Deutschland feststellen. Es zeigt sich, dass Jungen in Deutschland sowohl in Mathematik als auch in den Naturwissenschaften signifikant höhere Leistungen erzielen als Mädchen, wobei die Unterschiede vor allem in den Naturwissenschaften vergleichsweise hoch sind. Im Vergleich mit den teilnehmenden Mitgliedsstaaten der EU und OECD zeigt sich, dass lediglich in der Tschechischen Republik die geschlechtsspezifischen Unterschiede in den Naturwissenschaften um drei Leistungspunkte höher ausfallen als in Deutschland (12). Im europäischen Vergleich zeigt sich, dass für Staaten wie England, Irland, Litauen, Nordirland, Portugal, Rumänien, Schweden und Ungarn sowohl für die Gesamtskala Mathematik als auch für die Gesamtskala Naturwissenschaften keine signifikanten geschlechtsspezifischen Leistungsunterschiede festzustellen sind. Für Dänemark, Finnland, Norwegen und Slowenien zeigen sich lediglich im Bereich Mathematik signifikante Leistungsunterschiede, die zugunsten der Jungen ausfallen.

Relative Leistungsstärken zeigen die Jungen in Deutschland in den mathematischen Inhaltsbereichen *Arithmetik* und *Geometrie/Messen*. Auch in den mathematischen Anforderungsbereichen *Reproduzieren* und *Problemlösen* erreichen Jungen in Deutschland höhere Leistungswerte als Mädchen. Für die naturwissenschaftlichen Inhaltsbereiche zeigen sich relative Leistungsstärken der Jungen in *Physik/Chemie* und *Geographie*. In den Naturwissenschaften erzielen Jungen in Deutschland in allen kognitiven Anforderungsbereichen signifikant bessere Leistungen als Mädchen.

Im Vergleich der beiden Zyklen TIMSS 2007 und 2011 scheint sich auf den ersten Blick weder für die Mädchen noch für die Jungen in Deutschland eine Veränderung ergeben zu haben. Bei genauerer Betrachtung der Inhalts- und Anforderungsbereiche zeigt sich jedoch, dass für Jungen im Bereich *Geometrie/Messen* und für Mädchen im Bereich *Umgang mit Daten* Leistungsverbesserungen gefunden werden können. Allerdings haben die Jungen im Inhaltsbereich *Biologie* im Vergleich signifikant niedrigere Leistungen erzielt als 2007. Im Anforderungsbereich *Reproduzieren* in Mathematik zeigen sich sowohl für die Jungen als auch für die Mädchen signifikant höhere Kompetenzen. Außerdem können die Mädchen in den Naturwissenschaften im Bereich *Anwenden* positive Leistungsveränderungen verzeichnen. Im internationalen Vergleich zeigt sich, dass sich in einigen Teilnehmerstaaten zwar signifikante Veränderungen der mittleren Leistungen von Jungen und Mädchen feststellen lassen, es aber keinem teilnehmenden Mitgliedsstaat der EU oder OECD gelungen ist, im Vergleich zu TIMSS 2007 Leistungsunterschiede zwischen den Geschlechtern signifikant zu verringern.

Wie bereits 2007 zeigen die Grundschülerinnen und Grundschüler in Deutschland 2011 sowohl zur Mathematik als auch zum Fach Sachunterricht ausgesprochen positive Einstellungen. Für das Fach Mathematik wird deutlich, dass Jungen sowohl eine höhere positive Einstellung zur Mathematik haben als auch ein höheres mathematikbezogenes Selbstkonzept aufweisen. Die geschlechtsspezifischen Unterschiede in den positiven Einstellungen zur Mathematik sind aber als eher gering zu bewerten. Allerdings zeigt sich im Vergleich zu 2007, dass der Anteil an

Mädchen mit hohen positiven Einstellungen um knapp 4 Prozentpunkte gesunken ist. In Bezug auf das mathematische Selbstkonzept lassen sich keine signifikanten Veränderungen zu TIMSS 2007 beobachten. Weiterhin zeigt sich, dass der Anteil an Mädchen in der Gruppe der Schülerinnen und Schüler mit einem niedrigen Selbstkonzept auch 2011 doppelt so hoch ist wie der Anteil der Jungen.

Unterschiede in den Einstellungen zum Sachunterricht und dem naturwissenschaftsbezogenen Selbstkonzept sind hingegen nicht signifikant. Etwa 80 Prozent der Mädchen und 79 Prozent der Jungen berichten von hohen positiven Einstellungen zum Fach Sachunterricht. Ebenso hoch liegen die Anteile von Jungen und Mädchen mit hohen positiven sachunterrichtsbezogenen Selbstkonzepten.

Insgesamt lässt sich festhalten, dass sowohl Leistungsunterschiede als auch Differenzen in Einstellungen und Selbstkonzepten zwischen Jungen und Mädchen in der Grundschule bestehen. Multivariate Analysen unter Einbezug der Variablen Geschlecht, Selbstkonzept, Note, Bildungsniveau der Eltern sowie sozioökonomischer Status zeigen in Mathematik, dass nach den Noten insbesondere das Selbstkonzept einen entscheidenden Anteil der Leistungsdifferenzen zwischen Jungen und Mädchen erklärt. Für die Naturwissenschaften zeigt sich hingegen, dass Geschlechterunterschiede sich auch unter Hinzunahme des Selbstkonzepts und der Noten als signifikant erweisen und sogar um etwa 4 Punkte ansteigen. Unter Hinzunahme der Information zur Zugehörigkeit zu der Gruppe mit hohem Bildungsniveau der Eltern sowie der Information zum familiären Berufsstatus verändert sich der Zusammenhang zwischen Geschlecht, Selbstkonzept und mathematischen wie naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern nur geringfügig. Diese Befunde verdeutlichen, dass hier Wechselwirkungen zwischen den fachspezifischen Leistungen und den Einstellungen und Selbstkonzepten bestehen, die sich zugunsten der Jungen auswirken. Dieser Befund verweist darauf, dass Programme, die auf einen optimierten Kompetenzerwerb von Mädchen wie Jungen abzielen, auch die fachspezifischen Einstellungen und Selbstkonzepte der Schülerinnen und Schüler berücksichtigen sollten, da die unterschiedliche Entwicklung von Selbstkonzepten von Mädchen und Jungen insbesondere im weiteren Verlauf der Bildungskarriere an Bedeutung gewinnen und beispielsweise zu unterschiedlichem Kurswahlverhalten in der Oberstufe führen kann.

Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften

In TIMSS 2011 zeigen sich – wie auch in TIMSS 2007 – für viele Teilnehmerstaaten zum Teil sehr stark ausgeprägte soziale Disparitäten in den mathematischen und naturwissenschaftlichen Leistungen der Schülerinnen und Schüler. In Deutschland haben Schülerinnen und Schüler, die berichten, dass ihre Familien mehr als 100 Bücher besitzen, in den mathematischen Kompetenzen gegenüber jenen, die angeben, maximal hundert Bücher zu besitzen, einen signifikanten Vorsprung, der etwa einem Lernjahr entspricht. Dieser Befund zeigt sich ebenfalls – mit wenigen Modifikationen – für den Kompetenzbereich der Naturwissenschaften. Im internationalen Vergleich zeigen sich signifikant höhere Disparitäten in den Schülerleistungen in Rumänien, Ungarn, England, Nordirland und Irland. Bedeutsam geringer ist der Leistungsvorsprung in beiden Kompetenzdomänen beispielsweise in den Niederlanden, Finnland, Portugal und

der Flämischen Gemeinschaft in Belgien. Positiv hervorzuheben ist in diesem Kontext, dass in Deutschland der Anteil der Schülerinnen und Schüler, die angeben, mehr als hundert Bücher zu haben, relativ hoch ausgeprägt ist. Vergleicht man diese Befunde zur sozialen Benachteiligung bei TIMSS 2011 mit denen aus 2007, so stellen sich in drei Staaten (für Mathematik) beziehungsweise in zwei Staaten (für Naturwissenschaften) statistisch bedeutsame Veränderungen im Leistungsvorsprung zugunsten der Kinder aus günstigen sozialen Lagen heraus. In den meisten Teilnehmerstaaten, die an beiden Erhebungen teilnahmen (so auch in Deutschland), zeigt sich keine bedeutsame Veränderung in Bezug auf das Ausmaß der sozialen Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler. Keinem Teilnehmerstaat der EU und der OECD ist es gelungen, im Vergleich zu TIMSS 2007 die sozialen Disparitäten in den Schülerleistungen zu verringern.

Da in Deutschland neben der Frage zum Buchbesitz in der Familie zusätzliche Indikatoren zur Messung der Benachteiligung im schulischen Kompetenzerwerb erhoben wurden, konnten für eine differenzierte Betrachtung ebenfalls Analysen in Bezug auf den Berufsstatus (nach EGP-Klassen) und die Armutsgefährdung (auch im Vergleich zu TIMSS 2007) durchgeführt werden. Es bestätigen sich die internationalen Befunde dahingehend, dass ausgeprägte soziale Disparitäten in den Schülerleistungen im deutschen Bildungssystem vorhanden sind, wobei die Leistungsunterschiede zwischen Kindern, deren Eltern den oberen Dienstklassen angehören, und Kindern, deren Eltern der Gruppe der Arbeiter angehören, etwa ein Lernjahr betragen. Selbiges gilt, wenn Viertklässlerinnen und Viertklässler aus armutsgefährdeten Elternhäusern mit Schülerinnen und Schülern verglichen werden, die nicht armutsgefährdet sind. Beim Vergleich mit den Ergebnissen aus TIMSS 2007 zeigen sich hinsichtlich der Unterschiede zwischen den EGP-Klassen zwar keine signifikanten Veränderungen, Regressionsanalysen mit einem anderen Indikator für den sozioökonomischen Status der Familie (HISEI) zeigen aber, dass das Ausmaß der sozialen Disparitäten in den Schülerleistungen 2011 etwas geringer ausfällt als noch vier Jahre zuvor.

Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund

Schülerinnen und Schüler, die zu Hause nicht die Testsprache sprechen, erzielen sowohl in Mathematik als auch in den Naturwissenschaften in fast allen Teilnehmerstaaten schlechtere Leistungen als ihre Mitschülerinnen und Mitschüler, die zu Hause die Testsprache sprechen. Dies ist auch in Deutschland nicht anders – hier sind die Disparitäten zwischen den Kompetenzen von Kindern mit und ohne Migrationshintergrund weiterhin Grund zur Sorge. Fast jedes dritte Kind in Deutschland, bei dem zu Hause manchmal oder nie Deutsch gesprochen wird, erzielt lediglich Leistungen, die unterhalb des Niveaus der Kompetenzstufe III liegen. Wird hingegen zu Hause Deutsch gesprochen, erzielten lediglich ein Sechstel der Kinder keine ausreichenden Leistungen in Mathematik. Auch in den Naturwissenschaften ist der Anteil der Kinder mit Migrationshintergrund auf den unteren beiden Kompetenzstufen mehr als doppelt so hoch im Vergleich zu den Kindern ohne Migrationshintergrund.

Betrachtet man die Leistungsstände, so zeigt sich, dass der Leistungsvorsprung von Schülerinnen und Schülern ohne Migrationshintergrund vor ihren Mit-

schülerinnen und Mitschülern mit Migrationshintergrund in Mathematik 2011 39 Punkte beträgt. Die Differenzen in den naturwissenschaftlichen Kompetenzen sind mit 57 Leistungspunkten noch stärker ausgeprägt als für die mathematischen Kompetenzen. Im Vergleich zu TIMSS 2007 zeigen sich jedoch positive Veränderungen: In Mathematik erzielten Schülerinnen und Schüler mit einem im Ausland geborenen Elternteil signifikant bessere Leistungen als bei TIMSS 2007. Auch in den Naturwissenschaften konnten die Schülerinnen und Schüler mit Migrationshintergrund aufholen. Hier erreichen nicht nur die Viertklässlerinnen und Viertklässler mit einem im Ausland geborenen Elternteil bessere Leistungen, sondern auch die Schülerinnen und Schüler mit beiden im Ausland geborenen Eltern. Dennoch ist Vorsicht geboten, diese Veränderung als einen gezielten Effekt der Förderung von Kindern mit Migrationshintergrund zu bewerten. Die großen Leistungsdifferenzen verweisen vielmehr darauf, dass die Förderung von Kindern mit Migrationshintergrund auch weiterhin ein wichtiges Handlungsfeld bilden muss. Es zeigt sich, dass die Unterschiede in den mathematischen und naturwissenschaftlichen Kompetenzen vor allem in den sozioökonomischen und soziokulturellen Lebensverhältnissen von Schülerfamilien begründet sind. Förderprogramme sollten entsprechend – auch um Stigmatisierungen zu vermeiden – konzeptionell primär an den konkreten Förderbedarfen beziehungsweise individuellen Leistungsschwächen der Schülerinnen und Schüler orientiert sein. Mögliche Teilnahmemhemmnisse, die in der besonderen ökonomischen und sozialen Situation von Schülerfamilien begründet sind, sollten gleichwohl ernst genommen werden.

Leistungsprofile von Viertklässlerinnen und Viertklässlern in Deutschland

Mit TIMSS werden in einem vierjährigen Zyklus die mathematischen und naturwissenschaftlichen Leistungen von Schülerinnen und Schülern – unter anderem am Ende der vierten Jahrgangsstufe – untersucht. Mit der *Internationalen Grundschul-Lese-Untersuchung (IGLU)/Progress in International Reading Literacy Study (PIRLS)* wird in einem fünfjährigen Zyklus das Leseverständnis von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe untersucht. Beide Studien werden von der *International Association for the Evaluation of Educational Achievement (IEA)* verantwortet. Im Jahr 2011 fiel der Erhebungszeitpunkt der beiden Studien IGLU und TIMSS erstmalig zusammen, so dass sich die Möglichkeit ergab, die Studien gemeinsam durchzuführen. In Deutschland wurde – wie in 37 weiteren Staaten und Regionen (davon drei als Benchmark-Teilnehmer) – eine gemeinsame Stichprobe von Schülerinnen und Schülern realisiert (zu den Einzelheiten der Administration vgl. Kapitel 2 in diesem Band). Die gemeinsame Stichprobe umfasst 197 Grundschulen und 4229 Schülerinnen und Schüler. Die Testung für IGLU und TIMSS fand in der Regel an zwei aufeinanderfolgenden Testtagen statt, wobei an 50 Prozent der Schulen am ersten Testtag TIMSS und an den anderen 50 Prozent der Schulen am ersten Testtag IGLU durchgeführt wurde (vgl. Kapitel 2). Diese besondere Datenlage eröffnet weiterführende Möglichkeiten der Auswertung.

Die Testleistungen in den drei Bereichen Lesen, Mathematik und Naturwissenschaften wurden für Deutschland gemeinsam skaliert, um so die Ergebnisse direkt miteinander vergleichen zu können. In die Skalierung wurde ein nationales Hintergrundmodell mit eingeschlossen und es wurde ein eindimen-

sionales (Referenzmodell) gegen ein drei- (IGLU- und TIMSS-Gesamtskalen) beziehungsweise achtdimensionales (IGLU- und TIMSS-Subskalen) Skalierungsmodell getestet. Da die üblichen Auswahlkriterien wie AIC, CAIC und BIC auf Grund der schnell ansteigenden Modellkomplexität mit wachsender Dimensionalität für die betrachtete Modellfamilie nicht unmittelbar einsetzbar zu sein schienen, wurde der *Likelihood Ratio Test* (LRT) als Alternative betrachtet. Der Vergleich der Modelle unter Nutzung der LRT-Statistik zeigte, dass sowohl das drei- als auch das achtdimensionale Modell gegenüber dem eindimensionalen Modell zu bevorzugen sind. Weitere Forschung wäre an dieser Stelle nötig, um geeignete informationstheoretische Auswahlmaße in dieser Modellfamilie zu entwickeln.

Die mehrdimensionale Modellierung ermöglicht auch die Bestimmung messfehlerkorrigierter Korrelationen. Zwischen den Leistungen in den drei Tests bestehen substantielle Zusammenhänge. Während die Leistungen in Mathematik und Naturwissenschaften mit .66 korrelieren, hängen Leistungen in den naturwissenschaftlichen Tests (.74) noch stärker als die in den Mathematiktests (.54) mit dem Leseverständnis zusammen. Dies ist auch auf das Aufgabenformat zurückzuführen, denn das Verständnis und die Lösung der naturwissenschaftlichen Aufgaben setzen die Lesefähigkeit voraus.

Mithilfe einer latenten Profilanalyse ließen sich sieben Typen von Schülerinnen und Schülern unterscheiden. Es zeigte sich, dass über alle Domänen hinweg die mittleren Leistungen der verschiedenen Typen zu den jeweils höheren Leistungstypen ansteigen. Die Leistungen in den drei Domänen sind bei der Mehrheit der Kinder, wenn auch auf unterschiedlichen Niveaus, recht ausgeglichen. Große Unterschiede in den Leistungen sind die Ausnahme – am deutlichsten treten sie bei Schülerinnen und Schülern mit geringen und sehr geringen Leistungen auf, bei denen die Leseleistungen am schlechtesten ausfallen.

Die sieben Typen mit unterscheidbaren Leistungsprofilen wurden in einem weiteren Auswertungsschritt daraufhin betrachtet, ob sie sich in relevanten Hintergrundmerkmalen unterscheiden. Der einzige Leistungstyp, in dem sich ein signifikanter Geschlechterunterschied zeigt, ist die Gruppe der Schülerinnen und Schüler mit herausragenden Leistungen in allen Bereichen, die allerdings mit nur 4 Prozent einen sehr kleinen Anteil der Schülerinnen und Schüler in Deutschland repräsentiert. Jungen sind in diesem Leistungstyp mit einem Anteil von 58,3 Prozent signifikant häufiger vertreten als Mädchen. Dieser Befund reflektiert das generelle Phänomen, wonach in der Gruppe der Schülerinnen und Schüler mit herausragenden Leistungen zumeist mehr Jungen als Mädchen vertreten sind. Es scheint der Grundschule folglich nicht zu gelingen, Mädchen in allen Kompetenzbereichen so zu fördern und zu motivieren, dass sie einen vergleichbaren Anteil in der Gruppe der Schülerinnen und Schüler mit extrem hohen Kompetenzen in allen Bereichen haben.

Die sieben Leistungstypen unterscheiden sich, zumeist signifikant, voneinander in Bezug auf kulturelle und soziale Hintergrundmerkmale. Insgesamt ist bei den Gruppen mit geringen und sehr geringen Leistungen eine Bildungsbenachteiligung zu konstatieren, die sich besonders deutlich durch den Kontrast mit den beiden leistungsstärksten Gruppen abzeichnet. Betrachtet man die Anteile der Kinder in der obersten und der untersten Leistungsgruppe, deren Eltern einen Hochschulabschluss haben, dann ergibt sich eine Differenz von etwa 61 Prozentpunkten. Es zeigt sich auch eine sehr enge Kopplung zwischen sozioökonomischen Merkmalen und der Zugehörigkeit der Schülerinnen und Schüler zu bestimmten Leistungstypen.

Etwa 12 Prozent der Schülerfamilien von Kindern des Leistungstyps mit geringen und sehr geringen Leistungen kann der dritten Sozialschicht zugeordnet werden. Bei den Leistungstypen mit sehr hohen und herausragenden Leistungen sind es mehr als die Hälfte der Schülerinnen und Schüler, deren Eltern, Akademiker, Techniker oder Führungskräfte sind. Während maximal jedes zehnte Kind in den Leistungstypen mit sehr hohen und herausragenden Leistungen von Armut gefährdet ist, betrifft dies etwa jedes zweite Kind aus den Gruppen mit geringen und sehr geringen Leistungen.


Auch die mit dem Migrationshintergrund verbundenen Merkmale unterscheiden sich zumeist signifikant in den sieben Leistungstypen. Schülerinnen und Schüler, die nur manchmal oder nie zu Hause die Testsprache sprechen, sind in den oberen Leistungstypen unter- und in den unteren Leistungstypen überrepräsentiert. Diese Befunde machen deutlich, dass hier weiterhin ein großer Bedarf an Maßnahmen zur Förderung der Unterrichtssprache besteht. Es kann jedoch nicht von einer monokausalen Beziehung zwischen Migrationsstatus und Kompetenzunterschieden ausgegangen werden, es zeigt sich vielmehr ein vielschichtiger Zusammenhang unter anderem zwischen der Migrationsgeschichte und den Sprachkenntnissen in den Familien, Bildungsabschlüssen, Berufen der Eltern und dem Buchbesitz in den Schülerfamilien.

Insgesamt ist bei den Leistungstypen mit geringen und sehr geringen Leistungen eine Bildungsbenachteiligung zu konstatieren, die sich besonders deutlich durch den Vergleich mit der begünstigten Gruppe von Kindern im Leistungstyp mit herausragenden Leistungen, aber auch im Typ mit sehr hohen Leistungen abzeichnet.

Die sieben Leistungstypen wurden nicht nur in Bezug auf kulturelle und soziale Hintergrundmerkmale untersucht, sondern auch in Bezug auf ihre fachspezifischen Einstellungen und Selbstkonzepte. In Gruppen mit höherer Leistung ist auch der Anteil an Schülerinnen und Schülern mit hohen fachbezogenen Selbstkonzepten signifikant höher. Für die meisten Vergleiche zwischen den Leistungstypen gilt, dass in Gruppen mit höherer Leistung der Anteil an Schülerinnen und Schülern mit hohen positiven Einstellungen zum Lesen signifikant höher ist, jedoch zeigt sich dieses Bild nicht für die Einstellungen zu Mathematik und Naturwissenschaften. Die Anteile von Schülerinnen und Schülern mit besonders positiven Einstellungen zu Mathematik beziehungsweise zu Naturwissenschaften unterscheiden sich zwischen den beiden Leistungstypen mit extrem niedrigen beziehungsweise extrem hohen Leistungswerten nicht signifikant. In der Grundschule scheinen sowohl Kinder mit herausragenden Leistungen als auch Schülerinnen und Schüler mit sehr geringen Leistungen gleichermaßen für Mathematik beziehungsweise Naturwissenschaften begeistert zu sein.

Literatur

- Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.). (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Tarelli, I., Bremerich-Vos, A. & Schwippert, K. (Hrsg.). (2012). *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Tarelli, I., Lankes, E.-M., Drossel, K. & Gegenfurtner, A. (2012). Lehr- und Lernbedingungen an Grundschulen im internationalen Vergleich. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.



Kapitel II

Ziele, Anlage und Durchführung der Trends in International Mathematics and Science Study (TIMSS 2011)

Heike Wendt, Irmela Tarelli, Wilfried Bos, Kristina Frey
und Mario Vennemann

1 Zielsetzung

Die *Trends in International Mathematics and Science Study* (TIMSS) ist eine international-vergleichende Schulleistungsuntersuchung, die von der *International Association for the Evaluation of Educational Achievement* (IEA) verantwortet wird. Mit TIMSS werden unter Berücksichtigung curricularer Vorgaben und anderer zentraler Rahmenbedingungen schulischer Lernumgebungen im internationalen Vergleich Fachleistungen von Schülerinnen und Schülern in den Domänen Mathematik und Naturwissenschaften am Ende der vierten und achten Jahrgangsstufe sowie am Ende der Sekundarstufe II (*TIMSS Advanced*) untersucht. Die IEA ist ein unabhängiger, internationaler Zusammenschluss von Wissenschaftlerinnen und Wissenschaftlern, Forschungseinrichtungen und Regierungsstellen und führt seit 1959 international-vergleichende Schulleistungsstudien durch. TIMSS gehört seit 1995 zu den Kernstudien der IEA und wird im Abstand von vier Jahren durchgeführt. Ein zentrales Anliegen ist es, langfristige Entwicklungen in den teilnehmenden Bildungssystemen zu dokumentieren.

Mit der Erhebung im Jahr 2011 beteiligte sich Deutschland zum zweiten Mal an der Grundschuluntersuchung von TIMSS. Die Teilnahme erfolgte auf Beschluss der *Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland* (KMK) und einer Vereinbarung zwischen der KMK und dem *Bundesministerium für Bildung und Forschung* (BMBF). Die erste Teilnahme an der Untersuchung mit Schülerinnen und Schülern der vierten Jahrgangsstufe erfolgte für den Studienzyklus 2007 (vgl. Bos et al., 2008). Darüber hinaus hat die Bundesrepublik Deutschland an der ersten TIMS-Studie im Jahr 1995 teilgenommen, damals *Third International Mathematics and Science Study*, jedoch nur Schülerinnen und Schüler der Sekundarstufen I und II einbezogen (vgl. Baumert, Bos & Lehmann, 2000).

Zu den aktuellen Bildungsstudien der IEA gehört auch die *Internationale Grundschul-Lese-Untersuchung (IGLU)/Progress in International Reading Literacy Study (PIRLS)*. Mit IGLU/PIRLS wird international-vergleichend das Leseverständnis von Viertklässlerinnen und Viertklässlern untersucht. IGLU/PIRLS wird seit 2001 in einem Rhythmus von fünf Jahren durchgeführt. Deutschland hat sich bisher an allen drei Studienzyklen 2001, 2006 sowie 2011 beteiligt.

Im Jahr 2011 fielen erstmalig die Erhebungszeitpunkte der beiden IEA-Grundschulstudien IGLU/PIRLS und TIMSS zusammen. Daraus ergab sich die Möglichkeit, beide Studien parallel oder gemeinsam durchzuführen. Den Teilnehmerstaaten und Regionen, die sich für eine gemeinsame Erhebung entschieden haben, war es dabei freigestellt, entweder dieselben Schülerinnen und Schüler in beiden Studien an zwei aufeinanderfolgenden Schultagen zu testen oder aber innerhalb einer Schule an einem Testtag verschiedene Klassen in jeweils nur einer Studie zu testen. In Deutschland wurde eine gemeinsame Schülerstichprobe realisiert (zu den Einzelheiten der Administration vgl. Abschnitt 6). Für die Bundesrepublik Deutschland – wie für weitere 37 Staaten und Regionen (davon drei als Benchmark-Teilnehmer) – ergeben sich hieraus aus TIMSS-Perspektive erweiterte Möglichkeiten, im internationalen Vergleich die Leistungsstände von Grundschulkindern in Deutschland in den Kompetenzbereichen Mathematik und Naturwissenschaften auch unter Berücksichtigung ihrer Lesefähigkeit differenzierter zu betrachten.

1.1 Erkenntnismöglichkeiten und zentrale Fragestellungen

Mit der Durchführung international-vergleichender Schulleistungsstudien verfolgt die IEA vorrangig das Ziel, den Ertrag von Bildungssystemen, das heißt die Schulleistungen und Einstellungen von Schülerinnen und Schülern am Ende einer vorgegebenen Zeitspanne festzustellen. Bei der Betrachtung werden gleichzeitig potentielle Einflussfaktoren berücksichtigt, um mehr über die Bedingungen und somit über Möglichkeiten der Verbesserung schulischer Förderung in Erfahrung zu bringen (IEA, 2012). In TIMSS werden als potentielle Einflussfaktoren zentrale Rahmenbedingungen des Kompetenzerwerbs und des Unterrichts in der Grundschule untersucht (vgl. Abschnitt 4.1.3). Zudem werden Aspekte berücksichtigt, welche die Nutzung der hieraus entstehenden Lerngelegenheiten beeinflussen.

Während die mathematischen und naturwissenschaftlichen Kompetenzen der Viertklässlerinnen und Viertklässler anhand von Leistungstests ermittelt werden, erfolgt die Erfassung der Rahmenbedingungen auf Grundlage der schriftlichen Befragung der Schülerinnen und Schüler, ihrer Eltern, Lehrkräfte und Schulleitungen sowie anhand der Einschätzungen von Experten (vgl. Abschnitt 4.1.3). Insofern bietet TIMSS vielfältige Erkenntnismöglichkeiten. Staaten, die regelmäßig an der alle vier Jahre stattfindenden Studie teilnehmen, erhalten Gelegenheit, Leistungsveränderungen im Zeitverlauf zu beobachten.

Aber auch Staaten, die zum ersten Mal an TIMSS teilnehmen, profitieren von der Teilnahme. Sie ermöglicht ihnen den Leistungsvergleich mit anderen Staaten und erlaubt es, einen Vergleichsmaßstab für zukünftige Trendanalysen zu etablieren (Mullis, Martin, Ruddock, O'Sullivan & Preuschoff, 2009). Die TIMS-Studien geben ebenso Einblick in die Bedingungen des Lehrens und Lernens von Mathematik und den Naturwissenschaften. Sie tragen zum

Verständnis von Bildungsprozessen innerhalb einzelner Staaten bei sowie auch zum Verständnis der Unterschiede verschiedener Bildungssysteme und ihrer Erträge. Aus diesen Erkenntnissen lassen sich Hinweise für Verbesserungen und Informationen für Entscheidungsträger ableiten. Für die Wissenschaft stellt die Studie wertvolle Referenzwerte zur Verfügung und zeigt gezielt weitergehenden Forschungsbedarf auf.

Die Teilnahme an TIMSS 2011 und die damit verbundene Generierung von umfassenden empirischen Daten, die repräsentative Rückschlüsse auf die Untersuchungspopulation (Schülerinnen und Schüler am Ende der vierten Jahrgangsstufe) erlauben, bietet aus Sicht der Bundesrepublik Deutschland besondere Erkenntnisgewinne:

- Die standardisierte Durchführung und statistisch angemessene Auswertung von elaborierten und erprobten Kompetenztestungen in den Bereichen Mathematik und naturwissenschaftliche Grundbildung ermöglichen eine zuverlässige Einschätzung des Leistungsniveaus von Viertklässlerinnen und Viertklässlern in Deutschland im internationalen Vergleich.
- Die Teilnahme an TIMSS 2011 ermöglicht es, im Vergleich zu TIMSS 2007 Veränderungen in den mathematischen und naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe zu beschreiben und international-vergleichend einzuordnen.
- Die standardisiert administrierte Befragung der Schülerinnen und Schüler zu ihren fachspezifischen Selbstkonzepten, ihren fachbezogenen Einstellungen sowie zu ihrem schulischen und außerschulischen Lernverhalten ermöglicht eine zuverlässige Einschätzung der Bedeutsamkeit dieser Aspekte für die Leistungsstände von Viertklässlerinnen und Viertklässlern in Deutschland.
- Die standardisiert administrierte Befragung weiterer schulischer Akteure (z.B. Fachlehrkräfte und Schulleitungen) zu zentralen Bedingungsfaktoren schulischen Lernens ermöglicht eine zuverlässige Einschätzung der Bedeutsamkeit dieser Aspekte für die Leistungsstände von Viertklässlerinnen und Viertklässlern in Deutschland im internationalen Vergleich.
- Die Stichprobenziehung ermöglicht darüber hinaus, belastbare Aussagen zum Leistungsniveau sowie zur Bedeutsamkeit von individuellen Lernvoraussetzungen und Kontextfaktoren für Subpopulationen (z.B. im Hinblick auf die Aspekte soziale Herkunft, Migrationshintergrund und Geschlecht) vorzunehmen. Hier bietet TIMSS 2011 erstmals die Gelegenheit, international-vergleichend Zusammenhänge zwischen familiärer Herkunft sowie elterlichem Erziehungsverhalten einerseits und den Mathematik- und Naturwissenschaftsleistungen andererseits differenzierter zu analysieren, da einige Teilnehmerstaaten das Instrumentarium um eine standardisierte Elternbefragung erweitert haben.

Der vorliegende Berichtsband zu Ergebnissen von TIMSS 2011 fokussiert die Ergebnisse des internationalen Vergleichs aus deutscher Perspektive. Vertiefende Fragestellungen sind Nachfolgepublikationen vorbehalten. Im Mittelpunkt dieses Berichts steht vorrangig die Beantwortung der folgenden Fragestellungen:

1. Welche Kompetenzniveaus zeigen Schülerinnen und Schüler in den Bereichen Mathematik und Naturwissenschaften am Ende der vierten Jahrgangsstufe in Deutschland im Jahr 2011? Wie lassen sich die Ergebnisse im internationalen Vergleich bewerten? Haben sich die Ergebnisse seit 2007 verändert?
2. Wie lassen sich die TIMSS-Leistungskennwerte auf Kompetenzstufen einordnen? Wie groß sind die Gruppen der auffällig leistungsschwachen und leistungsstarken Schülerinnen und Schüler? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und 2011?
3. Welche Ergebnisse erzielen die Viertklässlerinnen und Viertklässler in Deutschland in den domänenspezifischen Inhaltsbereichen und den kognitiven Anforderungsbereichen der TIMS-Studie?
4. Welche Bedeutung haben die individuellen Lernvoraussetzungen und Kontextfaktoren für Kompetenzen in den Bereichen Mathematik und Naturwissenschaften? Welche Veränderungen von Lehr- und Lernbedingungen lassen sich seit 2007 beobachten?

2 TIMSS 2011 – ein kooperatives Unternehmen

Die aufwendige Realisierung internationaler Vergleichsstudien wie TIMSS erfordert die Zusammenarbeit verschiedener Organisationen, Institutionen und Personen auf nationaler und internationaler Ebene. Die IEA hat mit dem internationalen Management der Studie das *TIMSS & PIRLS International Study Center* (ISC) am *Boston College* in Chestnut Hill, Massachusetts, USA beauftragt, welches unter Leitung von Ina V. S. Mullis, *Professor* an der *Lynch School of Education, Boston College* sowie Michael O. Martin, *Research Professor* an der *Lynch School of Education, Boston College* steht. Das ISC ist verantwortlich für das Design und die Implementation der Studie, für die internationale Koordination der Entwicklung der Instrumente und der Erhebungsprozeduren sowie für die Qualität der Datenerhebung. Darüber hinaus werden am ISC die internationale Skalierung (vgl. Abschnitt 9) durchgeführt und der internationale Ergebnisbericht verfasst. Für die Stichprobenziehung, die Dokumentation der nationalen Stichproben und die Berechnung der internationalen Stichprobengewichte kooperiert die internationale Studienleitung mit *Statistics Canada* (Marc Joncas) in Ottawa, Ontario und der Abteilung ‚Stichprobenziehung‘ am *IEA Data Processing and Research Center* (DPC) in Hamburg (Dr. Sabine Meinck).

In jedem der an TIMSS 2011 beteiligten Staaten trägt ein nationaler Projektkoordinator (*National Research Coordinator*, NRC) die Verantwortung für die Vorbereitung und Durchführung der Studie gemäß den international vorgegebenen Richtlinien. In Deutschland wurde Prof. Dr. Wilfried Bos am Institut für Schulentwicklungsforschung (IFS) der Technischen Universität Dortmund mit dieser Aufgabe betraut. Die Projektleitung und Koordination der Studie verantworten für Deutschland Dr. Kristina Frey (bis März 2011) und Dr. Heike Wendt (ab März 2011). Die Durchführung von TIMSS in Deutschland wird zu gleichen Teilen vom BMBF und durch die KMK finanziert.

Für die Analyse der Studienergebnisse und die Berichtslegung in Deutschland ist ein nationales Konsortium unter Federführung des IFS verantwortlich, dem folgende Wissenschaftlerinnen und Wissenschaftler angehören:

Prof. Dr. Wilfried Bos	(NRC, Wissenschaftlicher Leiter von TIMSS 2011 in Deutschland und Sprecher des Konsortiums) – Professor für Bildungsforschung und Qualitätssicherung an der Technischen Universität Dortmund, Direktor des Instituts für Schulentwicklungsforschung (IFS)
Prof. Dr. Christoph Selter	Professor für Mathematikdidaktik an der Technischen Universität Dortmund; Leiter des Instituts für Entwicklung und Erforschung des Mathematikunterrichts
Prof. Dr. Olaf Köller	Professor für Empirische Bildungsforschung an der Christian-Albrechts-Universität zu Kiel, Geschäftsführender Direktor des Leibniz-Instituts für die Pädagogik der Naturwissenschaften und Mathematik (IPN) und Direktor der Abteilung Erziehungswissenschaft
Prof. Dr. Gerd Walther	(Kooptiertes Mitglied des Konsortiums) – Professor für Mathematik und ihre Didaktik an der Christian-Albrechts-Universität zu Kiel
Prof. Dr. Manfred Prenzel	(Kooptiertes Mitglied des Konsortiums) – Professor für Empirische Bildungsforschung an der Technischen Universität München (TUM) sowie Dekan der TUM School of Education
Dr. Thilo Kleickmann	(Kooptiertes Mitglied des Konsortiums) – Wissenschaftlicher Mitarbeiter am Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN)
Dr. Kristina Frey	(Kooptiertes Mitglied des Konsortiums) – Wissenschaftliche Mitarbeiterin am Institut für Erziehungswissenschaft der Westfälischen Wilhelms-Universität Münster
Dr. Heike Wendt	(Projektleitung) Wissenschaftliche Mitarbeiterin am Institut für Schulentwicklungsforschung (IFS) an der Technischen Universität Dortmund

Für die fachdidaktische Mitgestaltung der Studie waren zwei Projektgruppen verantwortlich. Die Arbeitsgruppe Mathematik besteht aus Prof. Dr. Christoph Selter, Prof. Dr. Gerd Walther und Jan Wessel. Der Arbeitsgruppe Naturwissenschaften gehören Prof. Dr. Olaf Köller, Dr. Thilo Kleickmann, Steffani Saß und Prof. Dr. Manfred Prenzel an.

Ein großer Teil der Projektarbeit erfolgte hauptverantwortlich am IFS unter der Leitung von Dr. Heike Wendt. Aufgrund der großen Schnittmenge der beiden Studien wurde zur Haupterhebung 2011 eine gemeinsame Arbeitsgruppe für IGLU/TIMSS 2011 eingerichtet. Zu den beteiligten wissenschaftlichen Mitarbeiterinnen und Mitarbeitern zählen: Dr. Irmela Tarelli (Projektleitung IGLU/PIRLS), Theresia Brehl, Kerstin Drossel, Benjamin Euen, Magdalena Kowoll, Silvia Schmitz, Mario Vennemann und Anika Zylowski. Das Projektteam wurde zudem von einer Vielzahl studentischer Mitarbeiterinnen und Mitarbeiter unterstützt.

Mit der Organisation der nationalen Datenerhebung und -verarbeitung sowie der Aufgabenkodierung hat das IFS das *IEA Data Processing and Research Center* (DPC) in Hamburg beauftragt.

3 Zur Teilnahme der Bundesrepublik Deutschland an international-vergleichenden Schulleistungsuntersuchungen in den Domänen Mathematik und Naturwissenschaften am Ende der vierten Jahrgangsstufe

TIMSS ist als Trendstudie konzipiert und wird in regelmäßigen Abständen wiederholt. Von Seiten der IEA ist TIMSS 2011 die vierte Untersuchung des mathematischen und naturwissenschaftlichen Grundverständnisses von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe: Als Vorläuferstudie gelten (1) die erste und zweite Mathematikstudie – *First and Second International Mathematics Study* (FIMS, SIMS) – sowie (2) die *First and Second International Science Study* (FISS, SISS). Während FIMS und SIMS in den Jahren 1964 beziehungsweise 1980-1982 lediglich mit einem Fokus auf die Sekundarstufen I und II realisiert wurden, fanden die beiden entsprechenden Naturwissenschaftsstudien in den Jahren 1970-1971 und 1983-1984 mit einem Fokus auf die vierte Jahrgangsstufe sowie auf die Sekundarstufen I und II statt (vgl. Leung, 2010). Deutschland nahm zwar mit einer Substichprobe (Hessen und Schleswig-Holstein) an FIMS und mit einer für Deutschland repräsentativen Stichprobe an FISS, nicht jedoch an den Folgestudien SIMS und SISS teil (vgl. u.a. Goy, van Ackeren & Schwippert, 2008). Erst im Jahr 1995 nahm die Bundesrepublik Deutschland dann mit Schülerinnen und Schülern der Sekundarstufen I und II an (3) TIMSS 1995 teil (vgl. Baumert et al., 2000).

Auf die überraschend mittelmäßigen Ergebnisse der Schülerinnen und Schüler der Sekundarstufen I und II in Deutschland (Baumert et al., 1997) hat die KMK mit dem Konstanzer Beschluss vom 24. Oktober 1997 reagiert und unter anderem vereinbart, künftig an international-vergleichenden Schulleistungsstudien teilzunehmen, um weitere Erkenntnisse über Kompetenzen von Schülerinnen und Schülern an zentralen Schnittstellen im Schulsystem zu erhalten. Seit dem Jahr 2000 nimmt die Bundesrepublik Deutschland daher an dem im Dreijahresrhythmus stattfindenden *Programme for International Student Assessment* (PISA) für die Sekundarstufe I teil, seit 2001 an IGLU/PIRLS, die für die Grundschule im Fünfjahresabstand erhoben wird, und seit 2007 im Vierjahreszyklus an der TIMS-Studie (Grundschule).

In der TIMS-Studie aus dem Jahr 1995 wurden zwar erstmalig in einem breiten internationalen Rahmen die Mathematik- und Naturwissenschaftsleistungen von Schülerinnen und Schülern am Ende der dritten und vierten Jahrgangsstufe erhoben, die Bundesrepublik Deutschland hatte sich jedoch am Grundschulteil von TIMSS 1995 nicht beteiligt. Im Jahr 2001 wurden dann erstmals in einem Erweiterungsteil von IGLU/PIRLS (IGLU-E) die Mathematik- und Naturwissenschaftsleistungen von Viertklässlerinnen und Viertklässlern in Deutschland untersucht (vgl. Bos et al., 2003). Die Leistungstests waren so konzipiert, dass eine vorsichtige Einordnung der Ergebnisse von Schülerinnen und Schülern in Deutschland in die damals gut fünf Jahre alte internationale Leistungsskala von TIMSS 1995 möglich war (vgl. Prenzel, Geiser, Langeheine & Lobemeier, 2003; Walther, Geiser, Langeheine & Lobemeier, 2003). IGLU-E 2001 war damit die erste umfassende Studie, die für die Bundesrepublik Deutschland mathematische Kompetenzen und naturwissenschaftliches Verständnis von Kindern im Grundschulalter untersucht hat. Die Berichterstattung zum internationalen Vergleich sowie die Ergebnisse zur nationalen Ergänzung (IGLU-E)

liegen in publizierter Form vor (Bos et al., 2004, 2003; Bos, Lankes, Prenzel, Valtin & Walther, 2005; Mullis, Martin, Gonzalez & Kennedy, 2003).

Im Jahr 2007 nahm die Bundesrepublik Deutschland erstmalig regulär an TIMSS für die vierte Jahrgangsstufe teil (Bos et al., 2008). Neben Deutschland waren weitere 35 Staaten sowie 7 Benchmark-Teilnehmer beteiligt, hiervon mit Deutschland 14 Mitgliedsstaaten der *Europäischen Union* (EU) und 16 Mitgliedsstaaten der *Organisation for Economic Co-operation and Development* (OECD). Besonderheiten der TIMSS 2007-Erhebung in Deutschland sind Bensen, Lintorf, Bos & Frey (2008) zu entnehmen.

Im Vergleich zum vorangegangenen Studienzyklus zeichnet sich TIMSS 2011 durch drei Besonderheiten aus:

- Während bei TIMSS 2007 für die Einordnung der Leistungsergebnisse der Grundschul Kinder primär der internationale Bezugsrahmen zur Verfügung stand, ist es nun mit TIMSS 2011 möglich, Kompetenzveränderungen zu TIMSS 2007 auf solider Datengrundlage zu beschreiben.
- In 39 Teilnehmerstaaten und Regionen wurde das Instrumentarium der Studie erstmals um eine standardisierte Elternbefragung erweitert. TIMSS 2011 bietet damit erstmalig die Gelegenheit, international-vergleichend Zusammenhänge zwischen familiärer Herkunft sowie elterlichem Erziehungsverhalten einerseits und den Mathematik- und Naturwissenschaftsleistungen andererseits differenziert zu analysieren.
- Zudem wurde die Studie gemeinsam mit der ebenfalls von der IEA initiierten Schwesterstudie IGLU/PIRLS realisiert (vgl. Bos, Tarelli, Bremerich-Vos & Schwippert, 2012).

4 Anlage und Durchführung von TIMSS 2011

TIMSS betrachtet unter Berücksichtigung curricularer Vorgaben und anderer zentraler Rahmenbedingungen schulischer Lernumgebungen im internationalen Vergleich Fachleistungen von Schülerinnen und Schülern in den Domänen Mathematik und Naturwissenschaften am Ende der vierten Jahrgangsstufe. Im vorliegenden Bericht werden zentrale Ergebnisse aus TIMSS 2011 vorgestellt. Dabei stehen die Ergebnisse des internationalen Vergleichs im Vordergrund; es werden aber auch nationale Analysen zu weiterführenden Fragestellungen referiert. Zum besseren Verständnis der Ergebnisse der Studie wird in den folgenden Abschnitten zunächst das Design von TIMSS 2011 dargestellt. Daran anschließend werden Besonderheiten der für diesen Band genutzten Darstellungen erläutert.

4.1 Die Rahmenkonzeption der Studie

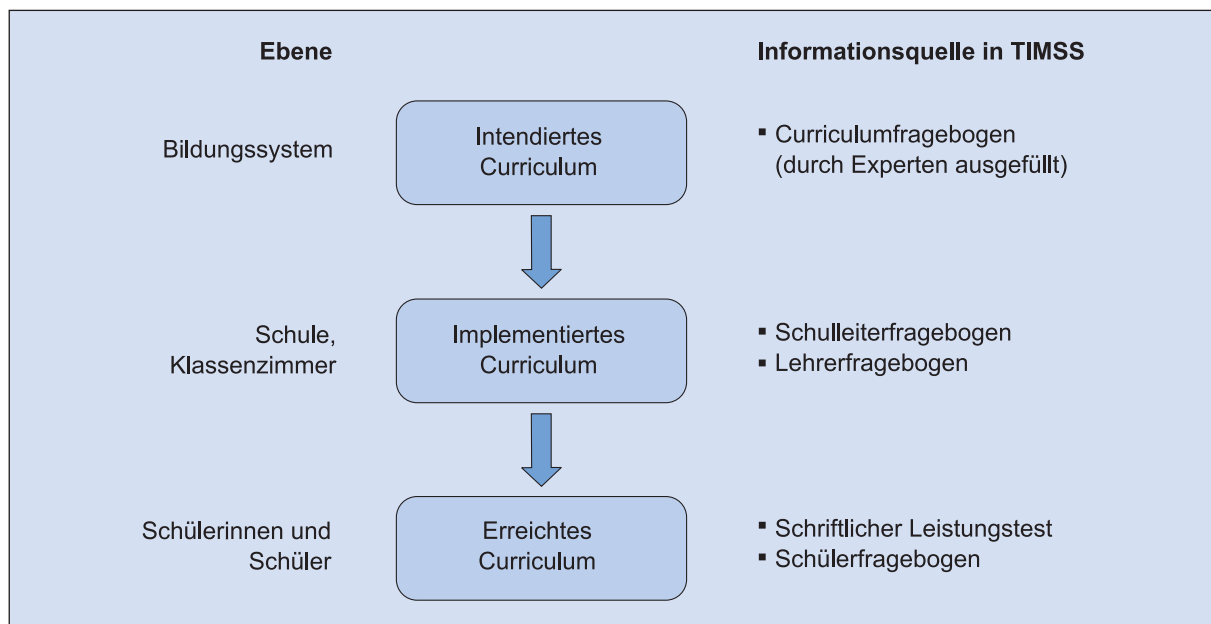
Im Mittelpunkt von TIMSS steht die Frage, wie erfolgreich Schülerinnen und Schüler in den Bildungssystemen der teilnehmenden Staaten und Regionen am Ende der vierten Jahrgangsstufe grundlegende Kompetenzen erworben haben. Die Konzeption von TIMSS muss somit belastbare Aussagen über die Kompetenzen von Schülerinnen und Schülern im internationalen Vergleich erlauben. Zusätzlich sollte es die Konzeption ermöglichen, auf der Ebene einzelner Staaten differenzierte Erkenntnisse über Stärken und Schwächen der Schülerinnen und Schüler

in unterschiedlichen Teilkompetenzen zu gewinnen. Dabei ist stets auch von Interesse, ob sich Teilgruppen von Schülerinnen und Schülern identifizieren lassen, die sich in ihren Kompetenzen systematisch voneinander unterscheiden. Die theoretischen und konzeptionellen Grundlagen der Studie sind in einem umfassenden Rahmenkonzept festgelegt, das Expertinnen und Experten aus vielen Teilnehmerstaaten erarbeitet haben (Mullis, Drucker, Preuschoff, Arora & Stanco, 2011; Mullis et al., 2009). Die Rahmenkonzeption von TIMSS 2011 lehnt sich aus Gründen der Kontinuität stark an die von TIMSS 2007 (Mullis et al., 2005) an. Weiterentwicklungen betreffen Aspekte der Messqualität und der Verwertbarkeit der Ergebnisse für Teilnehmerstaaten (Mullis et al., 2009).

4.1.1 Das Curriculum-Modell

Kern der Rahmenkonzeption von TIMSS 2011 ist das Curriculum-Modell, das bereits für die früheren Durchführungen der TIMS-Studie die Grundlage bildete (z.B. Baumert et al., 2000; Robitaille et al., 1993). Anders als IGLU/PIRLS und PISA, die konzeptionell am *Literacy*-Konzept ausgerichtet sind, folgt TIMSS einem Ansatz, nach dem auch eine partielle Gültigkeit der Leistungstests für die Curricula der teilnehmenden Staaten und Regionen angestrebt wird. Die Rahmenkonzeption von TIMSS sieht entsprechend vor, dass Schülerleistungen im jeweils spezifischen curricularen Kontext interpretiert werden. Das TIMSS-Curriculum-Modell ist in Abbildung 2.1 illustriert. Konzeptionell wird bei der Beschreibung des Curriculums eines Landes dreistufig zwischen dem *intendierten Curriculum*, dem *implementierten Curriculum* und dem *erreichten Curriculum* unterschieden. Das intendierte Curriculum repräsentiert dabei die mathematischen und naturwissenschaftlichen Inhalte beziehungsweise Prozesse, welche die Schülerinnen und Schüler nach staatlicher Vorgabe lernen sollten. Als implementiertes Curriculum wird der tatsächlich unterrichtete Lernstoff verstanden.

Abbildung 2.1: Das TIMSS-Curriculum-Modell



Als das erreichte Curriculum werden schließlich das von den Schülerinnen und Schülern Gelernte sowie ihre Einstellungen zum unterrichteten Fach gefasst. Dies wird durch die erzielten Testleistungen und Befragungsergebnisse aufgezeigt.

4.1.2 Untersuchte Kompetenzbereiche in TIMSS

Mit TIMSS werden mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe untersucht. Der Testentwicklung liegt dabei ein Kompetenzmodell zugrunde, das für die Domänen Mathematik und naturwissenschaftliche Grundbildung getrennt spezifiziert, welche grundlegenden Stoffgebiete (Inhaltsbereiche) und fachspezifisch zu aktivierenden kognitiven Denkprozesse und -tätigkeiten (kognitive Anforderungsbereiche) durch die Leistungstests abgebildet werden sollen (Mullis et al., 2009). Diese zweidimensionale Differenzierung soll es ermöglichen, mathematische und naturwissenschaftliche Kompetenzen möglichst umfassend jedoch so spezifisch abzubilden, dass sich sowohl ausreichende Schnittmengen mit den Curricula der beteiligten Bildungssysteme ergeben als auch unterschiedliche Schülerleistungen als Teilkompetenzen differenziert beschreiben lassen. Zudem soll durch eine Unterscheidung von inhaltlichen Bereichen und unterschiedlichen Niveaus kognitiver Anforderungen vermieden werden, dass überwiegend Aufgaben zum Einsatz kommen, die nur die Aktivierung von abrufbarem Faktenwissen oder die Anwendung von Grundfertigkeiten und die Nutzung von Grundvorstellungen erfordern.

Das Kompetenzmodell von TIMSS 2011 differenziert fachspezifisch je drei Inhaltsbereiche und drei kognitive Anforderungsbereiche (vgl. Tabelle 2.1). Das Modell ist hier zunächst jahrgangsübergreifend für die Erfassung von Lernständen in der vierten und achten Jahrgangsstufe konzipiert. Allerdings werden sowohl die Inhaltsbereiche als auch die kognitiven Anforderungen dem Alter und Entwicklungsstand angemessen differenziert und in unterschiedlichen Anforderungsgraden und Zeitumfängen berücksichtigt. Inhaltlich wird für die Erfassung mathematischer Kompetenzen am Ende der vierten Jahrgangsstufe zwischen den Bereichen *Arithmetik*, *Geometrie/Messen* und dem *Umgang mit Daten* unterschieden. Für die Naturwissenschaften werden die Bereiche *Biologie*, *Physik* und *Geographie* differenziert, wobei der Inhaltsbereich *Physik* auch Themenstellungen umfasst, die in Deutschland traditionell der Chemie zugeordnet werden. Die einzelnen Inhaltsbereiche werden zudem fachbezogen in verschiedene Themengebiete aufgegliedert. Um grundlegende Denkprozesse zu beschreiben, die für eine erfolgreiche Aufgabenbearbeitung zu bewältigen sind, werden die kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen* unterschieden, denen wiederum konkrete Verhaltensweisen zuzuordnen sind (Mullis et al., 2009; siehe auch Kapitel 3 und 4 in diesem Band).

Tabelle 2.1: Inhaltsbereiche und kognitive Anforderungsbereiche in TIMSS 2011

Mathematik		Naturwissenschaften	
<i>Inhaltsbereiche</i>	<i>Kognitive Anforderungsbereiche</i>	<i>Inhaltsbereiche</i>	<i>Kognitive Anforderungsbereiche</i>
Arithmetik	Reproduzieren	Biologie	Reproduzieren
Geometrie/Messen	Anwenden	Physik/Chemie	Anwenden
Umgang mit Daten	Problemlösen	Geographie	Problemlösen

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Das für TIMSS 2011 von der internationalen Studienleitung in Kooperation mit Expertinnen und Experten aus verschiedenen Fachdisziplinen und Ländern entwickelte Kompetenzmodell bildet die Grundlage für die Entwicklung der in TIMSS 2011 eingesetzten Testaufgaben (vgl. Abschnitt 5; Martin & Mullis, 2011a, 2011d). Ein zentrales Kriterium für die Eignung von Testaufgaben ist hierbei, dass jede Aufgabe möglichst eindeutig einem bestimmten Inhaltsbereich und einem kognitiven Anforderungsniveau zuzuordnen ist. Die Gewichtung der für unterschiedliche Inhalts- und Anforderungsbereiche charakteristischen Aufgaben erfolgt nach einem vorgegebenen Schlüssel (Mullis et al., 2009; siehe auch Kapitel 3 und 4 in diesem Band).

4.1.3 Erfassung von Bedingungsfaktoren des schulischen Lernens und des Kompetenzerwerbs in Mathematik und Naturwissenschaften

Das Lernen von Schülerinnen und Schülern ist in verschiedene soziale Kontexte eingebettet. Während ihrer Grundschulzeit entwickeln Kinder vielfältige fachbezogene Kompetenzen und Einstellungen sowie ein individuelles Lernverhalten, welche allesamt durch ihre Erfahrungen in der Schule und in außerschulischen Kontexten geprägt sind. Je nach Umfeld werden Lernprozesse, die für die Ausbildung von fachspezifischen Kompetenzen relevant sind, durch unterschiedliche Ressourcen und Aktivitäten unterstützt. Zahlreiche individuelle Faktoren, aber auch Merkmale des Unterrichts sowie übergeordnete Merkmale der Schule als soziale Organisation beeinflussen die Kompetenzentwicklung. Für die Analyse und statistische Betrachtung von Unterschieden in der Schülerleistung sind vielerlei Bedingungsfaktoren und deren komplexe wechselseitige Verflechtungen zu berücksichtigen.

Mit TIMSS sollen nicht nur die Leistungsstände von Grundschülerinnen und Grundschulern am Ende der vierten Jahrgangsstufe untersucht, sondern darüber hinaus auch Faktoren identifiziert werden, die Leistungen und Einstellungen beeinflussen können. Um dies zu ermöglichen, werden – wie auch in vergleichbaren anderen Studien – die institutionellen und gesellschaftlichen Rahmenbedingungen, die die Bereitstellung und Nutzung schulischer Lerngelegenheiten beeinflussen, durch die standardisierte schriftliche Befragung der teilnehmenden Schülerinnen und Schüler, der unterrichtenden Lehrkräfte und der Schulleitungen erfasst. Ein Rückgriff auf diese Rahmendaten ermöglicht vertiefende Analysen zu den Zusammenhängen zwischen den Leistungsergebnissen und Hintergrundmerkmalen, die wertvolle Informationen und Erklärungsansätze ins-

besondere zu sozioökonomischen, kulturellen und sozialen Disparitäten der Schülerleistungen liefern.

Um ein möglichst umfassendes Bild der Rahmenbedingungen schulischen Lernens aufzuzeigen, werden im Rahmen von TIMSS zusätzlich zu den von den beteiligten Akteuren ausgefüllten Fragebögen mit Hilfe weiterer Instrumente ergänzende Informationen erfasst (Mullis et al., 2009). Die so gewonnenen Informationen sollen die Interpretation und Einordnung der Ergebnisse zusätzlich unterstützen. Das TIMSS-Rahmenmodell sieht vor, für einen internationalen Vergleich von Schülerleistungen neben den individuellen und familiären sowie den inner- und außerschulischen Lernkontexten auch Systemmerkmale zu berücksichtigen. Die standardisierte Beschreibung zentraler Aspekte des Bildungssystems ermöglicht es, die Leistungsergebnisse der Schülerinnen und Schüler in den einzelnen Teilnehmerstaaten auch unter Berücksichtigung dieser Faktoren zu interpretieren. Im Rahmen von TIMSS wird die internationale Berichterstattung zu den Leistungsergebnissen durch die sogenannte *Enzyklopädie* komplementiert (Mullis et al., 2012c). Diese Publikation beinhaltet für jeden Teilnehmerstaat eine Beschreibung des jeweiligen Schulsystems. Die Kapitel werden anhand einheitlicher Kriterien durch die nationalen Studienkoordinatoren verfasst. Die Schwerpunkte liegen dabei auf der Erläuterung der Organisation und Struktur des Bildungssystems und einer Beschreibung der fachspezifischen Curricula. Die Kapitel der Enzyklopädie werden durch tabellarische Darstellungen von demographischen Charakteristika der beteiligten Bildungssysteme, Inhalt und Aufbau der jeweiligen Curricula sowie Spezifika der Lehrerbildung ergänzt. Diese Informationen werden von der internationalen Studienleitung durch die Administration standardisierter Fragebögen (*Curriculum Questionnaire*) erfasst, die von Fachexpertinnen und Fachexperten der Teilnehmerstaaten bearbeitet werden.

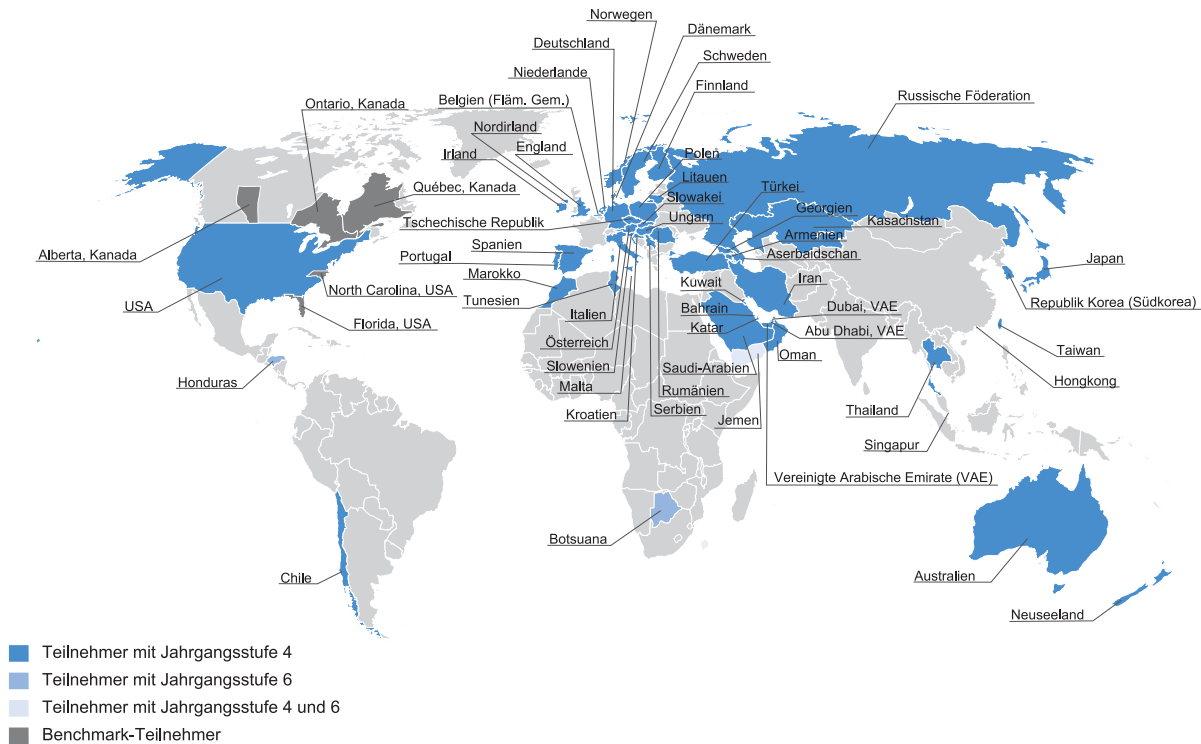
4.2 Zu den Teilnehmern der Studie

4.2.1 Teilnehmer 2011

An TIMSS 2011 haben sich weltweit 59 Bildungssysteme beteiligt (vgl. Abbildung 2.2), wobei es sich hierbei überwiegend um souveräne Staaten handelt. Hinzu kommen Regionen, die in der internationalen Berichterstattung als einzelne Teilnehmer aufgeführt werden, da sie für Bildungsfragen weitgehend selbstverantwortlich sind und auch an den vorangegangenen Studienzyklen separat teilgenommen haben (z.B. die Flämische Gemeinschaft in Belgien). Insgesamt nahmen 50 Staaten und Regionen als reguläre Teilnehmer mit der vierten Jahrgangsstufe (oder der national äquivalenten Klassenstufe) an der Studie teil (vgl. Abbildung 2.2). Zudem werden sieben Regionen als sogenannte Benchmark-Teilnehmer berichtet. Hierbei handelt es sich um Regionen, die ihre Ergebnisse auf der internationalen Leistungsskala verorten möchten, deren Leistungsdaten jedoch nicht in die Berechnung des internationalen Mittelwerts eingehen.

Darüber hinaus führten drei Staaten die Untersuchung auch oder ausschließlich in der sechsten Jahrgangsstufe durch. Die Erweiterung der Studie um die Möglichkeit, sich auch mit einer anderen Jahrgangsstufe zu beteiligen als die international definierte Zielpopulation vorsieht (vgl. Abschnitt 4.2.4), rich-

Abbildung 2.2: Staaten und Regionen, die an TIMSS 2011 mit der vierten Jahrgangsstufe teilnehmen



IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

tet sich an Staaten, die davon ausgehen, dass die mathematischen und naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler in der vierten Jahrgangsstufe nicht vergleichbar mit dem internationalen mittleren Niveau sind. Durch die Teilnahme mit der sechsten Jahrgangsstufe können diese Staaten ihr Leistungsniveau im internationalen Bezugsrahmen verorten.

In diesem Berichtsband werden die Ergebnisse der Schülerinnen und Schüler auf den Gesamtskalen für Mathematik und Naturwissenschaften für alle beteiligten Staaten, Regionen und Benchmark-Teilnehmer berichtet (vgl. Abbildung 3.5 und Abbildung 4.5). Um die Mittelwerte der Teilnehmer berichten zu können, die Mitglieder der EU und/oder der OECD sind, wurden wie auch im vorangegangenen Zyklus zwei Vergleichsgruppen (VG) gebildet, für die jeweils auch Mittelwerte berichtet werden: VG_{EU} und VG_{OECD} . Dies ermöglicht einen Vergleich Deutschlands mit jeweils einer Gruppe von Staaten mit ähnlichem kulturellen Hintergrund oder vergleichbarer wirtschaftlicher Situation. Welche Teilnehmer zu den Vergleichsgruppen zählen, ist Tabelle 2.2 zu entnehmen. Um einen für Deutschland spezifischen Fokus einzunehmen, werden in den meisten Abbildungen und Tabellen der einzelnen Kapitel lediglich die Ergebnisse für diejenigen Teilnehmerstaaten und Regionen berichtet, die regulär mit der vierten Jahrgangsstufe teilgenommen haben und für die zudem mindestens eines der folgenden Kriterien gilt: (1) Teilnehmer, die Mitglieder der EU sind, (2) Teilnehmer, die der OECD angehören, (3) Teilnehmer, die auf der jeweiligen Gesamtskala (vgl. Abbildung 3.5 oder Abbildung 4.5) signifikant bessere Ergebnisse erzielt haben als Deutschland.

Tabelle 2.2: Vergleichsgruppen in TIMSS 2011

VG _{EU}		VG _{OECD}	
Belgien (Fläm. Gem.)	Portugal	Australien	Norwegen
Dänemark	Rumänien	Belgien (Fläm. Gem.)	Österreich
Deutschland	Schweden	Chile	Polen
England	Slowakei	Dänemark	Portugal
Finnland	Slowenien	Deutschland	Republik Korea (Südkorea)
Irland	Spanien	England	Schweden
Italien	Tschechische Republik	Finnland	Slowakei
Litauen	Ungarn	Irland	Slowenien
Malta		Italien	Spanien
Niederlande		Japan	Tschechische Republik
Nordirland		Neuseeland	Türkei
Österreich		Niederlande	Ungarn
Polen		Nordirland	USA
21		26	

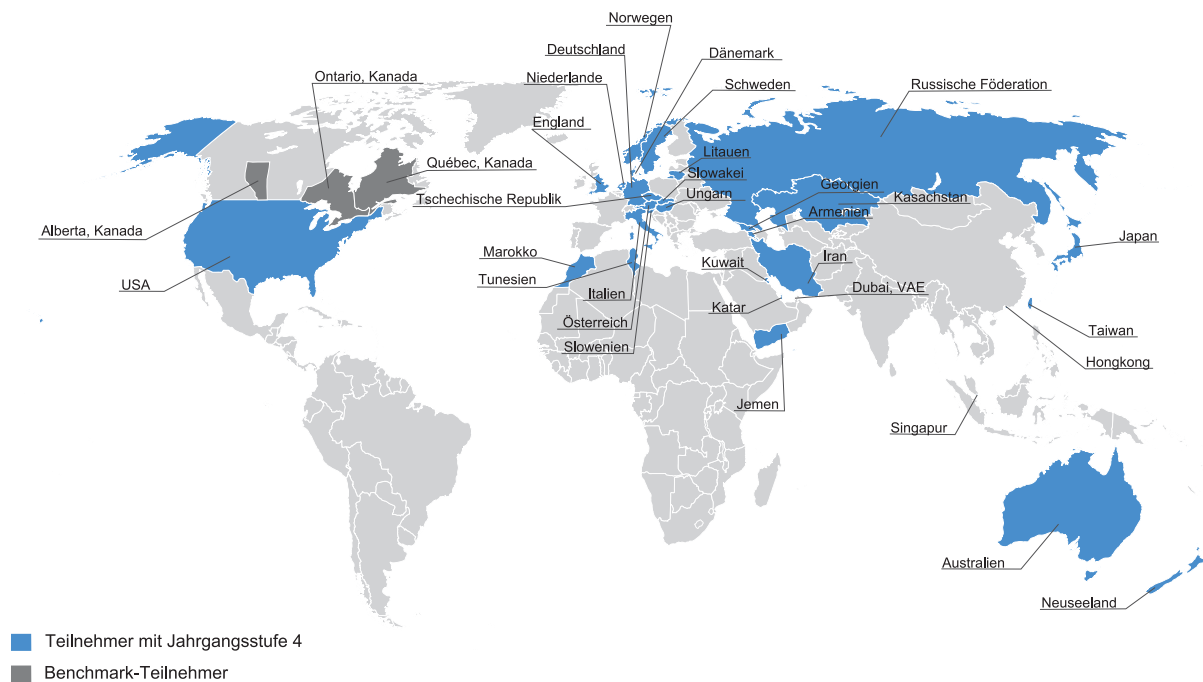
IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

4.2.2 Teilnehmer 2007 und 2011

Wie Abbildung 2.3 zu entnehmen ist, haben sich 30 Teilnehmer der TIMS-Studie 2011 (Grundschule) sowie 4 Benchmark-Teilnehmer bereits am Studienzyklus 2007 mit der vierten Jahrgangsstufe beteiligt.

Abbildung 2.3: Teilnehmer an TIMSS 2007 und TIMSS 2011



IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

4.2.3 Teilnahmebedingungen der Studie

Sinnvolle Vergleiche von Bildungssystemen und ihren Erträgen setzen die Einhaltung strenger methodischer Standards voraus. Hierzu gehören auch eindeutige Vorgaben, unter welchen Bedingungen und mit welchen Schülerinnen und Schülern sich Bildungssysteme an der Studie beteiligen dürfen (siehe Mullis et al., 2009). Im Folgenden werden die internationalen Teilnahmebedingungen für TIMSS 2011 aufgezeigt.

4.2.4 Definition der Untersuchungspopulation

TIMSS hat zum Ziel, schulisches Lernen zu verbessern, und möchte dazu Empfehlungen für bildungspolitische Bestimmungen und schulische Praktiken aussprechen. Da diese oft klassenstufenspezifisch sind, wird mit TIMSS angestrebt, jeweils komplette Schülerjahrgänge zu vergleichen. Für die Definition der zu untersuchenden Schülerpopulation (*Zielpopulation*) steht dabei zunächst das Kriterium der formalen Beschulungszeit im Vordergrund:

Zu der für alle Teilnehmerstaaten verbindlich festgelegten Zielpopulation zählen alle Schülerinnen und Schüler, die sich nach der *International Standard Classification of Education* im vierten Jahr formaler Beschulung befinden (ISCED Level 1; vgl. UIS, 2006). In den meisten Teilnehmerstaaten trifft dies auf die Schülerinnen und Schüler der Jahrgangsstufe 4 zu. Um eine entwicklungsgerechte Passung der Leistungstests und der Durchführungsbedingungen der Tests sicherzustellen, wird darüber hinaus ein Alterskriterium gesetzt, nach dem in jedem Teilnehmerstaat das durchschnittliche Alter der Zielpopulation zum Testzeitpunkt mindestens 9 Jahre und 6 Monate betragen soll (Mullis et al., 2009).

Ausschöpfung der Zielpopulation

Um internationale Vergleichbarkeit zu ermöglichen, muss die international definierte Zielpopulation von allen Teilnehmerstaaten möglichst vollständig ausgeschöpft werden. Dies ist dann gegeben, wenn alle Schülerinnen und Schüler, die unter die Definition der internationalen Zielpopulation fallen, potentiell an der Studie teilnehmen können, also für die zu untersuchende Stichprobe ausgewählt werden können. Ausgehend von der international definierten Zielpopulation muss jeder Teilnehmer die international vorgegebenen Richtlinien zur Stichprobenziehung an das eigene System anpassen. Grundlage hierfür ist ein Handbuch zur Stichprobenziehung. Zudem steht umfassende Unterstützung durch Expertinnen und Experten von *Statistics Canada* und dem DPC zur Verfügung, die von der internationalen Studienleitung unter anderem mit der Stichprobenziehung beauftragt wurden.

Vor dem Hintergrund nationaler Besonderheiten kann es jedoch einzelnen Teilnehmern nicht möglich sein, den internationalen Vorgaben zu entsprechen. In fast jedem Schulsystem gibt es einzelne Schulen oder Schülergruppen, die aus verschiedenen Gründen nicht an der Erhebung teilnehmen können oder teilnehmen wollen. Innerhalb der nationalen Zielpopulation kann jeder Teilnehmerstaat einen geringen Prozentsatz an Schülerinnen und Schülern beziehungsweise an Schulen festlegen, der bei der Stichprobenziehung ausgeschlossen wird. Die internationale Studienleitung legt für *Ausschlüsse von der Zielpopulation* eindeutige Kriterien fest:

Auf *Schülerebene* dürfen Schülerinnen und Schüler ausgeschlossen werden, die aus körperlichen, emotionalen oder geistigen Gründen nicht in der Lage sind, den Test selbstständig zu bearbeiten, sowie Schülerinnen und Schüler, die weniger als ein Jahr in der Testsprache unterrichtet wurden und deren Muttersprache nicht die Testsprache ist.

Auf *Schulebene* können Schulen ausgeschlossen werden, die in besonders schwer zugänglichen Regionen liegen, die weniger als vier Schülerinnen und Schüler im Jahrgang 4 haben, die in Bezug auf den Lehrplan oder die Schulstruktur vom nationalen Schulsystem abweichen oder die ausschließlich Schülerinnen und Schüler der auf Schülerebene genannten Ausschlusskriterien unterrichten. Ausschlüsse auf Schul- und Schülerebene dürfen insgesamt nicht mehr als 5 Prozent der Zielpopulation betragen (Joncas & Foy, 2011).

Nachdem alle Ausschlussquellen in Betracht gezogen wurden, lässt sich die *national effektiv erreichte Zielpopulation (national effective target population)* für jeden Teilnehmer beschreiben. Diese verweist darauf, welche Schülerinnen und Schüler einer Population durch die Stichprobe der teilnehmenden Schülerinnen und Schüler tatsächlich repräsentiert werden. Auf Grundlage dieser Beschreibung werden alle Abweichungen von der international definierten Zielpopulation in den internationalen Berichten genau dokumentiert und bedeutsame Abweichungen besonders hervorgehoben.

4.3 Verfahren und Kriterien der Stichprobenziehung

Die Stichprobenziehung in TIMSS erfolgt nach einem sogenannten zweistufigen stratifizierten Clusterdesign (Joncas & Foy, 2011). Der erste Sampling-Schritt besteht aus einer Zufallsauswahl aus allen Schulen der national effektiv erreichten Zielpopulation; der zweite Schritt umfasst die Ermittlung einer oder mehrerer Klassen in den ausgewählten Schulen.

Auf der *ersten Stufe* wird aus einer Liste (*Sampling Frame*), in der alle Schulen aufgeführt sind, die Teil der national effektiv erreichten Zielpopulation sind, eine Zufallsstichprobe von Schulen gezogen. Nach den Vorgaben der internationalen Studienleitung beträgt die Mindestanzahl 150 Schulen. Da Schulen mit einer größeren Anzahl an Schülerinnen und Schülern in der vierten Jahrgangsstufe auch einen höheren Anteil an der Population haben, wird die Information zur Größe der Jahrgangsstufe 4 systematisch bei der Stichprobenziehung berücksichtigt. Um die Effizienz der Stichprobe und die Präzision der Ergebnisse zu steigern, kommt zudem die Methode der Stratifizierung zur Anwendung: Alle Schulen, die als Teil der national effektiv erreichten Zielpopulation gelten, werden nach bestimmten Merkmalen (*Strata*) kategorisiert (*geschichtet*). Mögliche Strata sind beispielsweise die geographische Lage (z.B. Bundesländer) oder der Schultyp (z.B. Förder- vs. Regelschule, privat vs. öffentlich). Diese zur Verfügung stehenden Informationen über die Schulen können nun bei der Stichprobenziehung berücksichtigt werden, so dass sichergestellt wird, dass sich die tatsächliche Verteilung der Schulen innerhalb dieser Strata auch in der Stichprobe widerspiegelt und dadurch spezifische Gruppen der Population adäquat repräsentiert werden.

Auf der *zweiten Stufe* werden innerhalb der ausgewählten Schulen Klassenstichproben gezogen. Dabei hat jede Klasse dieselbe Wahrscheinlichkeit, in die Stichprobe zu gelangen. Mit Auswahl der Klassen sind automatisch auch die Schülerinnen und Schüler bestimmt, da – abgesehen von oben beschriebenen

Ausnahmen – alle Schülerinnen und Schüler einer Klasse am Test teilnehmen sollen. Laut Vorgaben der internationalen Studienleitung sind mindestens 4000 Schülerinnen und Schüler auszuwählen (Joncas & Foy, 2011).

4.3.1 Schul- und Schülerteilnahmequoten

Wie in jeder Studie, so ist auch in TIMSS mit Stichprobenausfällen zu rechnen. Um das Problem von Datenverzerrungen aufgrund nicht teilnehmender Schulen oder Personengruppen möglichst gering zu halten, macht die internationale Studienleitung Vorgaben für minimale Beteiligungsquoten (Joncas & Foy, 2011). In der internationalen Berichterstattung werden Teilnehmerstaaten und Regionen, die die vorgeschriebenen Rücklaufquoten nicht erreichen, bei weniger gravierenden Abweichungen mit einer Fußnote und einer entsprechenden Erklärung in den Ländervergleich einbezogen. Sollten sich Rücklaufquoten als zu gering erweisen, werden die Ergebnisse gesondert berichtet.

Auf *Schulebene* ist vorgesehen, dass mindestens 85 Prozent der ursprünglich ausgewählten Schulen an der Erhebung teilnehmen müssen. Fehlen weniger als 15 Prozent der ursprünglich ausgewählten Schulen, können nicht teilnehmende Schulen durch die Heranziehung von nachrückenden Schulen ersetzt werden, so dass die Anforderungen der Stichprobengröße insgesamt erfüllt werden können. Ersatzschulen sind den ursprünglich gezogenen Schulen in den für die Stratifizierung bestimmten Merkmalen möglichst ähnlich.

Auf *Klassenebene* ist eine Teilnahme an der Erhebung von mindestens 95 Prozent der ursprünglich ausgewählten Klassen vorgesehen. Nicht teilnehmende Klassen können nicht durch Parallelklassen ersetzt werden, vielmehr müssen auch für sie Ersatzschulen gefunden werden. Auch für die Befragung der Lehrkräfte wird eine Mindestbeteiligung von 85 Prozent vorgegeben. Wie im Falle von Klassen dürfen auch Lehrkräfte nicht ersetzt werden (Joncas & Foy, 2011).

Für die *Schülerebene* ist festgelegt, dass innerhalb der teilnehmenden Schulen mindestens eine 85-prozentige Beteiligung der Schülerschaft vorliegen muss, wobei die Beteiligung in den einzelnen Klassen nicht unter 50 Prozent sinken darf. Um die geforderten Rücklaufquoten zu erreichen, wird an Schulen, an denen mehr als 10 Prozent der Schülerinnen und Schüler am Testtag nicht anwesend sind, ein Nachttest durchgeführt. Wird trotz Nachttest die Mindestrücklaufquote nicht erreicht, wird die Klasse und somit die ganze Schule ausgeschlossen. Übergeordnet gilt zudem, dass die kombinierte Schul- und Schülerteilnahmequote nicht unter 75 Prozent sinken sollte.

Als weiteres Kriterium zur Bewertung der Qualität der Stichprobe auf Schülerebene wird darüber hinaus berücksichtigt, inwieweit die Teilnahme an der Leistungsmessung tatsächlich erfolgreich war. Hierbei ist als Grundvoraussetzung für eine Interpretation der Kompetenzmittelwerte vorgesehen, dass mindestens 85 Prozent der Schülerinnen und Schüler innerhalb der Stichprobe über skalierbare Leistungswerte verfügen. Als Minimal Kriterium für die Skalierung der Leistungstests gilt hier, dass eine Mindestanzahl an Testaufgaben und von diesen mindestens eine erfolgreich bearbeitet worden sein muss.

4.3.2 Besonderheiten der Stichproben im internationalen Vergleich

Die Tabelle A.2 (im Anhang) bietet einen Überblick über die Stichproben aller Teilnehmer an TIMSS 2011 und listet Besonderheiten bezüglich der nationalen Zielpopulationen, der Schul- und Schülerteilnahmequoten sowie der Leistungsmessungen auf. Benchmark-Teilnehmer sowie Teilnehmer mit der sechsten Jahrgangsstufe (Botsuana, Honduras und Jemen) werden gesondert aufgeführt. In Tabelle A.1 (im Anhang) sind diese Besonderheiten entsprechend für TIMSS 2007 angegeben. Auf diese Besonderheiten der Stichproben, aus denen für eine international-vergleichende Interpretation von Ergebnissen Einschränkungen der Vergleichbarkeit resultieren können, weisen zusätzlich die Fußnoten in der ersten Tabellenspalte hin (vgl. Abschnitt 4.3.4).

Der dritten Tabellenspalte ist zu entnehmen, welche Teilnehmer von TIMSS 2011 bereits am Studienzyklus 2007 beteiligt waren. Die Tabellenspalten 4 bis 7 verweisen auf Besonderheiten bezüglich der Definition und Ausschöpfung der nationalen Zielpopulation. Die weiteren Tabellenspalten 8 bis 14 weisen auf Besonderheiten bezüglich der *Schul- und Schülerteilnahmequoten* hin.

Wie aus der Spalte 4 (*Getestete Jahrgangsstufe*) deutlich wird, wählten Neuseeland, Malta und England als äquivalente nationale Zielpopulation gemäß der Kriterien ‚formale Beschulungszeit‘ und ‚Durchschnittsalter des Schülerjahrgangs‘ eine höhere Jahrgangsstufe. Das Durchschnittsalter der getesteten Schülerinnen und Schüler ist der fünften Tabellenspalte zu entnehmen.

In Spalte 6 wird der *Ausschöpfungsgrad der nationalen Zielpopulation in Prozent bezogen auf die internationale Vorgabe* (100%) illustriert. Hier zeigt sich für drei Teilnehmer und zwei Benchmark-Teilnehmer, dass die internationale Zielpopulation nicht zu 100 Prozent durch die nationale Zielpopulation abgedeckt wird. Für Georgien und Litauen ist der geringe Ausschöpfungsgrad dadurch bedingt, dass hier nur Schülerinnen und Schüler getestet wurden, die in Georgisch beziehungsweise Litauisch unterrichtet werden. In Kuwait, Florida und North Carolina nahmen nur diejenigen Schülerinnen und Schüler an der Erhebung teil, die staatliche Schulen besuchen.

In Spalte 7 werden die *Ausschlüsse von der Zielpopulation* als Gesamtausschlussquote in Prozent aufgeführt. Hier zeigt sich, dass die internationale Vorgabe von maximal 5 Prozent von 13 Teilnehmern und fünf Benchmark-Teilnehmern überschritten wird. Deutschland (vgl. Abschnitt 4.3.3) gehört zu den acht von insgesamt 60 Teilnehmern, deren Gesamtausschlussquote auf Ebene der nationalen Zielpopulation unter 2 Prozent liegt.

Die *Schulteilnahmequoten* (ohne und mit Ersatzschulen), die *Schülerteilnahmequoten* sowie die *Gesamtteilnahmequoten* (ohne und mit Ersatzschulen) der einzelnen Teilnehmer können den Tabellenspalten 8 bis 12 entnommen werden. Während eine Vielzahl der Teilnehmer die internationalen Kriterien erfüllt, zeigt sich für Aserbaidschan, die Flämische Gemeinschaft in Belgien, Dänemark, Italien, Neuseeland und Nordirland, dass sie die *Schulteilnahmequote* von 85 Prozent nur unter Hinzunahme der nachrückenden Schulen erreichen. England, den Niederlanden, Norwegen und Ungarn hingegen gelingt dies auch mit Berücksichtigung der Ersatzschulen nicht. Die *Schülerteilnahmequote* von 85 Prozent erreichen alle Teilnehmer. Die *kombinierte Schüler- und Schulgesamttteilnahmequote* von 75 Prozent erreichen die Niederlande und Nordirland nur unter Hinzunahme der Ersatzschulen. Norwegen liegt auch mit Ersatzschulen noch unterhalb der Schüler- und Schulgesamttteilnahmequote.

In den letzten beiden Tabellenspalten ist für TIMSS 2011 der Anteil der Schülerinnen und Schüler ohne skalierbare Leistungswerte in den Naturwissenschaften (Tabellenspalte 13) und in Mathematik (Tabellenspalte 14) angegeben. Da diese Angaben erst seit der Erhebung im Jahr 2011 für TIMSS berichtet werden, finden sich diese Tabellenspalten nicht in den Tabellen zum vorangegangenen Studienzyklus. Hier weisen Jemen, Kuwait, Marokko und Tunesien einen hohen Anteil von Schülerinnen und Schülern ohne auswertbare Angaben in den Leistungstests sowohl in den Naturwissenschaften als auch in Mathematik auf. In Oman sowie in Botsuana und Honduras, die alle mit der sechsten Jahrgangsstufe an TIMSS 2011 teilgenommen haben, liegt der Anteil der Schülerinnen und Schüler ohne gültige Leistungswerte in Mathematik über 15 Prozent.

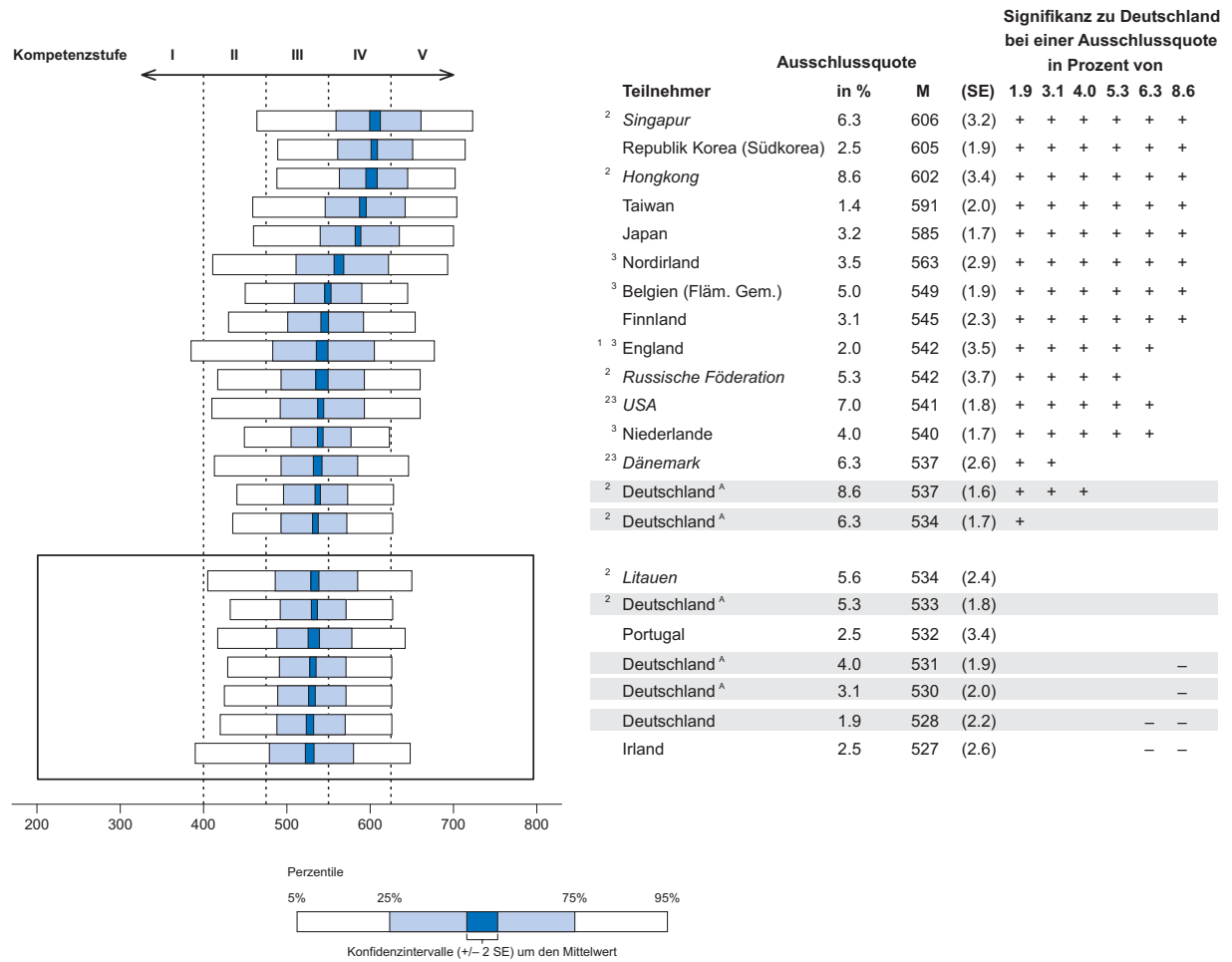
4.3.3 Bedeutsamkeit der Ausschlussquote für den internationalen Vergleich

Wie im vorangegangenen Abschnitt beschrieben sind die Ausschlüsse von der Zielpopulation für Deutschland mit einer Gesamtausschlussquote von 1.9 Prozent im internationalen Vergleich sehr gering, während andere Teilnehmer wie Hongkong (8.6%), die USA (7.0%), Dänemark (6.3%), Singapur (6.3%) oder Litauen (5.6%) Ausschlussquoten erreichen, die über der internationalen Vorgabe liegen.

Unterstellt man, dass Ausschlüsse von der Zielpopulation in erster Linie Schülerinnen und Schüler mit eher geringen Kompetenzen betreffen, so stellt sich die Frage, welche Mittelwerte die Viertklässlerinnen und Viertklässler in Deutschland bei vergleichbaren Ausschlussquoten erzielt hätten. In der Abbildung 2.4 sind unterschiedliche Verortungen Deutschlands auf der Gesamtskala Mathematik (vgl. Abbildung 3.5) dargestellt, die sich ergeben würden, wenn für Deutschland Ausschlussraten zugrunde gelegt würden, die vergleichbar sind mit denen ausgewählter Teilnehmer, welche die Vorgabe von 5 Prozent überschreiten. Um die Abbildung nicht zu überfrachten werden ausgewählte Ausschlussquoten illustriert. Als Referenzpunkte sind zudem in der Abbildung zunächst die realen Kompetenzmittelwerte sowohl für Deutschland als auch für diejenigen Teilnehmerstaaten angegeben, die auf der Gesamtskala signifikant besser oder nicht signifikant different zu Deutschland abschneiden. Der rechten Seite der Abbildung ist zu entnehmen, ob die Leistungsmittelwerte der dargestellten Teilnehmer unter Berücksichtigung der unterschiedlichen Ausschlussquoten signifikant besser oder schlechter abschneiden als die Leistungsmittelwerte für Deutschland. Signifikant höhere Werte sind durch ein Plus (+), signifikant niedrigere Werte durch ein Minus (-) symbolisiert.

In Abbildung 2.4 ist die Veränderung des Mittelwerts der Viertklässlerinnen und Viertklässler in Deutschland beim Vergleich der Leistung auf der Gesamtskala Mathematik in Abhängigkeit von der Ausschlussquote dargestellt. Aus der Abbildung wird ersichtlich, dass eine Erhöhung der Ausschlussquoten für Deutschland im Rahmen der internationalen Vorgaben (max. 5.49%) keinen bedeutsamen Einfluss auf das Abschneiden der Viertklässlerinnen und Viertklässler in Deutschland im internationalen Vergleich hätte. Erst Ausschlussquoten von 8.6 Prozent (wie in Hongkong), 7.0 Prozent (wie in den USA) oder 6.3 Prozent (wie in Singapur und Dänemark) – die jedoch auch mit einer entsprechenden Fußnote im internationalen Berichtsband gekennzeichnet werden würden – würden den Leistungsmittelwert der Schülerinnen und Schüler in Deutschland und

Abbildung 2.4: Veränderung des Mittelwerts von Viertklässlerinnen und Viertklässlern in Deutschland beim Vergleich der Leistung auf der Gesamtskala Mathematik in Abhängigkeit von der Ausschlussquote



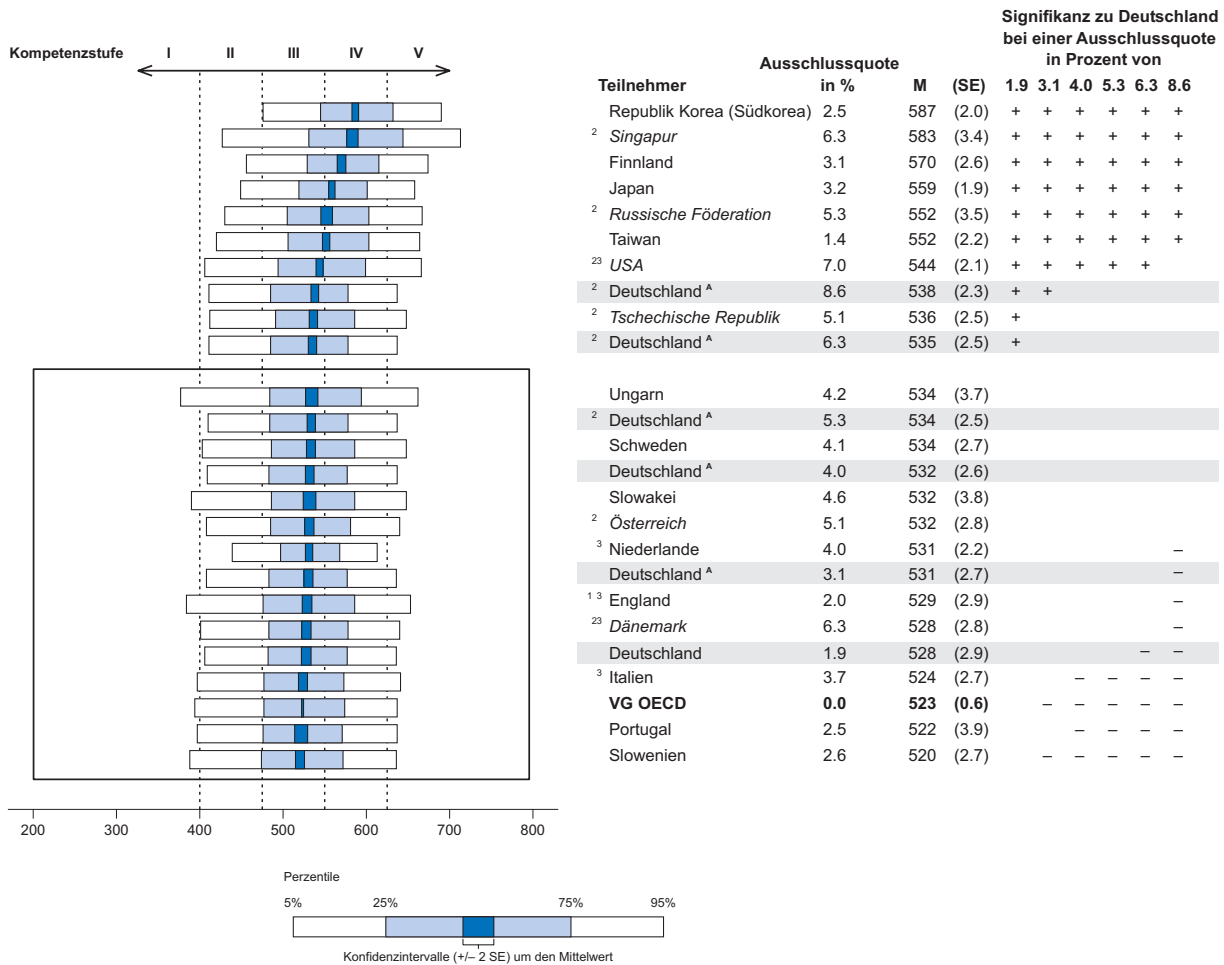
□ Nicht statistisch signifikant vom tatsächlichen deutschen Mittelwert abweichende Staaten ($p > .05$).
 Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 A = Berechnung für Deutschland nach einem höheren Ausschluss von Schülerinnen und Schülern mit geringen Kompetenzwerten.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

damit auch die Platzierung Deutschlands im Ländervergleich signifikant positiv beeinflussen. Umgekehrt ist auch zu vermuten, dass die Leistungsmittelwerte der Teilnehmer Hongkong, USA, Singapur und Dänemark bei geringeren Ausschlussquoten wahrscheinlich unterhalb der für diese Teilnehmer gefundenen Leistungsmittelwerte liegen würden. Festzustellen bleibt, dass die Ausschlussquoten anderer Teilnehmer nur begrenzt als Begründung für das Abschneiden Deutschlands im internationalen Vergleich herangezogen werden können. Selbst unter Ausschluss von 8.6 Prozent der leistungsschwächsten Viertklässlerinnen und Viertklässler aus der deutschen Stichprobe würden die Schülerinnen und Schüler in der Republik Korea, in Taiwan, Japan, Nordirland, der Flämischen Gemeinschaft in Belgien, Singapur und Hongkong weiterhin signifikant bessere Leistungsergebnisse erzielen. Ein vergleichbarer Befund lässt sich für die Naturwissenschaften festhalten (vgl. Abbildung 2.5).

Abbildung 2.5: Veränderung des Mittelwerts von Viertklässlerinnen und Viertklässlern in Deutschland beim Vergleich der Leistung auf der Gesamtskala Naturwissenschaften in Abhängigkeit von der Ausschlussquote



□ Nicht statistisch signifikant vom tatsächlichen deutschen Mittelwert abweichende Staaten ($p > .05$).
 Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3= Die Teilnahmequoten auf Schul- und/oder Schülerbene erreichen nicht die internationalen Vorgaben.
 A = Berechnung für Deutschland nach einem höheren Ausschluss von Schülerinnen und Schülern mit geringen Kompetenzwerten.

4.3.4 TIMSS 2011 Fußnotensystem zur Klassifikation von Teilnahmebedingungen

In den Tabellen A.1 und A.2 (im Anhang) wird für jeden Teilnehmer detailliert auf Besonderheiten der Teilnahmebedingungen verwiesen.

Für diesen Berichtsband wurde für die Darstellung von Ergebnissen in Abbildungen und Tabellen in Anlehnung an die internationale Berichterstattung ein ausdifferenziertes Fußnotensystem entwickelt, das auf beide TIMSS-Erhebungen angewendet werden kann. Die Fußnoten weisen auf Besonderheiten der Stichproben einzelner Teilnehmer hin. Damit wird eine aus wissenschaftlicher Sicht wünschenswerte Grundlage geboten, um für eine international-vergleichende Interpretation von Ergebnissen Einschränkungen der Vergleichbarkeit, die aus Besonderheiten der nationalen Stichproben resultieren, stets im Blick

behalten zu können. Folgende Fußnoten werden in diesem Berichtsband zur Kennzeichnung von Besonderheiten von nationalen Stichproben verwendet:

1	=	Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
2	=	Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
3	=	Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
4	=	Sehr hoher Anteil an Schülerinnen und Schülern mit nicht skalierbaren Leistungswerten.

Die Fußnote 4 ist nur für den Studienzyklus 2011 relevant und wird, sofern eine Unterscheidung notwendig ist, fachspezifisch in den Fußnoten 4 (nicht skalierbare Leistungswerte in Mathematik) und 5 (nicht skalierbare Leistungswerte in den Naturwissenschaften) differenziert.

Teilnehmer, die eine Fußnote 4 erhalten, werden nur einmal in den Gesamtskalen berichtet und dann von allen weiteren Darstellungen im nationalen Berichtsband ausgespart.

Für den Vergleich zu TIMSS 2007 wird zudem folgende Fußnote ergänzt:

6	=	Abweichender Testzeitpunkt (In Kuwait und Dubai, VAE erfolgte die Testung zeitlich verzögert.)
---	---	--

Kursivschreibung der Ländernamen und Kennzahlen

Ergänzend zu den Fußnoten wird durch Kursivschreibung der Staatennamen auf eine eingeschränkte Vergleichbarkeit hingewiesen, wenn für Teilnehmer mindestens eines der folgenden Kriterien zutrifft:

- die Gesamtausschlussquote liegt über der internationalen Vorgabe von 5 Prozent (vgl. Tabelle A.2; Spalte 7; *Ausschlüsse von der nationalen Zielpopulation (Gesamtquote) in %*).
- die Schüler- und Schulgesamtteilnahmequote (mit Ersatzschulen) liegt unter 75 Prozent (vgl. Tabelle A.2; Spalte 12; *Gesamtteilnahmequote in %*).
- der Anteil der Schülerinnen und Schüler ohne gültige Leistungswerte liegt über 15 Prozent (vgl. Fußnote 4 und 5).

In den Abbildungen und Tabellen werden kursive Formatierungen der Staatennamen durch folgende Fußnote erläutert:

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

5 Entwicklung und Charakteristika der Instrumente

In TIMSS kommen zwei Arten von Erhebungsinstrumenten zum Einsatz: Zur Ermittlung von Schülerkompetenzen in Mathematik und den Naturwissenschaften werden Leistungstests eingesetzt. Zur Erfassung umfangreicher Hintergrundinformationen werden Fragebögen auf System-, Schul-, Lehrer- und Schülersebene eingesetzt. In Teilnehmerstaaten und Regionen, die mit derselben Schülerstichprobe an IGLU/PIRLS 2011 teilgenommen haben, wird zudem ein Elternfragebogen administriert.

5.1 Leistungstests

Die in TIMSS eingesetzten Leistungstests dienen der Ermittlung von Leistungsständen der Schülerinnen und Schüler in den Bereichen Mathematik und Naturwissenschaften mit dem Ziel, im internationalen Vergleich Aussagen über Leistungsunterschiede zwischen gesamten Schülerpopulationen sowie zwischen spezifischen Gruppen von Schülerinnen und Schülern für die übergreifenden Kompetenzskalen und die Subdomänen treffen zu können. Ein studienzyklenspezifischer Aufgabenpool umfasst rund 170 Aufgaben pro Kompetenzdomäne. In der Regel werden rund 60 Prozent der Aufgabenblöcke zum Zwecke der Verlinkung aus den vorangegangenen Studienzyklen übernommen, wobei davon rund 75 Prozent dieser Aufgaben aus dem jeweils letzten und gut 25 Prozent aus dem jeweils vorletzten Zyklus stammen (Martin & Mullis, 2011f). Gut 40 Prozent der Blöcke werden somit neu entwickelt. Die Entwicklung der Testaufgaben erfolgt in einem kooperativen Prozess der TIMSS-Teilnehmerstaaten. Die Aufgabenvorschläge stammen aus verschiedenen Teilnehmerstaaten und werden von Expertinnen und Experten aus verschiedenen Fachdisziplinen und Ländern begutachtet. Ein maßgebliches Kriterium bei der Entwicklung von TIMSS-Testaufgaben besteht darin, dass die Aufgaben repräsentativ für die in der Rahmenkonzeption formulierten Inhalte und Anforderungen sind. Bei diesem Überprüfungsprozess wird insbesondere sichergestellt, dass die Aufgaben die Testziele der TIMSS-Rahmenkonzeption repräsentieren (Mullis et al., 2011). Im Rahmen nationaler Feldtests werden ausgewählte Testaufgaben dann auf ihre empirische Qualität hin überprüft und es wird sichergestellt, dass die Testaufgaben für die Schülerinnen und Schüler aller Teilnehmerstaaten vergleichbare Messeigenschaften aufweisen (Martin & Mullis, 2011b).

5.1.1 Charakteristika der Testaufgaben

Alle Testaufgaben sind kindgerecht aufbereitet und zum Teil ansprechend illustriert. Etwas mehr als die Hälfte der Testaufgaben wird im *Multiple-Choice*-Format gehalten, bei dem die Schülerinnen und Schüler aus zwei, vier oder fünf vorgegebenen Antworten die richtige Lösung auswählen müssen. Etwas weniger als die Hälfte der Aufgaben ist durch ein offenes Antwortformat gekennzeichnet, welches den Schülerinnen und Schülern die Möglichkeit gibt, ihre Antworten auf spezifische Testaufgaben frei zu formulieren.

5.1.2 Rotation der Testaufgaben

In den Testheften, die die Schülerinnen und Schüler während der Testsitzung erhalten und bearbeiten, sind jeweils sowohl Mathematik- als auch Naturwissenschaftsaufgaben enthalten. Die Zusammenstellung der Testhefte erfolgt über domänenspezifische Blöcke von Aufgaben. Jeder Aufgabenblock umfasst zwischen 10 und 15 Einzelaufgaben. Die einzelnen Testaufgaben können unabhängig voneinander bearbeitet werden. Um die teilnehmenden Kinder nicht übermäßig zu belasten und dennoch einen umfassenden Einblick in ihre fachspezifischen Kompetenzen zu erhalten, liegt der Testkonzeption ein multiples Matrixdesign zugrunde, welches eine Rotation von 28 verschiedenen Testblöcken (jeweils 14 für Mathematik und die Naturwissenschaften) über 14 unterschiedliche Testhefte vorsieht (vgl. Tabelle 2.3). Mit diesem Verfahren bearbeitet jede Schülerin und jeder Schüler lediglich vier ausgewählte Aufgabenblöcke (zwei für Mathematik und zwei für Naturwissenschaften), gleichzeitig können aber aufgrund der ausreichend großen Stichprobe die Leistungswerte für die Population präzise geschätzt werden. Durch das Rotationsdesign wird auch sichergestellt, dass gut die Hälfte der Schülerinnen und Schüler zuerst zwei Aufgabenblöcke aus dem Kompetenzbereich Mathematik bearbeitet und im Anschluss die Aufgabenblöcke für den Bereich Naturwissenschaften, während für die andere Hälfte die Administration der domänenspezifischen Testblöcke in umgekehrter Reihenfolge erfolgt (Mullis et al., 2009).

Tabelle 2.3: Testheftdesign in TIMSS 2011

Dauer in Min.		Testheft													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
18	Teil 1	M 01	S 02	M 03	S 04	M 05	S 06	M 07	S 08	M 09	S 10	M 11	S 12	M 13	S 14
18		M 02	S 03	M 04	S 05	M 06	S 07	M 08	S 09	M 10	S 11	M 12	S 13	M 14	S 01
18	Teil 2	S 01	M 02	S 03	M 04	S 05	M 06	S 07	M 08	S 09	M 10	S 11	M 12	S 13	M 14
18		S 02	M 03	S 04	M 05	S 06	M 07	S 08	M 09	S 10	M 11	S 12	M 13	S 14	M 01

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

5.1.3 Curriculare Validität der Testaufgaben

Obwohl die Rahmenkonzeption von allen Teilnehmerstaaten verabschiedet wurde, ist es unvermeidlich, dass der Grad der Übereinstimmung zwischen den TIMSS-Tests und den national spezifischen Curricula von Staat zu Staat variiert. Eine Einschränkung der Tests auf Aufgaben, die ausnahmslos in allen Staaten curricular valide sind, hätte eine starke Einschränkung der Testinhalte und somit auch der bearbeitbaren Forschungsfragen zur Folge. TIMSS verzichtet daher auf eine Testreduktion, führt aber traditionsgemäß seit Beginn der TIMSS-Zyklen im Jahr 1995 eine sogenannte *Test-Curriculum Matching Analysis* (TCMA) durch. Alle Teilnehmerstaaten beurteilen hierzu jede einzelne Aufgabe hinsichtlich ihrer Validität für das nationale Curriculum. Die Ergebnisse der Beurteilung für Deutschland sind in den Kapiteln 3 und 4 beschrieben. Von Seiten der internationalen Studienleitung werden diese Informationen genutzt, um für jeden

Teilnehmerstaat vergleichend zu ermitteln, wie seine Leistungsergebnisse (dargestellt als durchschnittliche Lösungswahrscheinlichkeit) im Vergleich zu den Ergebnissen aller anderen Teilnehmer ausfallen, wenn nur jene Aufgaben berücksichtigt werden, die nach Experteneinschätzung Teil des eigenen Curriculums sind. Dieser multiple Vergleich der Leistungsergebnisse ist Teil der internationalen Ergebnisberichte (Mullis et al., 2009). Die Ergebnisse dieser Analysen zeigen, dass alle Staaten in dem für ihr jeweiliges Curriculum validen Subtest eine etwas höhere durchschnittliche Lösungswahrscheinlichkeit als im Gesamttest erreichen. Allerdings beträgt die Differenz bei den meisten Staaten nur ein bis zwei Prozentpunkte. Die Ergebnisse für Deutschland sind für Mathematik in Tabelle 3.11 und für die Naturwissenschaften in Tabelle 4.9 dargestellt. Die TIMSS-Aufgaben erlauben somit einen angemessenen und fairen Vergleich zwischen den Teilnehmerstaaten (Martin, Mullis, Foy & Stanco, 2012; Mullis, Martin, Foy & Arora, 2012a).

5.1.4 Freigegebene Aufgaben

Um der Öffentlichkeit Informationen über Art und Inhalt der TIMSS-Testaufgaben zugänglich zu machen, werden im Februar 2013 von der internationalen Studienleitung zwölf Aufgabenblöcke der TIMSS-2011-Testhefte veröffentlicht. Dabei handelt es sich um 6 der 14 Testblöcke für den Bereich Mathematik sowie um 6 der 14 Testblöcke für den Bereich Naturwissenschaften. Jeweils drei der sechs Aufgabenblöcke enthalten Aufgaben aus TIMSS 2003, zwei Blöcke enthalten Aufgaben aus TIMSS 2007 und ein Block enthält Aufgaben aus TIMSS 2011 (Martin & Mullis, 2011c). Alle in diesem Bericht verwendeten Aufgabenbeispiele stammen aus dieser Aufgabenauswahl.

5.2 Kontextfragebögen

Die an TIMSS teilnehmenden Schülerinnen und Schüler sowie die an ihrem Lernprozess beteiligten Akteure werden im Zusammenhang mit der Leistungserhebung von TIMSS mit Hilfe von Fragebögen um Auskunft zu Kontextmerkmalen des schulischen und häuslichen Lernens gebeten. Auch die Entwicklung der Fragebogenitems erfolgt in einem kooperativen Prozess der TIMSS-Teilnehmerstaaten. Die Vorschläge für konkrete Fragen zur Erfassung von Hintergrundmerkmalen stammen aus verschiedenen Teilnehmerstaaten und werden von Expertinnen und Experten aus verschiedenen Fachdisziplinen und Ländern begutachtet. Ein maßgebliches Kriterium bei der Entwicklung von TIMSS-Fragebogenitems besteht darin, dass sich gewonnene Informationen in der Auswertung dazu nutzen lassen, Unterschiede in schulischen Leistungen von Schülerinnen und Schülern verschiedener Teilnehmerstaaten und Regionen durch ihre Hinzunahme interpretieren zu können. Beim Überprüfungsprozess wird zudem insbesondere sichergestellt, dass die ausgewählten Fragen die Testziele der TIMSS-Rahmenkonzeption repräsentieren und sich in allen Teilnehmerstaaten einsetzen lassen. Eine Überprüfung der empirischen Qualität erfolgt ebenfalls im Rahmen des Feldtests (Martin & Mullis, 2011g). Die von Seiten der internationalen Studienleitung vorgesehenen Inhalte der Kontextfragebögen stellen für alle Teilnehmerstaaten verbindliche Vorgaben dar (Mullis et al., 2009). Als solche sind sie immer auch ein Kompromiss hinsichtlich der Forschungsinteressen

verschiedener Staaten und decken keinesfalls alle Fragen ab, welche die Wissenschaft und Politik einzelner Staaten vordringlich beschäftigen. Aus diesem Grund hat das deutsche TIMSS-Konsortium die Kontextbefragung ergänzt.

5.2.1 Datenschutzrechtliche Begutachtung der Kontextfragebögen

Die Datenschutzbeauftragten der Länder der Bundesrepublik Deutschland wurden über alle im Rahmen von TIMSS 2011 implementierten Verfahren der Datenerhebung ausführlich informiert. Durch spezifische Anpassung der Testinstrumente und Erhebungsverfahren wurde sichergestellt, dass der Datenschutz den gesetzlichen Vorgaben entsprechend eingehalten wurde.

5.2.2 Inhalte der Kontextfragebögen

In den in Deutschland eingesetzten TIMSS-2011-Fragebögen werden unter anderem folgende Aspekte thematisiert:

Der *Schülerfragebogen* enthält Fragen zu Lernaktivitäten, zu fachspezifischen Einstellungen, Selbstkonzepten sowie emotionalen und leistungsthematischen Verhaltensweisen, zum familiären Hintergrund, zu soziodemographischen Merkmalen (Alter, Geschlecht) und zu Freizeitaktivitäten.

Mit dem *Elternfragebogen* können die individuellen Schülerangaben um Informationen zur familiären Lernumwelt ergänzt und erweitert werden. Sie fokussieren unter anderem familiäre Ressourcen, soziodemographische Daten (Status, höchster Bildungsabschluss, Fragen zur Berufstätigkeit, Informationen über die Zusammensetzung der Familie), Bildungshintergrund der Eltern, Bildungsaspirationen der Eltern, lernunterstützende Aktivitäten und den kognitiven Anregungsgehalt der familiären Lernumwelt.

Um das Unterrichtsgeschehen und die Klassenmerkmale zu erfassen, werden im *Fragebogen für die Mathematik- und Sachunterrichtslehrkräfte* unter anderem Personenmerkmale (Alter, Geschlecht, Lehrerfahrung), ihre eigene Ausbildung und Qualifizierung, Merkmale des Klassenkontexts, Ausstattung der Klassenzimmer, Einstellungen zu unterrichtsrelevanten Aspekten, Unterrichtspraxis und Lehrmethoden, lernrelevante Voraussetzungen der Schülerinnen und Schüler sowie persönliche Einstellungen zur Mathematik und den Naturwissenschaften erfragt. Informationen, die für die individuellen Schülerinnen und Schüler benötigt werden (Noten, Schullaufbahneempfehlungen, Förderbedarfe) werden durch eine Schülerteilnahmeliste erhoben, die die Lehrkräfte ausfüllen.

Der *Schulfragebogen*, der von den Schulleitungen ausgefüllt wird, umfasst unter anderem folgende Aspekte: Schulcharakteristika (Schulgröße, Lehrkräfte, Unterrichtszeiten, soziale und geographische Lage), Organisation, Kontext und Ausstattung der Schulen, pädagogische Zielsetzungen, curriculare Gestaltung an der Schule, Ressourcen in Form von technischer und materieller Ausstattung, soziales Klima, Kooperationen im Kollegium, Einbindung der Eltern, außercurriculare Aktivitäten und Unterstützungssysteme der Schule, Rolle der Schulleitung sowie Fragen zum zusätzlichen Kursangebot in Mathematik beziehungsweise in den Naturwissenschaften.

Der *Curriculumfragebogen*, der von Fachexpertinnen und Fachexperten ausgefüllt wird, enthält Fragen dazu, inwieweit die Inhalte der TIMSS-Tests in den jeweiligen intendierten Curricula vorgesehen sind. Daneben werden auch

Rahmenbedingungen der Umsetzung des Curriculums erfragt, zum Beispiel wer über das jeweilige Curriculum entscheidet, ob und in welcher Form die Implementation des Curriculums evaluiert wird oder wie Lehrkräfte bei der Umsetzung des Curriculums unterstützt werden.

5.2.3 Übersetzung und Gestaltung der Testinstrumente

Bei TIMSS werden die Schülerinnen und Schüler in ihrer Unterrichtssprache getestet und befragt. Jede nationale Forschungsgruppe ist entsprechend dafür verantwortlich, Versionen der Testinstrumente und Fragebögen in den für den nationalen Kontext relevanten Unterrichtssprachen zu erstellen (Martin & Mullis, 2011e). Zur Sicherung der Qualität und einheitlicher Eigenschaften der Erhebungsinstrumente legt die internationale Studienleitung präzise fest, nach welchen Richtlinien und Abläufen die Übersetzung und Gestaltung der Erhebungsinstrumente erfolgen soll. Als wichtigstes Kriterium gilt, dass bei Übersetzung der Testaufgaben oder Fragen in die jeweilige(n) Unterrichtssprache(n) die internationale Vergleichbarkeit gewahrt bleiben muss. Insbesondere für die Leistungstests gilt, dass durch die Übersetzung die Aufgaben oder Antwortoptionen keinesfalls schwieriger oder einfacher werden dürfen als im Original. Für die Übersetzung der Kontextfragebögen ist zu berücksichtigen, dass die Fragen im nationalen Kontext Sinn ergeben. Entsprechend ist auf Besonderheiten des nationalen Schulsystems Rücksicht zu nehmen.

Die Einhaltung der Prozeduren sowie die Qualität der Übersetzung werden von der internationalen Studienleitung genau kontrolliert. Nationale Anpassungen der international vorgegebenen Instrumente sind nur zulässig, sofern sie durch kulturelle Unterschiede begründet werden können (Yu & Ebbs, 2011). Zur Überprüfung der Qualität der Übersetzung in andere Sprachen und möglicher nationaler Adaptionen werden die nationalen Versionen der TIMSS-Instrumente von Übersetzern der IEA ins Englische zurückübertragen. Unerwünschte Abweichungen vom Inhalt und Layout des englischen Originals werden dabei in Abstimmung mit der nationalen Studienleitung korrigiert und zulässige nationale Abweichungen dokumentiert.

In Deutschland erfolgte die Übersetzung der Testaufgaben und Fragebögen in Kooperation mit der TIMSS-Forscherguppe aus Österreich und wurde von einer professionellen Übersetzerin begleitet, die bereits für den vorangegangenen Studienzyklus die Übersetzung verantwortete. Im Anschluss an die Übersetzung wurden die Testinstrumente und Fragebögen von Expertinnen und Experten begutachtet und mit Übersetzungen aus dem vorangegangenen Studienzyklus abgeglichen. Um die Qualität der Übersetzung für den Haupttest noch zu steigern, wurden die Leistungstests noch einmal sprachlich überarbeitet.

Über die internationalen Vorgaben hinaus bestand Bedarf an Übersetzungen einzelner erhebungsrelevanter Dokumente. Dies betraf Elternanschreiben, in denen über die Studie und ihre Organisation, Komponenten der Leistungsmessung und Befragungen sowie die datenschutzrechtlichen Grundlagen informiert wurde und Eltern gegebenenfalls – je nach landesspezifischem Datenschutzrecht – um ihr Einverständnis zur Teilnahme ihres Kindes an Komponenten der Erhebung gebeten wurden. Um sicherzustellen, dass nichtdeutschsprachige Eltern entsprechende Schreiben auch informationssicher lesen können, wurden diese je nach bundeslandspezifischem Bedarf in eine oder mehrere Sprachen übersetzt. Als Zielsprachen wurden insgesamt die folgenden zwölf Optionen angefor-

dert: Arabisch, Bosnisch, Englisch, Griechisch, Italienisch, Kroatisch, Kurdisch, Polnisch, Russisch, Serbisch, Türkisch und Vietnamesisch.

6 Erhebung

Die Haupterhebung von TIMSS 2011 fand in der Zeit vom 9. Mai bis zum 30. Juni 2011 an den teilnehmenden Schulen statt.

6.1 Gemeinsame Erhebung mit IGLU/PIRLS 2011

Eine Besonderheit von TIMSS 2011 ist, dass der Erhebungszeitpunkt der alle vier Jahre stattfindenden Untersuchung im Jahr 2011 erstmalig mit dem der Schwesterstudie IGLU/PIRLS (fünfjähriger Zyklus) zusammengefallen ist. Ebenso wie 37 andere Teilnehmer hat auch Deutschland sich zu einer Durchführung der beiden Studien mit gemeinsamer Stichprobe entschieden. Die Umsetzung erfolgte an zwei in der Regel aufeinanderfolgenden Testtagen. Nach einem Samplingsystem wurden die Testkomponenten so rotiert, dass an 50.5 Prozent der Schulen die IGLU-Testkomponente am ersten Testtag durchgeführt und die TIMSS-Testkomponente am zweiten Testtag vorgelegt wurde. An 49.5 Prozent der Schulen erfolgte die Durchführung in umgekehrter Reihenfolge. Berechnungen der internationalen Studienleitung konnten zeigen, dass das Studiendesign keinen statistisch signifikanten Einfluss auf die Leistungsergebnisse der Schülerinnen und Schüler hatte (Foy, 2012).

6.2 Aufbau der Untersuchung

Der international vorgegebene Testablauf ist so gestaltet, dass die teilnehmenden Schülerinnen und Schüler genügend Zeit zur Bearbeitung der Testaufgaben haben. Das Testdesign ist insgesamt so angelegt, dass die Testaufgaben abwechslungsreich und sinnvoll aufgeteilt sind und von Pausen unterbrochen bearbeitet werden können.

In Tabelle 2.4 sind die beiden möglichen Abläufe der zwei Testtage dargestellt. An Schulen, an denen TIMSS am ersten Testtag durchgeführt wurde, erfolgte in einem ersten Teil die Administration der Leistungstests und im Anschluss daran wurde von den Schülerinnen und Schülern ein international einheitlicher Fragebogen bearbeitet.

Für die Bearbeitung von je zwei Aufgabenblöcken pro Kompetenzdomäne waren 36 Minuten vorgesehen. Somit ergab sich eine reine Testzeit von 72 Minuten.

Am zweiten Testtag fand dann die Erhebung von IGLU/PIRLS statt. Im Anschluss an die Administration der Leistungstests wurde ein weiterer Fragebogen an die Kinder ausgeteilt, der über den internationalen Vergleich hinausgehende Fragen beinhaltete. Schülerinnen und Schüler, die am ersten Testtag nicht anwesend waren, wurden am zweiten Testtag gebeten, nach der Bearbeitung des jeweiligen Leistungstests den Fragebogen des ersten Testtages auszufüllen, damit für den internationalen Vergleich möglichst wenige Ausfälle zu verzeichnen waren.

Tabelle 2.4: Untersuchungsablauf von TIMSS 2011 und IGLU/PIRLS 2011**Option A: Klasse, in der die Erhebung am ersten Testtag mit TIMSS beginnt**

TIMSS 1. Testtag		IGLU/PIRLS 2. Testtag	
Beginn der Testsitzung: Verteilung des Materials, Einweisung	ca. 10 min.	Beginn der Testsitzung: Verteilung des Materials, Einweisung	ca. 10 min.
Testheft Teil I (Mathematik oder Naturwissenschaften)	36 min.	Lesetest Teil I	40 min.
Pause	10 min.	Pause	10 min.
Einweisung in Teil II	ca. 5 min.	Einweisung in Teil II	ca. 5 min.
Testheft Teil II (Mathematik oder Naturwissenschaften)	36 min.	Lesetest Teil II	40 min.
Pause [<i>Austeilen der Schüler- und der Elternfragebögen</i>]	15 min.	Pause [<i>Austeilen der Schülerfragebögen</i>]	15 min.
Einweisung in die Bearbeitung des kognitiven Fähigkeitstests	ca. 5 min.	Einweisung in die Bearbeitung des kognitiven Fähigkeitstests	ca. 5 min.
Kognitiver Fähigkeitstest	7 min.	Kognitiver Fähigkeitstest	8 min.
Einweisung in die Bearbeitung des Schülerfragebogens 1	ca. 5 min.	Einweisung in die Bearbeitung des Schülerfragebogens 2	ca. 5 min.
Bearbeitung des Schülerfragebogens 1	30 min.	Bearbeitung des Schülerfragebogens 2	30 min.
Beenden der Testsitzung, Einsammeln der Materialien	ca. 5 min.	Beenden der Testsitzung, Einsammeln der Materialien	ca. 5 min.
Gesamtzeit ca. 165 min.		Gesamtzeit ca. 173 min.	

Option B: Klasse, in der die Erhebung am ersten Testtag mit IGLU/PIRLS beginnt

IGLU/PIRLS 1. Testtag		TIMSS 2. Testtag	
Beginn der Testsitzung: Verteilung des Materials, Einweisung	ca. 10 min.	Beginn der Testsitzung: Verteilung des Materials, Einweisung	ca. 10 min.
Lesetest Teil I	40 min.	Testheft Teil I (Mathematik oder Naturwissenschaften)	36 min.
Pause	10 min.	Pause	10 min.
Einweisung in Teil II	ca. 5 min.	Einweisung in Teil II	ca. 5 min.
Lesetest Teil II	40 min.	Testheft Teil II (Mathematik oder Naturwissenschaften)	36 min.
Pause [<i>Austeilen der Schüler- und der Elternfragebögen</i>]	15 min.	Pause [<i>Austeilen der Schülerfragebögen</i>]	15 min.
Einweisung in die Bearbeitung des kognitiven Fähigkeitstests	ca. 5 min.	Einweisung in die Bearbeitung des kognitiven Fähigkeitstests	ca. 5 min.
Kognitiver Fähigkeitstest	8 min.	Kognitiver Fähigkeitstest	7 min.
Einweisung in die Bearbeitung des Schülerfragebogens 1	ca. 5 min.	Einweisung in die Bearbeitung des Schülerfragebogens 2	ca. 5 min.
Bearbeitung des Schülerfragebogens 1	30 min.	Bearbeitung des Schülerfragebogens 2	30 min.
Beenden der Testsitzung, Einsammeln der Materialien	ca. 5 min.	Beenden der Testsitzung, Einsammeln der Materialien	ca. 5 min.
Gesamtzeit ca. 173 min.		Gesamtzeit ca. 165 min.	

6.3 Durchführung der Erhebung

Die Durchführung der Erhebung erfolgte hoch standardisiert und kontrolliert. In Deutschland wurde das *IEA Data Processing and Research Center* (DPC) in Hamburg mit der Durchführung der Datenerhebung beauftragt. Die Arbeit des DPC umfasste den Kontakt mit den Schulen, aber auch die Auswahl der Testleiterinnen und Testleiter. Diese waren zumeist Lehramts-, Erziehungswissenschafts- oder Psychologiestudierende höheren Semesters, die bereits in früheren Studien Erfahrungen im Bereich der Testdurchführung sammeln konnten. Zudem nahmen alle Testleiterinnen und Testleiter an einer Schulung für die Testdurchführung teil und erhielten ein schriftliches Manual mit genauen Vorgaben für die Testdurchführung. Eine zentrale Vorgabe war es, den international vorgegebenen Testablauf akkurat einzuhalten. Nach standardisierter Vorgabe wurde jeder Testteil den Kindern ausführlich erklärt und die Durchführung anhand von Beispielen erläutert. Die Schülerinnen und Schüler hatten jederzeit die Möglichkeit, nicht auf den Inhalt bezogene Verständnisfragen zur Testbearbeitung zu stellen.

6.4 Qualitätssicherung

Die Aussagekraft des internationalen Vergleichs setzt voraus, dass der Test in allen Staaten unter vergleichbaren Voraussetzungen durchgeführt wird. Mit der Teilnahme an TIMSS verpflichteten sich alle Staaten zur Einhaltung aller Vorgaben und zur Implementation umfassender qualitätssichernder Maßnahmen.

Die Einhaltung der internationalen Vorgaben wurde in den Teilnehmerstaaten stichprobenartig von unabhängigen Expertinnen und Experten überprüft, die von der IEA dazu beauftragt wurden. Mit dem sogenannten *International Quality Control Monitoring* in Deutschland wurde Jun.-Prof. Dr. Falk Radisch vom Institut für Bildungsforschung an der *School of Education* der Bergischen Universität Wuppertal beauftragt. Unabhängig von der Arbeit des deutschen TIMSS-Konsortiums kontrollierte er in einer Stichprobe von zehn Prozent aller Testklassen den Ablauf der Testdurchführung und dokumentierte seine Beobachtungen in von der internationalen Studienleitung vorgegebenen Beobachtungsbögen. Neben den Testbeobachtungen wurden auch die Schulkoordinatorinnen und Schulkoordinatoren und die Testleitungen mündlich befragt. Bei der Qualitätskontrolle in Deutschland konnten keine Mängel in der Einhaltung der Erhebungsstandards festgestellt werden.

Ergänzend wurde ein von der nationalen Studienleitung verantwortetes Qualitätsmonitoring an weiteren 21 zufällig ausgewählten Testschulen durchgeführt. Bei der Durchführung der Maßnahme wurde das IFS in Dortmund durch das Arbeitsteam von Prof. Dr. Knut Schwippert (Konsortium IGLU/PIRLS) an der Universität Hamburg unterstützt. Die konkrete Umsetzung und Koordination wurde dabei hauptverantwortlich von Jun.-Prof. Dr. Doren Prinz übernommen. Die Qualitätsbeobachtung, die von Mitarbeiterinnen und Mitarbeitern der Universität Hamburg und dem IFS durchgeführt wurden, umfasste dabei jeweils beide Testtage sowie ein anschließendes Interview mit der Schulkoordination zu Testqualität, Organisation und Belastungsempfinden der Schulen. Auch durch das *National Quality Control Monitoring* konnten keine Mängel in der Einhaltung der Erhebungsstandards festgestellt werden.

7 Stichprobe und Beteiligungsquoten in Deutschland

Die für Deutschland zufällig gezogene Schulstichprobe umfasste zunächst 200 Schulen. Durch eine explizite Stratifizierung nach den 16 Ländern der Bundesrepublik Deutschland wurde sichergestellt, dass die Länder gemäß ihrer Größe und Schulanzahl adäquat durch die Stichprobe abgebildet werden. Darüber hinaus erfolgte eine implizite Stratifizierung nach regionaler Zugehörigkeit und Schultyp. Die Schulstichprobe wurde schon vor der Testung um zwei Schulen reduziert: Bei einer Grundschule handelte es sich um eine internationale Schule mit Englisch als Unterrichtssprache, bei der anderen um eine Privatschule, die nicht zur Teilnahme verpflichtet war. Die endgültige deutsche TIMSS-Stichprobe umfasst demnach 198 Schulen und 4241 Schülerinnen und Schüler. Darunter befinden sich sieben nachgerückte Schulen. An 197 der 198 getesteten Schulen konnte die Leistungsmessung planmäßig durchgeführt werden. In Absprache mit der nationalen Projektleitung wurde an einer Förderschule die Testung abgebrochen.

Von den 4229 Schülerinnen und Schülern an den 197 Schulen der Stichprobe nahmen schließlich 3995 Schülerinnen und Schüler tatsächlich am TIMSS-Test teil. Die Nicht-Teilnahme von 234 Schülerinnen und Schülern ist auf verschiedene Gründe zurückzuführen (o.g. Ausschlussgründe, Verlassen der Schule vor dem Testtag oder Abwesenheit am Testtag). Deutschland erreicht damit, ähnlich wie in TIMSS 2007, eine Gesamtteilnahmequote (für Schüler *und* Schulen) von 92 Prozent exklusive beziehungsweise 95 Prozent inklusive Ersatzschulen (vgl. Tabelle A.2 im Anhang).

Auf erfreulich hohe Akzeptanz stieß die TIMS-Studie auch bei Lehrkräften, Schulleitungen und Eltern. Der Anteil der TIMSS-Klassen, aus denen mindestens ein Lehrerfragebogen vorliegt, beläuft sich auf 97 Prozent. Die Rücklaufquote der TIMSS-Schulfragebögen beträgt für Deutschland 95 Prozent. Die Rücklaufquote der Elternfragebögen liegt bei 80 Prozent.

8 Aufbereitung und Analyse der Daten

Mit der nationalen Datenverarbeitung und -aufbereitung und der Kodierung der Leistungstests war in Deutschland das DPC beauftragt. Die Dateneingabe und -verarbeitung konnte technologiegestützt erfolgen, was abgesehen von einer niedrigen Fehlerquote den Vorteil eines schnellen Zugriffs auf die Antworten einzelner Schülerinnen und Schüler hat, für den Fall, dass individuelle Antworten überprüft werden müssen. Antworten, die von der Scan-Software nicht erkannt werden konnten, wurden von ausgebildeten Mitarbeiterinnen und Mitarbeitern verifiziert. Die erstellten Datensätze wurden anschließend mit Hilfe speziell entwickelter Software auf ungültige Daten und Inkonsistenzen innerhalb der Daten geprüft.

8.1 Kodierung der Leistungstests

Die Bewertung der Testlösungen erfolgte nach eng umrissenen internationalen Vorgaben. Während bei Mehrfachwahlaufgaben die richtige Lösung eindeutig festgelegt ist, weisen Testfragen mit offenem Antwortformat (vgl. Abschnitt 5.1.1 sowie Kapitel 3 und 4) häufig einen Bewertungsspielraum auf. Die Kodierung

der Fragen mit offenem Antwortformat wurde in Deutschland ebenfalls dem DPC übertragen und dort von erfahrenen Kodierern übernommen, die bevorzugt Mathematik oder ein naturwissenschaftliches Fach studieren. Die Kodierer wurden in einem Training mit Beispiellösungen intensiv auf ihre Aufgabe vorbereitet und erhielten eine Kodieranweisung mit genauen Beschreibungen für richtige und falsche Lösungen zu jeder einzelnen Aufgabe. Die Güte der Kodierungen wurde mit Hilfe von Doppelkodierungen geprüft. Dazu wurden 200 zufällig ausgewählte Antworten zu jeder Aufgabe von einer zweiten Person beurteilt, und es wurde die Übereinstimmung zwischen dem Erst- und dem Zweitkodierer bestimmt. In TIMSS 2011 ist, wie auch in TIMSS 2007, die durchschnittliche Übereinstimmung über alle Aufgaben in allen Staaten sehr hoch (vgl. Martin & Mullis, 2011h).

9 Skalierung der Leistungstests

Mit der in TIMSS eingesetzten Leistungstestung soll es gelingen, von den in den Testheften gezeigten Schülerantworten auf Kompetenzausprägungen zu schließen. Die Verortung von Schülerinnen und Schülern auf der Kompetenzskala wird Skalierung genannt. In TIMSS wird für die Bestimmung individueller Leistungswerte auf statistische Modelle zurückgegriffen, die auf der *Item Response Theory* basieren (IRT, Boomsma, van Duijn & Snijders, 2000; van der Linden & Hambleton, 1997). In historischer Perspektive ist TIMSS ein Vorreiter der Etablierung dieser Verfahren in international-vergleichenden Schulleistungsstudien (Wendt, Bos & Goy, 2011). Die IRT geht von der Annahme aus, dass die beobachtbaren Antworten einer Person in einem Test durch eine nicht beobachtbare, testbezogene Fähigkeit erklärbar sind. Die beobachteten Antworten werden verwendet, um die Höhe der nicht beobachtbaren (*latenten*) Fähigkeit zu schätzen. Ein Vorzug der IRT besteht darin, dass sich die Schwierigkeit einer Aufgabe und die Fähigkeit einer Person auf derselben Skala abbilden lassen. In TIMSS wird die IRT-Skalierung nicht zuletzt aufgrund des implementierten Testheftdesigns (*Multi-Matrix-Design*, vgl. Tabelle 2.3) gewählt (vgl. Foy, Brossman & Galia, 2012). Dieses Untersuchungsdesign sieht vor, dass die einzelnen teilnehmenden Schülerinnen und Schüler immer nur eine kleine Auswahl aller Testaufgaben des TIMSS-Aufgabenpools bearbeiten. Da diese Ausschnitte nicht immer in ihrer Schwierigkeit übereinstimmen, sind auch die darauf basierenden Testleistungen nicht ohne weiteres vergleichbar. Durch die Skalierung wird dies ermöglicht – eine ausreichend große Stichprobe und Überlappungen der Testheftinhalte vorausgesetzt (vgl. z.B. Kolen, 1981; Kolen & Brennan, 2004; Lord, 1980). Allerdings hat die IRT auch Grenzen. Die eingeschränkte Aufgabenanzahl pro Testheft sowie Schwierigkeitsunterschiede zwischen den Testheften führen mitunter zu ungenauen Fähigkeitsschätzungen für einzelne Personen (vgl. Bos & Voss, 2008). Da TIMSS aber nicht auf Individualdiagnostik abzielt, sondern Aussagen über Gruppen treffen möchte, sind diese Ungenauigkeiten zu vernachlässigen.

Um sicherzustellen, dass sich auf der Grundlage individueller Testleistungen akkurate Rückschlüsse auf Populationsmerkmale ziehen lassen, wird in TIMSS zur Ermittlung der Schülerleistung (*Personenparameter*) der *Plausible-Values*-Ansatz gewählt (Mislevy, 1991; Mislevy, Beaton, Kaplan & Sheehan, 1992), ein Verfahren, dem die Theorie der Multiplen Imputation zugrunde liegt (Rubin, 1987). Die Grundidee dieses Ansatzes ist es, die nicht beobacht-

bare Fähigkeit einer Person als fehlenden Wert zu betrachten, der durch einen ‚plausiblen Wert‘ ersetzt werden kann. Dazu werden alle als relevant erachteten Informationen über eine Person herangezogen. Neben der Information über die Testleistung der Person wird dabei eine Vielzahl von Informationen aus den Hintergrundfragebögen berücksichtigt. Aus den Informationen wird eine bedingte Wahrscheinlichkeitsverteilung für die Fähigkeit der Person erstellt. Je mehr Informationen über eine Person vorliegen, desto geringer wird die Streuung dieser Verteilung. *Plausible Values* sind nun Fähigkeitswerte, die für jede Testperson zufällig aus diesen auf Grundlage von Hintergrundinformationen bedingten Verteilungen gezogen werden. Durch die Ziehung mehrerer *Plausible Values* wird berücksichtigt, dass die Ersetzung eines fehlenden Werts immer auch mit Unsicherheit behaftet ist. Als Konvention hat sich die Ziehung von fünf Werten etabliert. Analysen mit den Leistungsdaten müssen entsprechend immer fünfmal durchgeführt werden und die Ergebnisse dann zusammengefasst werden.

9.1 Skalierungsmodelle

Im vorliegenden Bericht werden die Kompetenzen der Schülerinnen und Schüler auf Grundlage der Skalierung berichtet, die die internationale Studienleitung durchgeführt hat (vgl. Foy et al., 2012). Die Skalierung der Leistungsdaten aller Teilnehmerstaaten ist eine komplexe und zeitintensive Aufgabe, der umfangreiche Datenprüfungs- und Aufbereitungsschritte vorausgehen. Für den Studienzyklus 2011 erfolgte erst nach Abschluss aller Prüfungen und einem Review der Ergebnisse durch alle beteiligten nationalen Studienleitungen eine Weitergabe der Datensätze aller Teilnehmerstaaten und Regionen an die nationalen Studienleitungen im September 2012 (vgl. Foy, 2012). Eine Veröffentlichung der Datensätze für Studienexterne ist für Anfang 2013 vorgesehen. Zentrale Modellparameter werden in der technischen Dokumentation zur Studie veröffentlicht (Foy et al., 2012).

Die internationale Skalierung erfolgte im Zeitraum von Herbst 2011 bis Sommer 2012 (vgl. Foy et al., 2012). Eine Qualitätssicherung dieser Berechnungen erfolgte durch Expertinnen und Experten vom US-amerikanischen *Educational Testing Service* (ETS). Ausgangspunkt der Skalierung ist eine umfassende Prüfung der Messeigenschaften einer jeden Testaufgabe nach festgelegten Kriterien (vgl. Martin & Mullis, 2011h). Trotz umfangreicher Prüfung der Testaufgaben im Anschluss an den Feldtest zeigten zwei Aufgaben für Mathematik und sieben Aufgaben für Naturwissenschaften ungenügende Messeigenschaften. Als Ergebnis wurden für die Ermittlung der Kompetenzmittelwerte für Mathematik 175 der 177 administrierten Testaufgaben genutzt. Die Kompetenzmittelwerte für die Naturwissenschaften wurden auf Basis von 168 der 175 administrierten Testaufgaben ermittelt (Foy et al., 2012).

Die Skalierung erfolgte unter Nutzung eines mehrdimensionalen dreiparametrischen logistischen Modells (3-PL Modell), in dem neben der Schwierigkeit und Trennschärfe auch eine spezifische Ratewahrscheinlichkeit für jede Testaufgabe berücksichtigt wird. Die Modellparameter wurden mit dem Computerprogramm *Parscale* geschätzt (Muraki & Bock, 1999). Für die Schätzung der Personenparameter wurde die Software *MGROUP* (Sheehan, 1985) verwendet. Um genauere Schätzer der Personenparameter zu erhalten, kommt in TIMSS 2011 – anders als im vorangegangenen Studienzyklus, in dem die Kompetenzwerte für die Subskalen (je drei Inhaltsbereiche und kogni-

tive Anforderungsbereiche) separat ermittelt wurden – für die Skalierung der Subskalen erstmalig ein mehrdimensionales Antwortmodell zur Anwendung, mit dem die in den Daten enthaltene Zusammenhangsstruktur besser abgebildet werden kann. Um eine Vergleichbarkeit zum vorangegangenen Studienzyklus sicherzustellen, wurden auch die Daten der Subskalen von TIMSS 2007 erneut skaliert. Abweichungen zur Berichtslegung für TIMSS 2007 (Bos et al., 2008; Martin et al., 2008; Mullis et al., 2008) sind entsprechend durch eine Veränderung des Skalierungsmodells bedingt.

Um für diesen Berichtsband schon frühzeitig mit vertiefenden Analysen beginnen zu können, wurde von der nationalen Studienleitung eine gesonderte Skalierung für die Leistungsdatensätze der Viertklässlerinnen und Viertklässler in Deutschland vorgenommen. Diese Skalierung diente unter anderem als Grundlage für ein umfassendes nationales *Itemreview*, welches auch die Grundlage für die nationale Interpretation der Kompetenzstufen lieferte. Für diese nationale Skalierung wurde das *Partial-Credit*-Modell (Masters, 1982) verwendet. Die Modellparameter wurden mit der Software *ConQuest* (Wu, Adams, Wilson & Haldane, 2007) geschätzt. Als Personenparameter wurden zunächst *Weighted Likelihood Estimates* (WLE, vgl. Warm, 1989) berechnet. Zu einem späteren Zeitpunkt wurde das Basismodell mehrdimensional sowie um ein nationales Hintergrundmodell erweitert (vgl. Kapitel 9). Als Grundlage für die Tabelle 3.11 und die Tabelle 4.9 (Kapitel 3 und 4) wurde von der nationalen Studienleitung die internationale Skalierung repliziert. Da der Veröffentlichungstermin internationaler Itemparameter und Rahmendaten zum internationalen Hintergrundmodell nach der Drucklegung für diesen Band lag, konnte diese Berechnung lediglich auf Basis des *Partial-Credit*-Modells (Masters, 1982) unter Nutzung der Software *ConQuest* (Wu et al., 2007) erfolgen.

10 Gewichtung und Schätzung von Stichproben- und Messfehlern

Das Ziel einer repräsentativen Studie ist es, von den Verhältnissen in einer Stichprobe auf die Grundgesamtheit zu schließen. Zuverlässige Rückschlüsse setzen jedoch voraus, dass die Stichprobe nicht verzerrt ist und die Grundgesamtheit angemessen repräsentiert. In TIMSS 2011 finden sich zwei Ursachen, die zu Stichprobenverzerrungen führen können (Joncas & Foy, 2011):

Erstens hat aufgrund des Designs der Stichprobenauswahl (vgl. Abschnitt 4.3) nicht jede Schülerin beziehungsweise jeder Schüler dieselbe Wahrscheinlichkeit, in die Stichprobe zu gelangen. So hängt die Wahrscheinlichkeit unter anderem von der Zugänglichkeit der Schulen ab: In einer Schule mit zwei Jahrgangsklassen hätte ein Schulkind eine Wahrscheinlichkeit von 50 Prozent in die Stichprobe zu gelangen. In einer Schule mit vier Jahrgangsklassen wäre es dagegen nur eine Wahrscheinlichkeit von 25 Prozent. Inferenzstatistiken setzen jedoch Zufallsstichproben voraus, bei denen jede Person dieselbe Wahrscheinlichkeit hat, in die Stichprobe zu gelangen.

Zweitens kann auch der Stichprobenausfall eine Verzerrung bedeuten. Nur selten sind Ausfälle zufällig; vielmehr weisen sie oft einen Zusammenhang mit studienrelevanten Merkmalen auf. So ist es denkbar, dass leistungsschwache Schülerinnen beziehungsweise Schüler mit höherer Wahrscheinlichkeit am Testtag fehlen als durchschnittliche oder leistungsstarke Schülerinnen beziehungsweise Schüler.

hungsweise Schüler. Werden Stichprobenverzerrungen nicht angemessen korrigiert, können inferenzstatistische Methoden zu falschen Schlussfolgerungen führen. In TIMSS wird diesem Problem dadurch begegnet, dass für jedes getestete Schulkind ein statistisches Gewicht berechnet wird (siehe Joncas & Foy, 2011). Die Gewichte werden dann bei der Berechnung aller Statistiken verwendet, zum Beispiel bei einem Mittelwert oder bei Prozentangaben. Die Daten der teilnehmenden Schülerinnen und Schüler tragen dadurch in kontrollierter Art und Weise unterschiedlich stark zu der Berechnung der Statistiken bei.

Da in TIMSS keine reinen Zufallsstichproben vorliegen, sondern sogenannte Clusterstichproben gezogen werden, kann die Bestimmung des Standardfehlers nicht mit sonst üblichen Analyseverfahren vorgenommen werden, da sonst der Standardfehler systematisch unterschätzt werden würde (vgl. Joncas & Foy, 2011). Eine präzisere Bestimmung des Standardfehlers erlauben sogenannte *Jackknife*-Verfahren. Diese Verfahren bestimmen die Variabilität der Schätzung von Populationskennwerten (wie z.B. die der Mathematikleistung der in TIMSS 2011 getesteten Grundschülerinnen und Grundschüler) durch ein wiederholtes Schätzen dieser Werte aus Substichproben, was die Möglichkeit bietet, Stichprobenfehler zu schätzen, ohne zugleich die Annahme einfacher Zufallsstichproben voraussetzen zu müssen. Durch die Reduzierung der Freiheitsgrade führt diese Methode zu akkurateren Schätzungen der Standardfehler.

Die korrekte Bestimmung des Standardfehlers ist sehr wichtig, weil dieser genutzt wird, um zu ermitteln, ob sich zwei Gruppen signifikant voneinander unterscheiden. Im vorliegenden Bericht werden entsprechend alle Standardfehler mit solch einem Verfahren (*Jackknife Repeated Replication Technique*, JRR) geschätzt (Joncas & Foy, 2011). In diesem Bericht wird die *Signifikanz* von Unterschieden aufgrund der großen Stichproben in der Regel über *z*-Werte bestimmt. Für die Entscheidung über die Signifikanz wird eine Irrtumswahrscheinlichkeit von $\alpha = .05$ zugrunde gelegt. Bei besonders vielen Signifikanztests, zum Beispiel beim Vergleich jedes Staats mit allen anderen Staaten, werden bei $\alpha = .05$ unabhängig von den echten Verhältnissen in der Population allein schon aus Zufall 5 Prozent der Vergleiche signifikant. Dieser unerwünschte Effekt wird üblicherweise mit einer Verringerung der Irrtumswahrscheinlichkeit ausgeräumt, zum Beispiel über die Bonferroni-Korrektur. Die Höhe der Korrektur hängt von der Anzahl der Vergleiche und somit von der Anzahl der Teilnehmerstaaten in TIMSS ab. Dies würde über die einzelnen TIMSS-Zyklen hinweg, aber auch in Tabellen oder Abbildungen, die sich jeweils nur auf ausgewählte Staaten eines Zyklus beziehen, zu einer je unterschiedlichen Korrektur führen. Die verschiedenen Signifikanztests wären dann nicht mehr vergleichbar. Die Entscheidung, im internationalen Ergebnisbericht keine Anpassung der Irrtumswahrscheinlichkeit vorzunehmen, wird daher auch für den vorliegenden deutschen Bericht übernommen (Martin & Mullis, 2011h).

11 Zur Darstellung und Interpretation der Ergebnisse

Für die Darstellung der in diesem Band vorgestellten Ergebnisse in Tabellen und Abbildungen werden verschiedene statistische Kennwerte verwendet. In diesem Abschnitt werden die wichtigsten Kennwerte in knapper Form erläutert. Für darüber hinausgehende Begriffserklärungen und technische Grundlagen sei auf den technischen Bericht der internationalen Studienleitung (Martin & Mullis, 2011h) sowie die einschlägige Fachliteratur verwiesen.

Mittelwerte und Standardabweichungen von Leistungsdaten

Die in diesem Bericht angeführten Ergebnisse und Vergleiche basieren auf der internationalen Stichprobe und den internationalen Kompetenz- und Fragebogenskalen von TIMSS 2011 (Martin et al., 2012; Mullis et al., 2012a). Zur Darstellung der Leistungswerte wurde für die erste TIMSS-Erhebung im Jahr 1995 ein Mittelwert (M für arithmetisches Mittel) von 500 Punkten und eine Standardabweichung (SD für *Standard Deviation*) von 100 Punkten festgelegt (Martin & Kelly, 1998; Mullis et al., 1998). Dabei beruht die Wahl der Einheiten für diese Skala ausschließlich auf Konventionen. Werte, die nahe beim Mittelwert liegen, kommen häufiger vor als Extremwerte. Oft ergibt sich eine Normalverteilung, wie sie in Abbildung 2.6 dargestellt ist.

Die Standardabweichung ist ein statistischer Kennwert, der die durchschnittliche Streuung der Werte um den Mittelwert quantifiziert. Im Bereich einer Standardabweichung über und unter dem Mittelwert (d.h. in Abbildung 2.6 im Bereich von 400 bis 600 Punkten) liegen rund zwei Drittel (68,3%) aller Testwerte der internationalen Population. Bei zwei Standardabweichungen erhöht sich dies auf 95,5 Prozent und bei drei Standardabweichungen auf 99,7 Prozent.

Internationaler Mittelwert vs. Skalenmittelwert

TIMSS ist als Trendstudie konzipiert, das heißt, dass in regelmäßigen Abständen die Erhebungen wiederholt werden, um so Veränderungen in den mittleren Leistungsniveaus der Teilnehmerstaaten über die Zeit hinweg darstellen zu können.

Um die Ergebnisse der Schülerinnen und Schüler über die verschiedenen Zyklen vergleichen zu können, wird für die Darstellung der Ergebnisse dieselbe Skala zugrunde gelegt. Die Vergleichbarkeit der Ergebnisse wird gewährleistet, indem in jedem Zyklus Aufgaben aus den vorangegangenen Studienzyklen erneut eingesetzt werden. Durch die gemeinsamen Testaufgaben können die Daten von TIMSS 2011 in einer gemeinsamen Skalierung mit den Daten von 2007 verankert werden (Foy et al., 2012). Auf gleiche Weise lässt sich TIMSS 2007 wiederum mit dem vorangegangenen Studienzyklus in Beziehung setzen. Somit werden die Daten der ersten TIMSS-Erhebung im Jahr 1995 zum Referenzpunkt für alle Folgezyklen. Veränderungen über die Zyklen hinweg können so immer mit Bezug zu einem festen Punkt, dem Mittelwert in TIMSS 1995 mit $M = 500$, beurteilt werden. Dieser Referenzwert wird in TIMSS als *Skalenmittelwert* bezeichnet.

Die Bezeichnung *internationaler Mittelwert* wird dagegen für einen Wert benutzt, der mit jeder TIMSS-Erhebung neu berechnet wird: Der Mittelwert über alle Mittelwerte der jeweiligen Teilnehmerstaaten, wobei die Benchmark-Teilnehmer (vgl. Abschnitt 4.2.1) nicht berücksichtigt werden. Im Gegensatz zum Skalenmittelwert variiert der internationale Mittelwert von Untersuchungszyklus zu Untersuchungszyklus. Setzen sich zum Beispiel die Teilnehmerstaaten einer Erhebung aus Staaten zusammen, die im Vergleich zu den Staaten, die an TIMSS 1995 teilgenommen haben, insgesamt leistungsstärker sind, ergibt sich ein internationaler Mittelwert der entsprechend größer als 500 ist. In TIMSS 2011 betragen die internationalen Mittelwerte für Mathematik 491 Punkte ($SD = 81$) und für Naturwissenschaften 486 Punkte ($SD = 85$). Somit ist die Gesamtheit der an TIMSS 2011 teilnehmenden Staaten leistungsschwächer als die Gesamtheit der Teilnehmerstaaten im Jahr 1995.

Bezüglich der berichteten Hintergrunddaten wird nicht zwischen einem internationalen Mittelwert und einem Skalenmittelwert unterschieden, da die

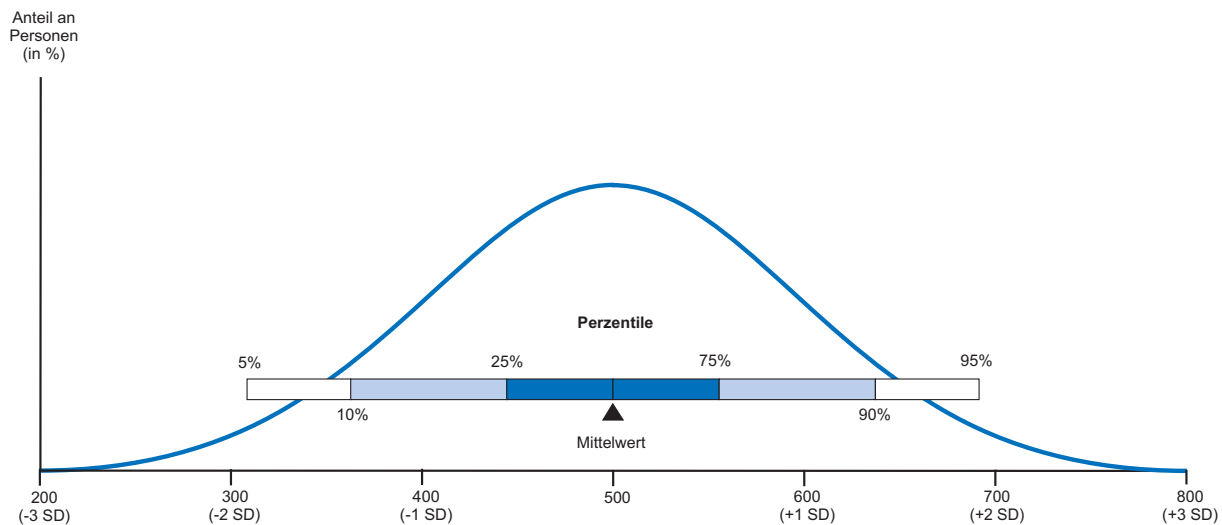
Fragebogeninhalte im Gegensatz zu den Leistungstests in sich sehr heterogene Inhalte abdecken, die zudem von Erhebung zu Erhebung in Teilen variieren.

Median, Perzentile und Perzentilbänder

In einigen Abbildungen werden neben dem Mittelwert und der Standardabweichung auch Perzentilwerte berichtet. Perzentilwerte geben, wie auch die Standardabweichung, eine Information über die Werteverteilung. Ein bestimmter Perzentilwert gibt an, wie viel Prozent der untersuchten Personen den genannten Wert erreichen oder darunter bleiben. Liegt beispielsweise der Leistungswert zum 5. Perzentil bei 318, bedeutet dies, dass 5 Prozent der untersuchten Schülerinnen und Schüler einen Punktwert von 318 oder geringer erreichen. Gleichzeitig bedeutet dies auch, dass 95 Prozent der Schülerinnen und Schüler einen Wert erreichen, der besser als 318 ist. Entsprechend trennt der Punktwert des 25. Perzentils das untere Leistungsviertel ab und der des 75. Perzentils das obere Leistungsviertel. Das 50. Perzentil, auch Median genannt, trennt die Verteilung in zwei Hälften mit gleicher Personenanzahl.

Im vorliegenden Bericht werden die Perzentilwerte entweder tabellarisch oder aber graphisch in Form von Perzentilbändern (vgl. Abbildung 2.6) dargestellt. Wenn sich die Werte, wie in Abbildung 2.6, einer Normalverteilung anpassen, ergibt sich ein symmetrisches Perzentilband und der Median fällt mit dem Mittelwert zusammen.

Abbildung 2.6: Normalverteilung mit Perzentilen



IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

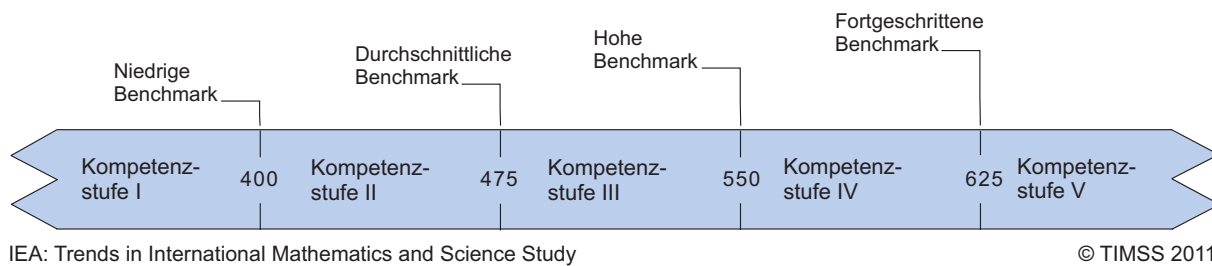
Kompetenzstufen

Die Leistungsskala in TIMSS deckt ein breites Kompetenzspektrum ab, da die Testaufgaben sehr vielseitige Inhalte ansprechen und verschiedene Prozesse kognitiver Anforderungen abbilden. Um eine inhaltliche Interpretation der von den getesteten Schülerinnen und Schülern erreichten Kompetenzwerte zu ermöglichen, werden in TIMSS sogenannte Kompetenzstufen gebildet. Inhaltlich werden die Kompetenzstufen in den Kapiteln 3 und 4 sowie den Anhängen B und C beschrieben.

Die Entwicklung und Beschreibung der Kompetenzstufen wird von einem international zusammengesetzten Expertengremium, dem *Science and Mathematics Item Review Committee* (SMIRC), unter Leitung der internationalen Studienleitung vorgenommen. Ausgangspunkt der Entwicklung der Kompetenzstufen ist zunächst die Festlegung von zentralen Bezugspunkten, sogenannten *Benchmarks*, auf den Leistungsskalen (Martin et al., 2012; Mullis et al., 2012a). Durch die Setzung von vier *Benchmarks* (400, 475, 550, 625) wird die Leistungsskala in fünf Abschnitte eingeteilt, die in Deutschland als *Kompetenzstufen* bezeichnet werden. Die niedrige Benchmark liegt bei 400 Punkten und bildet die Grenze zwischen Kompetenzstufe I und Kompetenzstufe II. Die durchschnittliche Benchmark liegt bei 475 Punkten und markiert den Beginn von Kompetenzstufe III. Die hohe und die fortgeschrittene Benchmark liegen bei 550 beziehungsweise 625 Punkten. Mit ihrer Überschreitung beginnen Kompetenzstufe IV beziehungsweise Kompetenzstufe V (vgl. Abbildung 2.7).

Abbildung 2.7 veranschaulicht die Setzung der internationalen Benchmarks und die sich daraus ergebenden Intervalle. Das unter der niedrigen Benchmark liegende Intervall wird in Deutschland als Kompetenzstufe I definiert. Die nur Kompetenzstufe I erreichenden Schülerinnen und Schüler sind nicht in der Lage, die für Kompetenzstufe II charakteristischen, relativ einfachen Aufgaben zu lösen. Die untere Grenze der Kompetenzstufe I fällt mit dem theoretischen Nullpunkt der TIMSS-Skala zusammen. Aufgrund der geringen Anzahl richtig gelöster Aufgaben lässt sich das Kompetenzniveau der Schülerinnen und Schüler am unteren Ende der Leistungsskala auf der Basis der eingesetzten Leistungstests nicht mehr differenziert beschreiben. Dies soll jedoch nicht zu der Annahme verleiten, die Schülerinnen und Schüler auf Kompetenzstufe I seien „ohne Kompetenz“. Sie erreichen schlicht nicht die mit dem internationalen TIMSS-Test festgelegten Mindestanforderungen.

Der konkreten Beschreibung und Entwicklung der Kompetenzstufen liegt jeweils eine spezifische Auswahl von Aufgaben zugrunde, die für das auf den vier Benchmarks gezeigte Leistungsniveau charakteristisch sind. Um zu dieser Aufgabenauswahl zu gelangen, werden jeweils all diejenigen Schülerinnen und Schüler ausgewählt, die folgende mittlere Leistungswerte erzielten: 395–405, 470–480, 545–555, 620–630 und sich damit im Bereich von fünf Punkten unter bis fünf Punkten über einem Benchmark-Wert verorten lassen. Diese Zuordnung ist möglich, da ein Verfahren genutzt wird, das eine Darstellung von Schülerfähigkeiten und Aufgabenschwierigkeiten auf einer gemeinsamen Skala erlaubt (vgl. Abschnitt 9). Für die Beschreibung der mittleren Benchmarks werden dann jene Aufgaben herangezogen, die von mindestens 65 Prozent der zugeordneten Schülerinnen und Schüler gelöst werden, aber von weniger als 50 Prozent der Schülerinnen und Schüler der darunter liegenden Benchmark. Für die Kompetenzstufen am oberen und unteren Ende der Leistungsskala gelten leicht modifizierte Kriterien (Martin & Mullis, 2011h).

Abbildung 2.7: Beziehung von Benchmarks und Kompetenzstufen

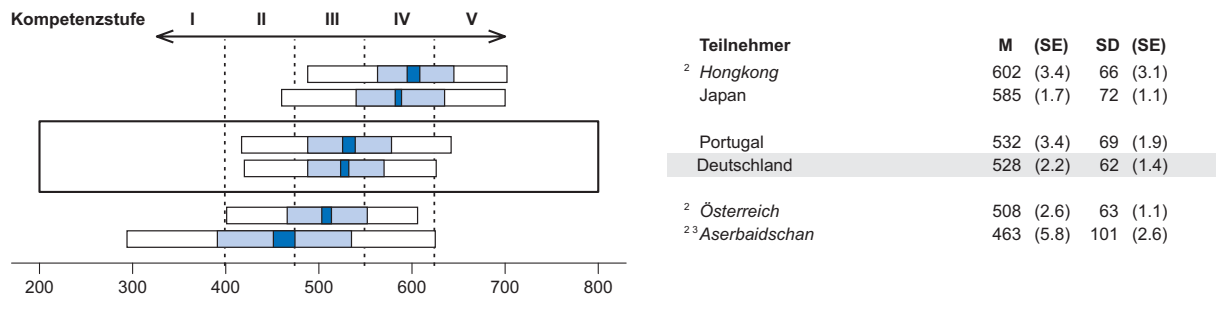
Standardfehler und Signifikanz

Ziel von TIMSS ist es, die auf Grundlage einer Stichprobe gewonnenen Erkenntnisse auf die Grundgesamtheit zu übertragen. Allerdings weicht jede Stichprobe von der Population ab, aus der sie stammt. Dabei variiert der Grad der Abweichung von Stichprobe zu Stichprobe. Sie ist in der Regel umso geringer, je größer die verwendeten Stichproben sind. Für Stichprobenuntersuchungen kann der jeweilige wahre Mittelwert der Grundgesamtheit – wenn auch relativ genau – nur geschätzt werden. Bei der Berechnung von Mittelwerten lässt sich, ähnlich wie für die Messwerte innerhalb einer Stichprobe, eine Streuung berechnen. Dieses Streuungsmaß ist der Standardfehler (*SE* für *Standard Error*) oder auch Standardschätzfehler. Der Standardfehler quantifiziert die Unsicherheit von aus Stichprobendaten geschätzten Populationswerten. Für die Interpretation gilt zunächst: Je kleiner der Standardfehler ausfällt, desto genauer ist ein geschätzter Populationskennwert. Wie bei der Standardabweichung gilt beim Standardfehler, dass im Bereich von zwei Standardfehlern unter bis zu zwei Standardfehlern über einem errechneten Mittelwert der „wahre“ Populationswert mit 95-prozentiger Wahrscheinlichkeit liegt.

Der Standardfehler lässt sich auch nutzen, um auf einen Blick einzuschätzen, ob sich zwei oder mehr Gruppen signifikant voneinander unterscheiden. Daher werden die Perzentilbänder in diesem Bericht um Konfidenzintervalle (Vertrauensintervalle) der Mittelwerte ergänzt (vgl. Abbildung 2.8). Diese Konfidenzintervalle geben an, in welchem Wertebereich der wahre Populationswert mit 95-prozentiger Wahrscheinlichkeit liegt. Zugrunde gelegt wird das Intervall von über beziehungsweise unter 1.96 (gerundet 2) Standardfehlern um die jeweilige Statistik. Um zu prüfen, ob sich zum Beispiel die Mittelwerte zweier Staaten überzufällig unterscheiden, können die Konfidenzintervalle um die jeweiligen Mittelwerte betrachtet werden.

Wenn sich zwei Konfidenzintervalle nicht überlappen, wie in Abbildung 2.8 die Konfidenzintervalle um die mittlere Mathematikleistung von Hongkong beziehungsweise Japan und Deutschland, so entspricht dies einem signifikanten Unterschied mit einer Irrtumswahrscheinlichkeit von $\alpha = .05$.

Abbildung 2.8: Darstellung von Perzentilbändern mit Konfidenzintervallen am Beispiel der Mathematikleistung in TIMSS 2011



Perzentile
5% 25% 75% 95%

Konfidenzintervalle (+/- 2 SE) um den Mittelwert

- Nicht statistisch signifikant vom deutschen Mittelwert abweichende Staaten ($p > .05$).
- Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
- 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
- 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
- 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
- 4 = Sehr hoher Anteil an Schülerinnen und Schülern mit nicht skalierbaren Leistungswerten.

Staatenvergleiche

In TIMSS 2011 bieten sich vielfältige Möglichkeiten, Deutschland mit anderen Staaten oder Regionen zu vergleichen. Inwieweit Vergleiche sinnvoll oder aufschlussreich sind, hängt von einer Reihe von inhaltlichen Aspekten ab. Einen Überblick zu den Bildungssystemen der einzelnen Teilnehmerstaaten bietet die von der internationalen Studienleitung veröffentlichte Enzyklopädie (Mullis et al., 2012c). Darüber hinaus sollten die in Abschnitt 4.3.2 beschriebenen Besonderheiten der einzelnen Stichproben bei der Interpretation von Ergebnissen nicht unberücksichtigt bleiben.

Ein Vergleich mit einzelnen Regionen wie den Benchmark-Teilnehmern (vgl. Abschnitt 4.2.1) erscheint wenig sinnvoll, da hier einzelne Besonderheiten eines Schulwesens mehr in den Vordergrund treten als bei einem ganzen Staat. Aus diesem Grund hat die internationale Studienleitung in Absprache mit den nationalen Studienleitungen wie auch in TIMSS 2007 Benchmark-Teilnehmer getrennt ausgewiesen.

Ebenso dürfte ein Vergleich mit Staaten, deren kultureller und ökonomischer Hintergrund sich stark von dem Deutschlands unterscheidet, wenig ergiebig sein. Aus diesem Grund wurden auch schon für die Berichtslegung zu TIMSS 2007 Vergleichsgruppen gebildet (Bos et al., 2008). Der vorliegende Bericht schließt sich dieser Vorgehensweise an und wählt die Teilnehmerstaaten, die der *Europäischen Union* (VG_{EU}) beziehungsweise der *Organisation for Economic Co-operation and Development* (VG_{OECD}) angehören, als Vergleichsgruppen (siehe Tabelle 2.2).

Die Ergebnisdarstellung in den Kapiteln dieses Bandes erfolgt folgendermaßen: In den Kapiteln 3 und 4 werden zunächst für alle Teilnehmerstaaten und Regionen, für die Teilnehmer mit der 6. Jahrgangsstufe sowie für die Benchmark-Teilnehmer die Ergebnisse auf den Gesamtskalen für Mathematik und Naturwissenschaften dargestellt. Die Teilnehmer mit der 6. Jahrgangsstufe

sowie die Benchmark-Teilnehmer werden graphisch nach unten abgesetzt berichtet. Auch fließen ihre Leistungswerte nicht in die Berechnung des internationalen Mittelwerts ein. In allen weiteren Abbildungen werden jene Teilnehmer dargestellt, für die mindestens eines der folgenden drei Kriterien gilt: (1) Mitglied der EU, (2) Mitglied der OECD und (3) signifikant bessere Leistungen als Deutschland auf den Gesamtskalen.

Literatur

- Baumert, J., Bos, W. & Lehmann, R. (Hrsg.). (2000). *TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (2 Bände). Opladen: Leske + Budrich.
- Baumert, J., Rainer, L., 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.
- Bonsen, M., Lintorf, K., Bos, W. & Frey, K. A. (2008). TIMSS 2007 Grundschule – Eine Einführung in die Studie. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 19–48). Münster: Waxmann.
- Boomsma, A., van Duijn, M. A. J. & Snijders, T. A. B. (Hrsg.). (2000). *Essays on item response theory*. New York: Springer.
- Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.). (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R. & Walther, G. (Hrsg.). (2004). *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Valtin, R. & Walther, G. (Hrsg.). (2005). *IGLU. Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien*. Münster: Waxmann.
- 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. & Voss, A. (2008). Empirische Schulentwicklung auf Grundlage von Lernstandserhebung. *Die Deutsche Schule*, 100 (4), 449–458.
- Foy, P. (2012). *TIMSS and PIRLS 2011 Counterbalanced data collection design*. 8th Meeting of TIMSS and PIRLS 2011 National Research Coordinators, Singapur.
- Foy, P., Brossman, B. & Galia, J. (2012). Scaling TIMSS and PIRLS 2011 achievement data. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter <http://timssandpirls.bc.edu/methods/index.html>.
- Goy, M., van Ackeren, I. & Schwippert, K. (2008). Ein halbes Jahrhundert internationale Schulleistungsstudien. Eine systematisierende Übersicht. *Tertium Comparationis*, 14 (1), 77–107.
- IEA (The International Association for the Evaluation of Educational Achievement). (2012). *Studies*. Zugriff am 16.11.2012 unter <http://www.iea.nl/studies.html>.
- Joncas, M. & Foy, P. (2011). Sample design in TIMSS and PIRLS. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter http://timssandpirls.bc.edu/methods/pdf/TP_Sampling_Design.pdf.
- Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement*, 18 (1), 1–11.
- Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking. Methods and practices* (2. Aufl.). New York: Springer.

- Leung, F. K. S. (2010). IEA studies in mathematics and science. In P. L. Peterson, E. L. Baker & B. McGaw (Hrsg.), *International encyclopedia of education* (3. Aufl., S. 650–655). Oxford: Elsevier.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Martin, M. O. & Kelly, D. L. (Hrsg.). (1998). *TIMSS. Third International Mathematics and Science Study. Technical report. Volume III: Implementation and analysis. Final year of secondary school (Population 3)*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O. & Mullis, I. V. S. (2011a). Mathematics and science development details. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter <http://timssandpirls.bc.edu/methods/t-instrument.html>.
- Martin, M. O. & Mullis, I. V. S. (2011b). TIMSS 2011 field test of mathematics and science items. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter http://timssandpirls.bc.edu/methods/pdf/MP_FieldTest_T11.pdf.
- Martin, M. O. & Mullis, I. V. S. (2011c). TIMSS 2011 item release plan. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter http://timss.bc.edu/methods/pdf/MP_ReleasePlan_T11.pdf.
- Martin, M. O. & Mullis, I. V. S. (2011d). TIMSS 2011 Item writing guidelines. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter http://timssandpirls.bc.edu/methods/pdf/T11_Item_writing_guidelines.pdf.
- Martin, M. O. & Mullis, I. V. S. (2011e). TIMSS 2011 languages used for the assessment instruments. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter http://timss.bc.edu/methods/pdf/Languages_Assessment_T11.pdf.
- Martin, M. O. & Mullis, I. V. S. (2011f). TIMSS 2011 Number of items in the mathematics and science assessments. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter http://timssandpirls.bc.edu/methods/pdf/MP_Assessment_T11.pdf.
- Martin, M. O. & Mullis, I. V. S. (2011g). TIMSS and PIRLS 2011 field test analysis of background questionnaire scale items. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter http://timss.bc.edu/methods/pdf/MP_FT-AnalysisScale_TP11.pdf.
- Martin, M. O. & Mullis, I. V. S. (Hrsg.). (2011h). *TIMSS and PIRLS methods and procedures*. Zugriff am 18.10.2012. Zugriff am 16.11.2012 unter <http://timssandpirls.bc.edu/methods/index.html>.
- Martin, M. O., Mullis, I. V. S., Foy, P., Olson, J. F., Erberer, E., Preuschoff, C. & Galia, J. (2008). *TIMSS 2007. International science report: findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Foy, P. & Stanco, G. M. (2012). *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56, 177–196.
- Mislevy, R. J., Beaton, A. E., Kaplan, B. & Sheehan, K. M. (1992). Estimation population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29 (2), 133–161.
- Mullis, I. V. S., Drucker, K. T., Preuschoff, C., Arora, A. & Stanco, G. M. (2011). Assessment framework and instrument development. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter http://timss.bc.edu/methods/pdf/TP_Instrument_Devel.pdf.
- Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1998). *Mathematics achievement in the primary school years. IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.

- Mullis, I. V. S., Martin, M. O., Foy, P. & Arora, A. (2012a). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J. F., Preuschoff, C., Erberer, E., Arora, A. & Galia, J. (2008). *TIMSS 2007. International mathematics report: findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J. & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Minnich, C. A., Stanco, G. M., Arora, A., Centurino, V. A. S. & Castle, C. E. (Hrsg.). (2012c). *TIMSS 2011 Encyclopedia. Education policy and curriculum in mathematics and science, Volumes 1 and 2*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A. & Eberber, E. (2005). *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y. & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Muraki, E. & Bock, D. (1999). *PARSCALE 3.5: IRT item analysis and test scoring for rating-scale data* [Computer software]. Chicago, IL: Scientific Software, Inc.
- Prenzel, M., Geiser, H., Langeheine, R. & Lobemeier, K. (2003). Das naturwissenschaftliche Verständnis am Ende der Grundschule. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 143–187). Münster: Waxmann.
- Robitaille, D. F., Schmidt, W. H., Raizen, S., McKnight, C., Britton, E. & Nicol, C. (Hrsg.). (1993). *TIMSS monograph no. 1. Curriculum frameworks for mathematics and science*. Vancouver, BC: Pacific Educational Press.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley + Sons.
- Sheehan, K. M. (1985). *M-GROUP. Estimation of group effects in multivariate models* [Computer software]. Princeton, NJ: Educational Testing Service.
- UIS (UNESCO Institute for Statistics). (2006). *International Standard Classification of Education. ISCED 1997*. Zugriff am 09.08.2012 unter <http://www.uis.unesco.org/Library/Documents/isced97-en.pdf>.
- van der Linden, W. & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.
- Walther, G., Geiser, H., Langeheine, R. & Lobemeier, K. (2003). Mathematische Kompetenzen am Ende der vierten Jahrgangsstufe. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 189–226). Münster: Waxmann.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427–450.
- Wendt, H., Bos, W. & Goy, M. (2011). On applications of Rasch models in international-comparative large-scale assessments: a historical review. *Educational Research and Evaluation*, 17 (6), 419–446.
- Wu, M., Adams, R., Wilson, M. R. & Haldane, S. (2007). *ACER ConQuest 2.0. Generalised item response modelling software* [Computer software]. Camberwell: Acer Press.
- Yu, A. & Ebbs, D. (2011). Translation and translation verification. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter http://timssandpirls.bc.edu/methods/pdf/TP_Translation_Verif.pdf.



Kapitel III

Mathematische Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse

Christoph Selter, Gerd Walther, Jan Wessel und Heike Wendt

1 Einleitung

Die Institution Grundschule hat nicht nur den Auftrag, die Persönlichkeitsentwicklung von Heranwachsenden zu begleiten, sondern auch anschlussfähige Bildung in zentralen Kulturtechniken wie Schreiben, Lesen und Mathematik zu vermitteln. Für den Bereich Mathematik mündete eine jahrzehntelange Diskussion zur Frage nach den zentralen Zielstellungen eines grundlegenden Mathematikunterrichts schließlich in eine breit akzeptierte Konzeption von mathematischer Grundbildung (Winter, 1995). Hiernach soll der Mathematikunterricht stufenübergreifend Erfahrungen in folgenden, vielfältig miteinander verflochtenen Bereichen ermöglichen:

- Mathematik als eine Welt *sui generis* mit ihren eigenen Konzepten, Begriffen, Darstellungen, Theorien,
- Anwendung von Mathematik zur Bearbeitung insbesondere außermathematischer Fragestellungen sowie
- Mathematik als Herausforderung und Schule des Denkens beim Bearbeiten von außer- und innermathematischen Problemen.

Lange standen hauptsächlich Fragen der Konkretisierung und Implementierung dieser Rahmenkonzeption in Lehrplänen, in Schulbüchern und im Unterricht im Vordergrund. Seit Mitte der 1990er Jahre geriet zusehends die Frage in den Fokus, über welche Kompetenzen die Schülerinnen und Schüler in diesen Bereichen tatsächlich verfügen (Output-Orientierung) und welche Konsequenzen für das Bildungssystem aus den gewonnenen Leistungsergebnissen zu ziehen sind.

Mit dem Konstanzer Beschluss vom 24. Oktober 1997 hat die *Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland* (KMK) auf die im selben Jahr publizierten, 1995 im Rahmen der *Third*

International Mathematics and Science Study (TIMSS) erhobenen und überraschend mittelmäßigen Ergebnisse der Schülerinnen und Schüler der Sekundarstufen I und II in Deutschland (Baumert et al., 1997) reagiert und Qualitätssicherung im deutschen Schulwesen zu einem zentralen Thema gemacht. Es wurde unter anderem beschlossen, künftig an international-vergleichenden Schulleistungsstudien teilzunehmen, um weitere Erkenntnisse über Kompetenzen von Schülerinnen und Schülern an zentralen Schnittstellen im Schulsystem zu erhalten.

Seit 2000 nimmt die Bundesrepublik Deutschland daher an dem im Dreijahresrhythmus stattfindenden *Programme for International Student Assessment* (PISA) für die Sekundarstufe I teil, seit 2001 an der *Internationalen Grundschul-Lese-Untersuchung* (IGLU)/*Progress in International Reading Literacy Study* (PIRLS), die im Fünfjahresabstand durchgeführt wird, und seit 2007 für die Grundschule im Vierjahreszyklus an TIMSS – mittlerweile ein Akronym für *Trends in International Mathematics and Science Study*.

Als eine weitere Maßnahme hat die KMK nationale Bildungsstandards entwickelt, die für den Grundschulbereich und die Fächer Deutsch und Mathematik seit 2004 vorliegen. Sie beinhalten definierte Leistungserwartungen als Leistungsstandards, definierte Qualitätsvorstellungen und Normen für Unterrichtsgestaltung sowie Schulentwicklung und konkretisieren damit bundesweit verbindlich den Bildungsauftrag, den Schulen zu erfüllen haben.

Die Einhaltung dieser Standards wird landesweit und auch bundesländerübergreifend überprüft. Hierzu wurde am 4. Juni 2004 von der KMK die Einrichtung eines bundesweit tätigen, von den Ländern der Bundesrepublik Deutschland gemeinsam getragenen *Instituts zur Qualitätsentwicklung im Bildungswesen* (IQB) beschlossen. Das IQB wurde Anfang Juni 2006 mit der von der KMK verabschiedeten *Gesamtkonzeption für Maßnahmen zur Feststellung der Leistungsfähigkeit des Bildungssystems und der Schulen* mit der Aufgabe betraut, zentral und länderübergreifend die Einhaltung der Bildungsstandards in der Sekundarstufe I zu überprüfen – im Jahr 2009 für die Fächer Deutsch, Englisch und Französisch (vgl. Köller, Knigge & Tesch, 2010) sowie im Jahr 2012 für die Fächer Mathematik, Biologie, Chemie und Physik.

Nationale Vergleichsstudien zur Grundschule: Eine ländervergleichende Untersuchung zur Einhaltung der Bildungsstandards für den Primarbereich (4. Jahrgangsstufe) wurde parallel zu den Erhebungen von IGLU 2011 und TIMSS 2011 von Mai bis Juni 2011 für die Fächer Mathematik und Deutsch an rund 1300 Grund- und Förderschulen durchgeführt (vgl. Stanat, Pant, Böhme & Richter, 2012). Die Hauptergebnisse des Ländervergleichs für den Bereich Mathematik werden im Folgenden knapp skizziert (vgl. Haag & Roppelt, 2012). Bei der Interpretation der Ergebnisse – im Vergleich zu den in diesem Kapitel beschriebenen Ergebnissen – ist zu beachten, dass trotz gleicher Metrik die Punktwerte anders und ohne Bezug zu TIMSS ermittelt und auch die Kompetenzstufen (vgl. Abschnitt 2) unterschiedlich definiert worden sind. Ein direkter Vergleich der Punktwerte von TIMSS 2011 mit den Punktwerten des Ländervergleichs ist somit nicht möglich, eine interpretative Bezugnahme jedoch schon.

- Es ließ sich mit Bayern, Sachsen, Sachsen-Anhalt und Baden-Württemberg eine eng zusammenliegende Gruppe von vier Ländern identifizieren, in denen die mittleren Kompetenzstände der Schülerinnen und Schüler signifikant über dem deutschen Mittelwert von 500 Punkten liegen. Mit gewissem Abstand folgt eine Gruppe von sechs Ländern, deren Leistungsmittelwerte

sich vom Mittelwert für Deutschland nicht signifikant unterscheiden. Es folgen Brandenburg, Schleswig-Holstein und Hessen sowie die Stadtstaaten Hamburg, Bremen und Berlin mit Ergebnissen statistisch signifikant unterhalb des deutschen Mittelwerts.

- Im Vergleich des 5. und des 95. Perzentiles (vgl. Abschnitt 5.1) sind vergleichsweise geringe Differenzen von weniger als 315 Punkten in Sachsen, Schleswig-Holstein, Niedersachsen und Nordrhein-Westfalen zu konstatieren. Relativ große Differenzwerte mit mehr als 335 Punkten bestehen hingegen in Mecklenburg-Vorpommern, Sachsen-Anhalt und Bremen.
- Der Anteil der Schülerinnen und Schüler, die mindestens den Regelstandard erreichten (Kompetenzstufe III und höher, vgl. Abschnitt 2), ist landesabhängig sehr unterschiedlich. Während in Bremen und Berlin nur jeweils etwa die Hälfte der Viertklässlerinnen und Viertklässler entsprechende Kompetenzen zeigte, beträgt der entsprechende Prozentsatz in Baden-Württemberg, Sachsen-Anhalt, Sachsen und Bayern zwischen 73 und 77 Prozent.
- Erheblich fielen die Länderunterschiede in Bezug auf den Anteil der Kinder aus, die den Mindeststandard (Kompetenzstufe II) verfehlen. In Berlin und Bremen einerseits beträgt der Anteil 25 Prozent; in Baden-Württemberg, Sachsen-Anhalt, Bayern und Sachsen hingegen sind es 10 Prozent oder weniger.
- Kompetenzen, die der höchsten Kompetenzstufe V (Optimalstandard) entsprechen, zeigten weniger als 10 Prozent der Kinder in Bremen und Berlin, jedoch rund 21 Prozent der Kinder in Sachsen, Bayern und Sachsen-Anhalt.
- Für die mathematische Kompetenz am Ende der vierten Jahrgangsstufe ist ein statistisch signifikanter Vorsprung zugunsten der Jungen von 16 Punkten zu verzeichnen, was in etwa dem Lernzuwachs eines Fünftel Schuljahres entspricht.
- Schülerinnen und Schüler ohne Zuwanderungshintergrund erreichten im Durchschnitt 515 Punkte. Für Kinder mit zwei im Ausland geborenen Elternteilen ergab sich ein Kompetenzrückstand von 54 Punkten, der wesentlich stärker ausgeprägt ist als der Rückstand von Kindern mit nur einem im Ausland geborenen Elternteil (30 Punkte).
- Betrachtet man den Zusammenhang zwischen dem Anteil fachfremd unterrichtender Lehrkräfte und den durchschnittlich von Schülerinnen und Schülern erreichten Kompetenzen, so wird deutlich, dass Klassen, die von einer Lehrkraft mit einer Lehrbefähigung für Mathematik unterrichtet wurden, im Durchschnitt um 18 Punkte besser abschnitten.

Da in diesem Kapitel über die Ergebnisse einer *internationalen* Vergleichsstudie berichtet wird, werden im Folgenden in der gebotenen Kürze auch zentrale Ergebnisse aktueller international-vergleichender Schulleistungsuntersuchungen für das Fach Mathematik berichtet.

Internationale Vergleichsstudien zur Sekundarstufe I: PISA untersucht im internationalen Vergleich der Mitgliedsstaaten der *Organisation for Economic Co-Operation and Development* (OECD) alltags- und berufsrelevante Kompetenzen von fünfzehnjährigen Schülerinnen und Schülern in den Bereichen Lesen, Mathematik und Naturwissenschaften. Die Länder der Bundesrepublik Deutschland haben sich seit 2000 an fünf Zyklen von PISA (2000, 2003, 2006, 2009, 2012) beteiligt. Mit der Teilnahme Deutschlands an PISA 2009 bot sich die Chance, bezüglich der Leistungsfähigkeit der Sekundarstufe I und der Wirkungen

der auf sie bezogenen Reformbemühungen ‚nach einem Jahrzehnt‘ Bilanz zu ziehen (Klieme et al., 2010).

Für den Kompetenzbereich Mathematik galt allerdings, dass er in den Jahren 2000 und 2009 als Nebendomäne und damit lediglich mit einer vergleichsweise geringen Anzahl an Testaufgaben untersucht wurde. Infolgedessen konnte keine nach Inhaltsbereichen differenzierende Analyse der Schülerkompetenzen erfolgen. Auch eine Betrachtung von Trends in mathematischen Kompetenzen wird erst im Rahmen der Berichterstattung zu PISA 2012 möglich sein, da bei PISA 2012 wie auch in PISA 2003 Mathematik als Hauptdomäne mit einem umfassenderen Aufgabenpool untersucht wird. Mit der Berichterstattung der Ergebnisse ist ab Ende 2013 zu rechnen.

Im Rahmen der Berichterstattung zu PISA 2009 konnte – neben einer international-vergleichenden und für national relevante Teilgruppen vertiefenden Analyse der mathematischen Leistungsstände von Fünfzehnjährigen in Deutschland – zumindest eine vorsichtige Trendberichterstattung für die Jahre 2003–2009 erfolgen. Hervorzuheben sind insbesondere die folgenden Befunde:

- Das mittlere mathematische Leistungsniveau der fünfzehnjährigen Schülerinnen und Schüler in Deutschland hat sich von PISA 2003 zu PISA 2009 statistisch signifikant um 10 Punkte verbessert (Frey, Heinze, Mildner, Hochweber & Asseburg, 2010, S. 162). Mit einem Leistungsmittelwert von 513 Punkten gehört Deutschland 2009 damit erstmalig zu der Gruppe jener Staaten, die signifikant über dem OECD-Durchschnitt (496 Punkte) liegen. Allerdings liegt die Streuung der mathematischen Leistungsmittelwerte der Schülerinnen und Schüler in Deutschland mit einer Standardabweichung von 98 Punkten weiterhin signifikant über der mittleren Streuung aller OECD-Staaten (92 Punkte).
- Der Abstand von rund 30 Punkten zur mittleren mathematischen Kompetenz in den leistungsstärksten Staaten Korea (546 Punkte) und Finnland (541 Punkte) ist jedoch nach wie vor erheblich. Diese Differenz entspricht dem Kompetenzzuwachs von rund einem Schuljahr (Frey et al., 2010, S. 162).
- Besonders bedeutsam ist, dass der Anteil der Jugendlichen mit sehr geringer mathematischer Kompetenz (unter Kompetenzstufe I) von PISA 2003 zu PISA 2009 signifikant reduziert werden konnte (Frey et al., 2010, S. 165). Eine schwache Kompetenz (unter oder auf Kompetenzstufe I) zeigen bei PISA 2009 allerdings insgesamt noch knapp 19 Prozent der Jugendlichen in Deutschland. Dieser Anteil ist jedoch signifikant geringer als der OECD-Durchschnitt (22 Prozent).
- Auf der höchsten Kompetenzstufe befinden sich in den OECD-Staaten durchschnittlich 3.1 Prozent der fünfzehnjährigen Schülerinnen und Schüler. In Deutschland sind es mit 4.6 Prozent signifikant mehr als der OECD-Durchschnitt; der Anteil hat sich seit PISA 2003 nicht wesentlich verändert (Frey et al., 2010, S. 165).

Insgesamt ist somit einerseits beim Vergleich der Untersuchungen der Jahre 2003 und 2009 eine leicht positive Entwicklung zu konstatieren. Andererseits werden auch Entwicklungspotentiale offenkundig, insbesondere im Bereich der Förderung von leistungsstarken wie leistungsschwachen Schülerinnen und Schülern.

Internationale Vergleichsstudien zur Grundschule: Eine Datenbasis, die es nach einem Jahrzehnt auch für die Grundschule erlauben würde, die Leistungsfähigkeit der Schülerinnen und Schüler sowie die Wirkungen der Reformbemühungen

zu bilanzieren, liegt für die Bereiche Mathematik und Naturwissenschaften noch nicht vor. In der TIMS-Studie von 1995 wurden zwar erstmalig in einem breiten internationalen Vergleich die Mathematikleistungen von Schülerinnen und Schülern am Ende der dritten und vierten Jahrgangsstufe erhoben; die Bundesrepublik Deutschland hatte sich jedoch an der Grundschülerhebung von TIMSS 1995 nicht beteiligt.

Im Jahr 2001 wurden dann erstmals in einem Erweiterungsteil von IGLU/PIRLS (IGLU-E) die Mathematik- und Naturwissenschaftsleistungen von Viertklässlerinnen und Viertklässlern in Deutschland untersucht (vgl. Bos et al., 2003). Der Test war so konzipiert, dass eine vorsichtige Einordnung der Ergebnisse von Schülerinnen und Schülern in Deutschland in die damals gut fünf Jahre alte internationale Leistungsskala von TIMSS 1995 möglich war (Walther, Geiser, Langeheine & Lobemeier, 2003). IGLU-E 2001 war damit die erste umfassende Studie, die für die Bundesrepublik Deutschland mathematische Kompetenzen von Kindern im Grundschulalter untersucht hat.

Dabei stellte sich für den Bereich Mathematik heraus, dass die Leistung der Viertklässlerinnen und Viertklässler in Deutschland im oberen internationalen Mittelfeld platziert war. Allerdings erwies sich mit knapp 17 Prozent der Anteil von Schülerinnen und Schülern mit sehr großen Defiziten als vergleichsweise hoch und mit knapp 7 Prozent der Anteil von leistungsstarken Schülerinnen und Schülern, die in der Lage waren, selbstständig auch anspruchsvolle mathematische Probleme zu lösen, als vergleichsweise gering. Zudem ließen sich sowohl bedeutsame Leistungsunterschiede zwischen Mädchen und Jungen als auch ausgeprägte Zusammenhänge zwischen der sozialen Herkunft und den mathematischen Kompetenzen von Schülerinnen und Schülern am Ende der Grundschulzeit feststellen.

Im Jahr 2007 nahm die Bundesrepublik Deutschland erstmals regulär an TIMSS für die vierte Jahrgangsstufe teil (Bos et al., 2008). Mit dieser Studie konnten in weiten Teilen die Ergebnisse aus IGLU-E bestätigt werden. Hervorzuheben sind insbesondere die folgenden Befunde (Walther, Selter, Bonsen & Bos, 2008, S. 59 ff.):

- Im internationalen Vergleich lag die Mathematikleistung der Viertklässlerinnen und Viertklässler in Deutschland mit einem Mittelwert von 525 Punkten über dem internationalen Mittelwert von 473 Punkten und über den Leistungsmittelwerten der an TIMSS 2007 teilnehmenden Mitgliedsstaaten der OECD und der *Europäischen Union* (EU).
- Zu den Staaten an der Leistungsspitze, wie zum Beispiel Japan, bestand jedoch ein beachtlicher Abstand.
- Trotz einer als erfreulich einzuschätzenden relativen Leistungshomogenität zeigte sich in Bezug auf das Leistungsniveau, dass jedes sechste Kind in Deutschland am Ende der Grundschulzeit nur über elementare mathematische Fertigkeiten verfügte und 4 Prozent der Kinder dieses Leistungsniveau nicht erreichten.
- Der Anteil der leistungsstarken Schülerinnen und Schüler fiel in Deutschland vergleichsweise gering aus. Er war beispielsweise in Japan fast viermal so groß wie in Deutschland.
- In Bezug auf die Inhaltsbereiche ließen sich für die Schülerinnen und Schüler in Deutschland Stärken im Bereich *Umgang mit Daten* und relative Schwächen im Bereich *Arithmetik* feststellen. Während in den beiden kognitiven Anforderungsbereichen *Reproduzieren* und *Problemlösen* die Grundschülerinnen und Grundschüler in Deutschland mit ihren mittlere-

ren Leistungen knapp über dem internationalen Mittelwert lagen, fielen die Leistungen im Bereich *Anwenden* deutlich schwächer aus.

- Der überwiegende Teil der deutschen Grundschülerinnen und Grundschüler verfügte über eine positive Einstellung zur Mathematik.
- Auch das mathematische Fähigkeitsselbstkonzept der Grundschülerinnen und Grundschüler in Deutschland erwies sich als positiv.
- Zudem ließen sich bedeutsame Leistungsunterschiede zwischen Jungen und Mädchen als auch ausgeprägte Zusammenhänge zwischen der sozialen Herkunft und den mathematischen Kompetenzen am Ende der Grundschulzeit feststellen (Bos et al., 2008).

Auf der Grundlage der berichteten Befunde werden nachfolgend die Ergebnisse der deutschen Beteiligung an TIMSS 2011 für den Bereich der Mathematik dargestellt. Hierzu werden in Abschnitt 2 die *Mathematische Grundbildung in Deutschland*, in Abschnitt 3 die *TIMSS-Rahmenkonzeption zur Erfassung mathematischer Kompetenzen* und in Abschnitt 4 der *Test zur Erfassung der mathematischen Kompetenzen* beschrieben, bevor in Abschnitt 5 die *Ergebnisse* entlang folgender Fragen berichtet werden:

- Welche Leistungsmittelwerte zeigen Schülerinnen und Schüler in Deutschland am Ende der vierten Jahrgangsstufe im internationalen Vergleich? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und TIMSS 2011? (Abschnitt 5.1)
- Wie lassen sich die TIMSS-Leistungskennwerte auf Kompetenzstufen einordnen? Wie groß sind die Gruppen der auffällig leistungsschwachen und leistungsstarken Kinder? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und TIMSS 2011? (Abschnitt 5.2)
- Welche Ergebnisse erzielen die Viertklässlerinnen und Viertklässler in Deutschland in den drei mathematischen Inhaltsbereichen *Arithmetik*, *Geometrie/Messen* und *Umgang mit Daten*? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und TIMSS 2011? (Abschnitt 5.3)
- Welche Ergebnisse erzielen die Grundschulkinder in Deutschland in den drei kognitiven Anforderungsbereichen der TIMS-Studie *Reproduzieren*, *Anwenden* und *Problemlösen*? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und TIMSS 2011? (Abschnitt 5.4)
- Welche Einstellungen zur Mathematik und welches mathematikbezogene Selbstkonzept lassen Schülerinnen und Schüler am Ende der vierten Jahrgangsstufe erkennen? Gibt es einen Zusammenhang zwischen den Einstellungen zur Mathematik und der mathematischen Kompetenz beziehungsweise zwischen dem mathematikbezogenen Selbstkonzept und der mathematischen Kompetenz? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und TIMSS 2011? (Abschnitt 5.5)

Ergebnisse aus TIMSS 2011 zu Unterschieden in der mathematischen Kompetenz in Abhängigkeit von Geschlecht, sozialer Herkunft oder Migrationshintergrund werden in anderen Kapiteln dieses Bandes beschrieben (Kapitel 6, 7 und 8).

2 Mathematische Grundbildung in Deutschland

Um die mathematischen Leistungsergebnisse der Viertklässlerinnen und Viertklässler in Deutschland einordnen zu können, ist die Passung zwischen der Rahmenkonzeption von TIMSS 2011 und den Konzeptionen zum Lehren und Lernen von Mathematik in Deutschland zu überprüfen. Eine solche Analyse soll im Folgenden in zwei Schritten erfolgen.

Zunächst werden in diesem Abschnitt die curricularen Grundlagen mathematischer Grundbildung an deutschen Grundschulen erläutert, bevor in Abschnitt 3 die TIMSS-Rahmenkonzeption beschrieben wird. In Abschnitt 4.2 werden schließlich Gemeinsamkeiten und Unterschiede zwischen beiden Konzeptionen herausgearbeitet und es wird die Passung zwischen der Rahmenkonzeption von TIMSS 2011 und den Vorstellungen zur mathematischen Grundbildung in Deutschland eingeschätzt. Die Beschreibung erfolgt in Referenz zu den Bildungsstandards der KMK (2005), zu deren Implementation und Anwendung in Lehrplanarbeit, Schulentwicklung sowie Lehrerbildung und Lehrerfortbildung sich alle Länder der Bundesrepublik Deutschland verpflichtet haben. Die Bildungsstandards stellen – trotz einer weiterhin bestehenden länderübergreifenden Variabilität in der Gestaltung der einzelnen Mathematiklehrpläne für die Primarstufe – den besten Bezugspunkt dar, um die Konzeption der mathematischen Grundbildung in Deutschland zu beschreiben.

Die Bildungsstandards beschreiben mathematische Kompetenzen, die Schülerinnen und Schüler am Ende der vierten Jahrgangsstufe erreicht haben sollen. Unterschieden wird zwischen inhaltsbezogenen und allgemeinen mathematischen Kompetenzen. Die Kompetenzerwartungen sind in drei Anforderungsbereichen auf zentrale mathematische Konzepte bezogen, die sogenannten mathematischen Leitideen, und in einem feingliedrigen System von Standards ausformuliert (vgl. Walther, Selter & Neubrand, 2007). Die durch die Bildungsstandards zum Ausdruck kommende Konzeption einer mathematischen Grundbildung ist jedoch keine Erfindung des 21. Jahrhunderts. Wie in Walther et al. (2007) ausgeführt wird, geht sie auf eine lange Tradition der Auseinandersetzung in der Mathematikdidaktik mit Bildungszielen beziehungsweise zentralen mathematischen Anforderungen zurück, die zentral durch wegweisende Arbeiten von Winter (Winter, 1975, 1995) geprägt wurden.

Inhaltsbezogene mathematische Kompetenzen: Die inhaltsbezogenen mathematischen Kompetenzen sind mit den folgenden fünf mathematischen Leitideen (KMK, 2005, S. 8) verbunden:

- Zahlen und Operationen,
- Raum und Form,
- Muster und Strukturen,
- Größen und Messen sowie
- Daten, Häufigkeit und Wahrscheinlichkeit.

Zur Leitidee *Zahlen und Operationen* beispielsweise sind folgende drei inhaltsbezogene Kompetenzen formuliert:

- Zahldarstellungen und Zahlbeziehungen verstehen,
- Rechenoperationen verstehen und beherrschen sowie
- in Kontexten rechnen.

Diese Kompetenzen wiederum sind in Form von Standards konkretisiert. Die Kompetenz *Zahldarstellungen und Zahlbeziehungen verstehen* etwa ist wie folgt gegliedert:

- den Aufbau des dezimalen Stellenwertsystems verstehen;
- Zahlen bis 1 000 000 auf verschiedene Weise darstellen und zueinander in Beziehung setzen;
- sich im Zahlenraum bis 1 000 000 orientieren (z.B. Zahlen der Größe nach ordnen, runden).

Allgemeine mathematische Kompetenzen: Den fünf allgemeinen mathematischen Kompetenzen, die auch synonym als prozessbezogene Kompetenzen bezeichnet werden, kommt eine herausragende Rolle bei der Entwicklung von auf Verständnis gegründeten inhaltlichen mathematischen Kompetenzen zu (KMK, 2005, S. 7 ff.). Die Entwicklung der mathematischen Grundbildung bei Schülerinnen und Schülern, die den Wortteil ‚Bildung‘ zurecht enthält, hängt wesentlich davon ab, in welchem Maße für deren Erreichung im Unterricht Anlässe geschaffen werden,

- selbst oder gemeinsam *Probleme* mathematisch zu lösen;
- über das Verstehen und das Lösen von Aufgaben zu *kommunizieren*;
- über das Zutreffen von Vermutungen und das Erkennen von mathematischen Zusammenhängen zu *argumentieren*;
- Sachsituationen in der Sprache der Mathematik zu *modellieren* und
- für die Bearbeitung von Problemen geeignete *Darstellungen* zu *entwickeln* oder *auszuwählen* (vgl. Walther et al., 2007, S. 26 ff.).

Auch die allgemeinen mathematischen Kompetenzen sind in Form von Standards konkretisiert. Für das *Kommunizieren* beispielsweise lauten diese:

- eigene Vorgehensweisen beschreiben, Lösungswege anderer verstehen und gemeinsam darüber reflektieren;
- mathematische Fachbegriffe und Zeichen sachgerecht verwenden sowie
- Aufgaben gemeinsam bearbeiten, dabei Verabredungen treffen und einhalten.

Bedeutsam ist insgesamt das große Gewicht, das die Bildungsstandards der Entwicklung der allgemeinen mathematischen Kompetenzen in Verbindung mit substantiellen mathematischen Inhalten im Unterricht zuschreiben. Die Formulierung von kognitiven Anforderungsbereichen, die über rein reproduktive Tätigkeiten hinausreichen, gibt Orientierung sowohl für die Aufgabenauswahl im Unterricht als auch für die Konstruktion von Testaufgaben. Die Idee der kognitiven Anforderungsbereiche findet sich in etwas anderer Form auch in der TIMSS-Rahmenkonzeption (vgl. Abschnitt 3.2).

Anforderungsbereiche: Die Bildungsstandards formulieren drei Anforderungsbereiche, die die kognitiven Anforderungen an Schülerinnen und Schüler bei der Bearbeitung von Aufgaben beschreiben (KMK, 2005, S. 13):

- Anforderungsbereich I – *Reproduzieren*: Das Lösen der Aufgabe erfordert Grundwissen und das Ausführen von Routinetätigkeiten.
- Anforderungsbereich II – *Zusammenhänge herstellen*: Das Lösen der Aufgabe erfordert das Erkennen und Nutzen von Zusammenhängen.
- Anforderungsbereich III – *Verallgemeinern und Reflektieren*: Das Lösen der Aufgabe erfordert komplexe Tätigkeiten wie Strukturieren, Entwickeln von Strategien, Beurteilen und Verallgemeinern.

Kompetenzstufenmodell der Bildungsstandards: Zur inhaltlichen Interpretation der bei Schulleistungsstudien ermittelten numerischen Leistungsdaten werden Kompetenzstufenmodelle verwendet. Als Grundlage für die Auswertung der seit 2011 im Fünfjahresturnus stattfindenden Ländervergleiche zur Überprüfung der Umsetzung der Bildungsstandards wurde durch das IQB ein Kompetenzstufenmodell entwickelt (KMK, 2009, S. 11; sowie Reiss, Roppelt, Haag, Pant & Köller, 2012; Reiss & Winkelmann, 2009). Die grobe Struktur des Modells ist in Tabelle 3.1 dargestellt. Das Modell ist global definiert und auf alle inhaltlichen Kompetenzen anwendbar. Die allgemeinen mathematischen Kompetenzen gehen in die Beschreibung der Leistungen auf den verschiedenen Stufen ein. Wie auch in international-vergleichenden Schulleistungsuntersuchungen üblich wurde ein Maßstab mit einem Mittelwert von 500 und einer Standardabweichung von 100 für die vierte Jahrgangsstufe definiert. Allerdings sind die Punktwerte nicht direkt miteinander vergleichbar. Die Breite einer jeden Stufe beträgt 70 Punkte. Stufe I ist nach unten, Stufe V nach oben offen (KMK, 2009, S. 11).

Tabelle 3.1: IQB-Kompetenzstufenmodell für das Fach Mathematik in der Primarstufe (Jahrgangsstufe 4)

Kompetenzstufe V (ab 600): Die Maximalstandards werden erreicht. <i>Modellierung komplexer Probleme unter selbstständiger Entwicklung geeigneter Strategien.</i>
Kompetenzstufe IV (530-599): Die „Regelstandards plus“ werden erreicht. <i>Sicheres und flexibles Anwenden von begrifflichem Wissen und Prozeduren im curricularen Umfang.</i>
Kompetenzstufe III (460-529): Die Regelstandards werden erreicht. <i>Erkennen und Nutzen von Zusammenhängen in einem vertrauten (mathematischen und sachbezogenen) Kontext.</i>
Kompetenzstufe II (390-459): Die Mindeststandards werden erreicht. <i>Einfache Anwendungen von Grundlagenwissen (Routineprozeduren in einem klar strukturierten Kontext).</i>
Kompetenzstufe I (unter 390): Ziele des Unterrichts werden weitgehend verfehlt. <i>Routineprozeduren auf Grundlage einfachen begrifflichen Wissens</i>

Quelle: KMK, 2009, S. 11

3 TIMSS-Rahmenkonzeption zur Erfassung mathematischer Kompetenzen

In Kapitel 2 dieses Bandes wurden die allgemeinen theoretischen und konzeptionellen Grundlagen von TIMSS 2011 beschrieben. In diesem Abschnitt wird nun die TIMSS-Rahmenkonzeption (Mullis, Martin, Ruddock, O’Sullivan & Preuschoff, 2009) zur Erfassung mathematischer Kompetenzen am Ende der vierten Jahrgangsstufe erläutert. Die Zielsetzung einer solchen Rahmenkonzeption besteht darin, die Konstruktion von Leistungstests dahingehend zu fundieren, dass sowohl belastbare Aussagen über mathematische Kompetenzen von Schülerinnen und Schülern im internationalen Vergleich getroffen als auch differenzierte Erkenntnisse über deren Stärken und Schwächen in unterschiedlichen Teilbereichen gewonnen werden können.

Die Ansprüche der breiten Abdeckung mathematischer Kompetenzen einerseits und der Passung zu den Lehrplänen in den Teilnehmerstaaten andererseits waren bei der Entwicklung der TIMSS-Rahmenkonzeption notwendigerweise auszubalancieren. Um diesem Anspruch Rechnung zu tragen, unterscheidet die Rahmenkonzeption für die Beschreibung mathematischer Kompetenzen grundlegende Inhaltsbereiche und kognitive Anforderungsbereiche, die in den folgenden Abschnitten erläutert werden sollen.

3.1 Mathematische Inhaltsbereiche

Insbesondere in Bezug auf die inhaltliche Auswahl von Testaufgaben strebt TIMSS eine breite Kompatibilität mit den Curricula der teilnehmenden Staaten und Regionen an. Die TIMSS-Rahmenkonzeption nennt für die vierte Jahrgangsstufe drei Inhaltsbereiche, die für eine international vergleichende Betrachtung mathematischer Kompetenzen als bedeutsam erachtet werden:

- Arithmetik (*Number*),
- Geometrie/Messen (*Geometric Shapes and Measures*) sowie
- Umgang mit Daten (*Data Display*).

Jeder der drei Inhaltsbereiche ist weiter in mehrere Teilgebiete aufgeschlüsselt: *Arithmetik* beispielsweise in *Natürliche Zahlen*, *Brüche und Dezimalzahlen*, *Zahlsätze* sowie *Muster und Beziehungen*. Für die einzelnen Themengebiete werden in der Rahmenkonzeption zudem inhaltsbezogene Kompetenzen formuliert, über die Schülerinnen und Schüler am Ende der Grundschulzeit verfügen sollen (siehe Tabelle 3.2). Bei der Beschreibung und Interpretation der mathematischen Inhaltsbereiche ist jedoch zu beachten, dass – abweichend von der in Deutschland üblichen Gliederung – einerseits die Größenbereiche *Geldwerte*, *Gewichte* und *Zeitspannen* zur *Arithmetik* und andererseits *Längen*, *Flächeninhalte* und *Volumina* zur *Geometrie* gezählt werden.

Tabelle 3.2: Themengebiete der einzelnen mathematischen Inhaltsbereiche

Arithmetik	Geometrie/Messen	Umgang mit Daten
<p><i>Natürliche Zahlen</i></p> <ul style="list-style-type: none"> – Über Wissen zu Stellenwerten verfügen, einschließlich des Lesens und Schreibens von Zahlen und des Darstellens von Zahlen mit Hilfe von Worten, Zeichnungen und Symbolen – Zahlen ordnen und vergleichen – Die vier Grundrechenarten ausführen; überschlagend rechnen – Vielfache und Teiler erkennen – Aufgaben lösen, einschließlich Sachaufgaben und Aufgaben, die Messen, Geld und einfache Verhältnisse beinhalten <p><i>Brüche und Dezimalzahlen</i></p> <ul style="list-style-type: none"> – Über Bruchverständnis verfügen: Brüche als Teile eines Ganzen oder als Teil mehrerer Ganzer erkennen, am Zahlenstrahl einordnen; Brüche durch Worte, Symbole oder geometrisch darstellen – Einfache gleichwertige Brüche erkennen; einfache Brüche vergleichen und ordnen – Einfache Brüche addieren und subtrahieren – Über Dezimalzahlverständnis verfügen: Dezimalzahlen durch Worte, Symbole oder geometrisch darstellen – Dezimalzahlen addieren und subtrahieren – Aufgaben mit einfachen Brüchen und Dezimalzahlen lösen <p><i>Zahlensätze</i></p> <ul style="list-style-type: none"> – Fehlende Zahlen oder Operationen in einem Zahlensatz finden (z.B. $17 + \underline{\quad} = 29$) – Einfache Situationen mit Unbekannten durch Zahlensätze modellieren <p><i>Muster und Beziehungen</i></p> <ul style="list-style-type: none"> – Klar definierte Muster fortsetzen oder in ihnen fehlende Elemente finden; Beziehungen zwischen benachbarten Elementen in einer Folge und zwischen dem Element und seiner Platznummer beschreiben – Eine Regel für die Beziehung von Zahlenpaaren aufschreiben oder auswählen; Zahlenpaare erzeugen, die einer vorgegebenen Regel folgen (z.B.: „Multipliziere die erste Zahl mit 3 und addiere 2, um die zweite Zahl zu erhalten.“) 	<p><i>Punkte, Geraden und Winkel</i></p> <ul style="list-style-type: none"> – Längen messen und schätzen – Zueinander parallele und zueinander senkrechte Geraden identifizieren und zeichnen – Winkel der Größe nach vergleichen; Winkel zeichnen (rechte Winkel, spitze Winkel, stumpfe Winkel) – Informelle Koordinatensysteme nutzen, um Punkte in der Ebene zu lokalisieren <p><i>Zwei- und dreidimensionale Figuren</i></p> <ul style="list-style-type: none"> – Einfache geometrische Figuren identifizieren, klassifizieren und vergleichen (z.B. gemäß Form, Größe oder Eigenschaften) – Elementare Eigenschaften von geometrischen Figuren abrufen, beschreiben und nutzen, einschließlich Achsen- und Drehsymmetrie – Beziehungen zwischen zwei- und dreidimensionalen Darstellungen von Objekten erkennen – Umfang und Flächeninhalt von Quadraten und Rechtecken berechnen; Flächen- und Rauminhalte von geometrischen Figuren bestimmen und schätzen (z.B. durch Auslegen mit vorgegebenen Figuren oder Würfeln) 	<p><i>Daten lesen und interpretieren</i></p> <ul style="list-style-type: none"> – Skalierungen und Daten aus Tabellen, Piktogrammen, Säulen- und Kreisdiagrammen entnehmen – Informationen aus verbundenen Datenmengen vergleichen (z.B. gegebene Daten oder Darstellungen von Daten über Lieblingseiscremesorten: „Finde die Klassen, in der Schokolade die Lieblingsorte ist.“) – Informationen aus Diagrammen oder Tabellen nutzen, um Fragen zu beantworten, die über das bloße Entnehmen von Daten hinausgehen (z.B. Daten kombinieren, auf Daten gestützt rechnen und Schlussfolgerungen ziehen) <p><i>Ordnen und Darstellen</i></p> <ul style="list-style-type: none"> – Verschiedene Darstellungen derselben Daten vergleichen und zuordnen – Daten ordnen und darstellen mit Hilfe von Tabellen, Piktogrammen und Säulendiagrammen

3.2 Kognitive Anforderungsbereiche

Mathematische Kompetenzen werden auch in TIMSS nicht nur inhaltsbezogen, sondern auch in Bezug auf ihre relevanten Denkprozesse unterschieden.

Die Rahmenkonzeption nennt für die vierte Jahrgangsstufe drei kognitive Anforderungsbereiche, die für eine international-vergleichende Betrachtung mathematischer Kompetenzen als bedeutsam erachtet werden:

- Reproduzieren von Wissen, Fertigkeiten und Grundvorstellungen (*Knowing*),
- Anwenden von Wissen, Fertigkeiten und Grundvorstellungen beim Bearbeiten von Standardaufgaben (*Applying*) sowie
- Lösen von komplexeren Berechnungs-, Anwendungs- und Begründungsproblemen (*Reasoning*).

Mit dieser Unterscheidung von unterschiedlich komplexen kognitiven Anforderungen soll vermieden werden, dass überwiegend Aufgaben zum Einsatz kommen, die nur die Aktivierung von abrufbarem Grundwissen, die Anwendung von Grundfertigkeiten und die Nutzung von Grundvorstellungen erfordern (Mullis et al., 2009, S. 40 ff.). Welche konkreten kognitiven Leistungen den einzelnen Anforderungsbereichen zugeordnet werden, ist Tabelle 3.3 zu entnehmen.

Tabelle 3.3: Kognitive Aktivitäten der einzelnen Anforderungsbereiche

Reproduzieren	Anwenden	Problemlösen
– Standardwissen abrufen	– Mathematische Operationen auswählen	– Erkennen mathematischer Beziehungen
– Einfache mathematische Objekte und Beziehungen erkennen	– Informationen darstellen	– Verallgemeinern/Spezialisieren
– Rechnen	– Mathematische Modelle bilden/modellieren	– Integrieren und Kombinieren von Wissen
– Informationen ablesen	– Mathematische Operationen ausführen	– Begründen
– Messen	– Standardaufgaben lösen	– Nicht-Standardprobleme lösen
– Klassifizieren/Ordnen		

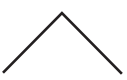



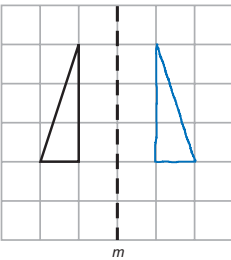






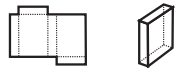


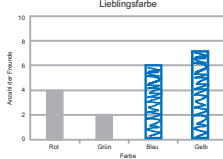

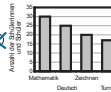
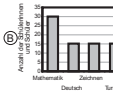
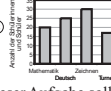
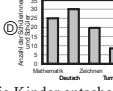
IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

3.3 Aufgabenbeispiele zu Inhaltsbereichen und kognitiven Anforderungsbereichen

In Abbildung 3.1 werden abschließend die drei Inhaltsbereiche und die drei Anforderungsbereiche anhand von Beispielaufgaben illustriert.

Abbildung 3.1: Zuordnung von Beispielaufgaben zu den Inhaltsbereichen und kognitiven Anforderungsbereichen

	Reproduzieren	Anwenden	Problemlösen																
Arithmetik	<p>Welche Zahl ist um 100 größer als 5 432?</p> <p>(A) 6 432 <input checked="" type="radio"/> (B) 5 532 (C) 5 442 (D) 5 433</p> <p>Bei dieser Aufgabe soll die Zahl angegeben werden, welche um 100 größer ist als 5 432. Um diese Aufgabe zu lösen, ist Routinewissen über das Stellenwertsystem abzurufen.</p>	<table border="1"> <tr><th colspan="2">Zutaten</th></tr> <tr><td>Eier</td><td>4</td></tr> <tr><td>Mehl</td><td>8 Tassen</td></tr> <tr><td>Milch</td><td>$\frac{1}{2}$ Tasse</td></tr> </table> <p>Bei einem Rezept für 6 Personen benötigt man die oben angegebenen Zutaten. Leo möchte dieses Rezept nur für 3 Personen machen. Trage in die Tabelle unten ein, wie viel Leo für 3 Personen braucht. Die Anzahl der Eier, die er braucht, ist schon eingetragen.</p> <table border="1"> <tr><th colspan="2">Zutaten</th></tr> <tr><td>Eier</td><td>2</td></tr> <tr><td>Mehl</td><td>4 Tassen</td></tr> <tr><td>Milch</td><td>$\frac{1}{2}$ Tasse</td></tr> </table> <p>Bei dieser Aufgabe sollen für ein Rezept Mengenangaben, die auf sechs Personen bezogen sind, für drei Personen umgerechnet werden. Zur Berechnung der fehlenden Angaben ist eine passende Operation und Vorgehensweise zu wählen.</p>	Zutaten		Eier	4	Mehl	8 Tassen	Milch	$\frac{1}{2}$ Tasse	Zutaten		Eier	2	Mehl	4 Tassen	Milch	$\frac{1}{2}$ Tasse	<p>Der Maßstab auf einer Landkarte zeigt an, dass 1 Zentimeter auf der Karte 4 Kilometer in der Wirklichkeit entspricht. Die Entfernung zwischen zwei Städten auf der Karte ist 8 Zentimeter. Wie viele Kilometer sind die zwei Städte voneinander entfernt?</p> <p>(A) 2 (B) 8 (C) 16 <input checked="" type="radio"/> (D) 32</p> <p>Bei dieser Aufgabe soll die Entfernung zwischen zwei Städten angegeben werden, wobei der Maßstab einer Landkarte zu berücksichtigen ist. Für die Lösung dieser Aufgabe kann kein Routinewissen angewendet werden.</p>
Zutaten																			
Eier	4																		
Mehl	8 Tassen																		
Milch	$\frac{1}{2}$ Tasse																		
Zutaten																			
Eier	2																		
Mehl	4 Tassen																		
Milch	$\frac{1}{2}$ Tasse																		
Geometrie/Messen	<p>Einer dieser Winkel ist ein rechter Winkel. Welcher?</p> <p>(A)  (B)  (C)  (D) </p> <p>Bei dieser Aufgabe soll angegeben werden, bei welchem der gegebenen Winkel es sich um einen rechten Winkel handelt. Um diese Aufgabe zu lösen ist Routinewissen über Eigenschaften geometrischer Grundbegriffe (hier des rechten Winkels) abzurufen.</p>	 <p>Zeichne das Spiegelbild dieses Dreiecks. Die gestrichelte Linie m ist die Spiegelachse.</p> <p>Bei dieser Aufgabe soll das Spiegelbild eines Dreiecks gezeichnet werden. Dabei sind Eigenschaften geometrischer Begriffe (hier der Achsensymmetrie) zu beachten.</p>	<p>Ina findet diese Faltbögen, aus denen man Verpackungen machen kann. Aus welchem Faltbogen kann man die daneben gezeichnete Verpackung tatsächlich falten?</p> <p>(A)   (B)   (C)   <input checked="" type="radio"/> (D)  </p> <p>Bei dieser Aufgabe soll angegeben werden, welche Oberfläche zu den gegebenen Körpern passt. Für die Lösung dieser Aufgabe sind jeweils zwei Darstellungen zueinander in Beziehung zu setzen.</p>																
Umgang mit Daten	<p>Diese Abbildung zeigt die Pizzasorten, die an einem Pizzastand verkauft werden.</p>  <p>Von welcher Pizzasorte wurde am meisten verkauft?</p> <p>(A) <input checked="" type="radio"/> Schinken (B) Pilze (C) Salami (D) Paprika</p> <p>Bei dieser Aufgabe ist aus einem Kreisdiagramm abzulesen, welche Pizzasorte am meisten verkauft wurde. Um diese Aufgabe zu lösen ist Routinewissen über das Ablesen von Daten aus Diagrammen (hier: Kreisdiagramm) abzurufen.</p>	<p>Daniel fragt seine Freunde nach ihrer Lieblingsfarbe. Er sammelt die Antworten in der Tabelle unten.</p> <table border="1"> <tr><th>Lieblingsfarbe</th><th>Anzahl der Freunde</th></tr> <tr><td>Rot</td><td>4</td></tr> <tr><td>Grün</td><td>2</td></tr> <tr><td>Blau</td><td>6</td></tr> <tr><td>Gelb</td><td>7</td></tr> </table> <p>Dann beginnt Daniel, eine Abbildung zu zeichnen, um die Antworten darzustellen. Ergänze Daniels Abbildung.</p>  <p>Bei dieser Aufgabe sollen in einer Tabelle gegebene Daten in einem teilweise gegebenen Säulendiagramm ergänzt werden. Dazu sind Informationen aus einer Tabelle zu entnehmen und anschließend fehlende Säulen in ein Säulendiagramm einzuzeichnen.</p>	Lieblingsfarbe	Anzahl der Freunde	Rot	4	Grün	2	Blau	6	Gelb	7	<p>Herr Jansen hat die Schülerinnen und Schüler seiner Schule nach ihren Lieblingsfächern gefragt. Dieses Kreisbild zeigt, wie viele Schülerinnen und Schüler jeweils eines der 5 Fächer genannt haben.</p>  <p>Welche dieser 4 Abbildungen zeigt die gleiche Information wie das Kreisbild?</p> <p>(A)   (B)  </p> <p>Bei dieser Aufgabe sollen die Kinder entscheiden, welches von vier gegebenen Säulendiagrammen die Informationen eines gegebenen Kreisdiagramms enthält. Für die Lösung dieser Aufgabe sind die einzelnen Diagramme zueinander in Beziehung zu setzen.</p>						
Lieblingsfarbe	Anzahl der Freunde																		
Rot	4																		
Grün	2																		
Blau	6																		
Gelb	7																		

4 Der TIMSS-Test zur Erfassung der mathematischen Kompetenzen

In diesem Abschnitt wird zunächst der Testaufbau beschrieben (Abschnitt 4.1), bevor Analysen und Einschätzungen zur curricularen Validität folgen (Abschnitt 4.2). In Abschnitt 4.3 wird das in dieser Studie verwendete Stufenmodell zur Beschreibung der mathematischen Kompetenzen erläutert.

4.1 Testaufbau

Die TIMSS-Aufgaben wurden in einem kooperativen Prozess entwickelt, an dem neben dem *TIMSS & PIRLS International Study Center* am Boston College und den nationalen Studienleitungen auch Expertinnen und Experten aus den Bereichen der Mathematikdidaktik beteiligt waren (vgl. Kapitel 2). Die nationalen Studienleitungen und die beteiligten Fachdidaktikerinnen und Fachdidaktiker erstellten zunächst Vorschläge für Aufgaben und sogenannte Kodiermanuale, die Vorgaben für die Auswertung der Schülerantworten beinhalten. Diese wurden anschließend von Expertinnen und Experten aus verschiedenen Bereichen begutachtet und überarbeitet. Dabei wurde unter anderem sichergestellt, dass die Aufgaben adäquat übersetzt wurden. Im Rahmen großer nationaler Feldtests wurden die Aufgaben erprobt und auf der Grundlage der gewonnenen Erkenntnisse Testaufgaben für die Haupterhebung ausgewählt.

Der für die Testung genutzte Aufgabenpool von TIMSS 2011 umfasste für den Bereich Mathematik 177 Aufgaben. Dabei wurden zwei Aufgaben aufgrund ihrer nicht zufriedenstellenden statistischen Kennwerte aus der Skalenbildung eliminiert, so dass sich im Weiteren alle leistungsbezogenen Aussagen auf 175 Aufgaben stützen (Näheres dazu im Kapitel 2).

Um die Ergebnisse von TIMSS 2011 mit denen vorangegangener Studienzyklen vergleichen zu können, wurden rund 60 Prozent der Aufgaben aus den vorangegangenen Studienzyklen übernommen, wovon etwa zwei Drittel aus dem letzten und ein Drittel dieser Aufgaben aus dem vorletzten Studienzyklus stammen. Etwas mehr als die Hälfte der Testaufgaben ist im *Multiple-Choice*-Format gehalten und etwas weniger als die Hälfte der Aufgaben ist durch ein offeneres Antwortformat gekennzeichnet.

Tabelle 3.4 zeigt, wie viele Testaufgaben jeweils den mathematischen Inhaltsbereichen zugeordnet sind. Mit einem Anteil von 51 Prozent beziehen sich die meisten Testaufgaben auf den Inhaltsbereich *Arithmetik*. Mit 34 Prozent liegen etwas weniger Testaufgaben zum Bereich *Geometrie/Messen* vor. Die geringste Anzahl an Testaufgaben ist mit 15 Prozent dem Inhaltsbereich *Umgang mit Daten* gewidmet.

Tabelle 3.4: Verteilung der Testaufgaben auf die mathematischen Inhaltsbereiche

Mathematische Inhaltsbereiche	Testaufgaben TIMSS 2011	
	Anzahl	Prozent
Arithmetik	90	51
Geometrie/Messen	61	34
Umgang mit Daten	26	15
Gesamt	177	100

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Der Test ist so konstruiert, dass zu jedem mathematischen Inhaltsbereich Aufgaben aus allen drei kognitiven Anforderungsbereichen zu bearbeiten sind. Der Hauptanteil liegt bei Aufgaben, welche das *Reproduzieren* und das *Anwenden* erfordern.

Tabelle 3.5 zeigt, wie viele Testaufgaben sich auf die jeweiligen kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen* beziehen.

Tabelle 3.5: Verteilung der Testaufgaben auf die kognitiven Anforderungsbereiche

Kognitive Anforderungsbereiche	Testaufgaben TIMSS 2011	
	Anzahl	Prozent
Reproduzieren	71	40
Anwenden	71	40
Problemlösen	35	20
Gesamt	177	100

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

4.2 Curriculare Validität

Aus den vorangegangenen Ausführungen wird deutlich, dass sich die TIMSS-Rahmenkonzeption sowohl in ihrer Struktur als auch in weiten Teilen der Inhaltsbereiche mit den Bildungsstandards als vereinbar erweist. Lediglich für die *allgemeinen mathematischen Kompetenzen*, die eine eigene Dimension der Bildungsstandards bilden, lässt sich keine direkte Entsprechung im TIMSS-Rahmenkonzept erkennen. Sie sind jedoch mit einer mehr oder weniger starken Akzentuierung und mit Ausnahme der Kompetenz *Kommunizieren* in den drei kognitiven Anforderungsbereichen von TIMSS vertreten. Bereits im Rahmen der Berichtslegung zu TIMSS 2007 wurde argumentiert, dass insgesamt von einer verhältnismäßig guten Passung der Rahmenkonzeption von TIMSS und den Vorstellungen zur mathematischen Grundbildung in Deutschland auszugehen ist (Walther et al., 2008, S. 52 ff.).

Auf der Ebene der Testitems ergab sich 2007 allerdings insofern eine Diskrepanz, als nach Einschätzung von Expertinnen und Experten die abgebildeten Inhalte von gut einem Viertel der Testaufgaben für einen Großteil der getesteten Grundschul Kinder in Deutschland bis dahin noch nicht zu den curricular vorgesehenen Unterrichtsinhalten gezählt hatten. Diese Feststellung verweist auf eine der zentralen Herausforderungen international-vergleichender Schulleistungsmessungen, nämlich aussagekräftige Leistungstests zu entwickeln. TIMSS verfolgt für die Testkonstruktion einen Ansatz, nach dem zumindest für alle teilnehmenden Staaten und Regionen eine partielle Gültigkeit der Leistungstests für die Curricula angestrebt wird.

Eine Einschränkung des Tests auf Aufgaben, die ausnahmslos in *allen* Teilnehmerstaaten und Regionen curricular valide sind, hätte eine starke Einschränkung der Testinhalte und somit auch der bearbeitbaren Forschungsfragen zur Folge. TIMSS verzichtet daher auf eine Testreduktion, führt aber traditionsgemäß seit Beginn der TIMSS-Zyklen im Jahr 1995 eine sogenannte *Test-Curriculum Matching Analysis* (TCMA) durch. Alle Teilnehmer beurteilen hierzu jede einzelne Aufgabe hinsichtlich ihrer Validität für das nationale Curriculum.

Für Deutschland konnte als ein Ergebnis dieser Analyse für TIMSS 2007 jedoch auch gezeigt werden, dass sich die konkrete Aufgabenauswahl (curricular valide oder nicht) für das Abschneiden der Viertklässlerinnen und Viertklässler in Deutschland im internationalen Vergleich als weitgehend irrelevant erwies. Im Folgenden werden Verfahren und Ergebnisse der TCMA zu TIMSS 2011 für die Bundesrepublik Deutschland erläutert und damit die konkreten Testinhalte von TIMSS 2011 vertiefend zu den curricularen Grundlagen der mathematischen Grundbildung in Deutschland in Beziehung gesetzt.

Die Überprüfung der curricularen Validität der für diesen Studienzyklus relevanten Testaufgaben erfolgte für die Bundesrepublik Deutschland durch eine Expertengruppe von Fachdidaktikerinnen und Fachdidaktikern auf der Grundlage der Bildungsstandards im Fach Mathematik für den Primarbereich (KMK, 2005), des Lehrplans Mathematik für die Grundschule des bevölkerungsstärksten Bundeslandes Nordrhein-Westfalen (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008) und einer Auswahl gängiger Schulbücher. Im Rahmen dieser Prüfung wurde jede Testaufgabe hinsichtlich ihrer *grafischen Darstellungsformen, mathematischen Begriffe, Verfahren und Sachverhalte* sowie *Bezeichnungen* untersucht. Im Rahmen der Analyse wurden jeweils für jede Aufgabe all jene Merkmale als unbekannt und damit *curricular nicht valide* bewertet, welche in den Referenzquellen keine Erwähnung finden. Die vorliegende Analyse bezog sich damit ausschließlich auf die Passung zwischen den Testaufgaben und zentralen curricularen Vorgaben, die für einen Großteil der Viertklässlerinnen und Viertklässler in Deutschland gelten.

Der Tabelle 3.6 sind – nach Merkmalen differenziert – die Inhalte der für Deutschland als curricular nicht valide klassifizierten Testaufgaben zu entnehmen. Abbildung 3.2 zeigt darüber hinaus zwei Beispiele curricular nicht valider Testaufgaben.

Tabelle 3.6: Merkmale von curricular nicht validen Testaufgaben

Aufgabenmerkmale	weitestgehend unbekannte Inhalte
Grafische Darstellungsformen	Kreisdiagramm
Mathematische Begriffe, Verfahren und Sachverhalte	<p><i>Brüche</i>: Bruchteile, Bruchvergleich (kleiner, gleich, größer), Addition/Subtraktion von Brüchen</p> <p><i>Dezimalbrüche</i>: Vergleich von Dezimalbrüchen (kleiner, gleich, größer), Addition/Subtraktion von Dezimalbrüchen</p> <p><i>Gleichungen</i>: der Form $a:b=b$ mit b als Unbekannter, der Form $a+b=c+d$ mit einer Unbekannten</p> <p><i>Winkel</i>: Maßzahlen (90°, 180°), Vergleich von Winkelgrößen (kleiner, größer)</p>
Bezeichnungen	Drehung, drehsymmetrisch, Quadratmeter, Quadratzentimeter und Seite A


IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Betrachtet man die Aufgabeninhalte und -merkmale, lassen sich hinsichtlich der Passung zwischen der TIMSS-Rahmenkonzeption und den inhaltlichen Vorgaben zur mathematischen Grundbildung in Deutschland folgende Aspekte als erste Ergebnisse der TCMA festhalten:

- In Bezug auf die *Darstellungsformen* zeigte sich, dass Kreisdiagramme im Gegensatz zu Säulen- und Balkendiagramme nur selten vorkommen.
- In Bezug auf die *mathematischen Begriffe, Verfahren und Sachverhalte* zeigte sich, dass *Brüche* meistens nur als Maßzahlen von Größen (z.B. $\frac{1}{2}$ Liter) behandelt und nicht ohne den Kontextbezug zu *Größen* thematisiert werden. Auch für *Dezimalbrüche* ließ sich feststellen, dass sie in deutschen Grundschulen zumeist als ‚Kommazahlen‘ im Kontext von Größenangaben behandelt werden (z.B. 2,58 m sind gleich 2 m und 58 cm). Das Dezimaltrennzeichen wird dabei größtenteils als Trennmarke von Einheiten (im Beispiel Meter und Zentimeter) gedeutet. Mit Bezug zu Größen können Viertklässlerinnen und Viertklässler in Deutschland zwar schon mit Dezimalbrüchen (Addition/Subtraktion, Dezimalbruchvergleich) operieren, eine Thematisierung von Dezimalbrüchen ohne Bezug zu Größen ist allerdings in der Grundschule in der Regel noch nicht vorgesehen. Zudem zeigte sich, dass *Gleichungen der Form $a+b=c+d$ mit einer Unbekannten* beziehungsweise *der Form $a:b=b$ mit b als Unbekannter* nur vereinzelt in deutschen Schulbüchern zu finden sind und nicht als Gegenstand des Mathematikunterrichts an deutschen Grundschulen gelten können. Das Thema *Winkel* ist ebenfalls nur teilweise Gegenstand des Unterrichts: Während in der Grundschule in der Regel nur rechte Winkel betrachtet werden, wird die Beschreibung von Winkeln durch Maßzahlen in der Regel erst in der Sekundarstufe thematisiert.
- In Bezug auf die *Bezeichnungen* zeigte sich, dass Konzepte und Begriffe zu *Drehung* beziehungsweise *Drehsymmetrie* für den Mathematikunterricht an deutschen Grundschulen in der Regel thematisch nicht vorgesehen sind. Für das Thema *Flächeninhalt* wurde meistens von einer Klassifizierung der Testaufgaben als curricular valide abgesehen, da eine sehr uneinheitliche Bezeichnung von Flächeneinheiten in Schulbüchern (z.B. Meterquadrat, Quadratmeter, Kästchen oder Fliesen) festgestellt werden konnte und infolgedessen nicht davon auszugehen war, dass ein Großteil der Viertklässlerinnen und Viertklässler in Deutschland mit den in den Testaufgaben verwendeten Begriffen vertraut ist.

Abbildung 3.2: Beispiele curricular nicht valider Testaufgaben

Aufgabenbeispiel zur Bruchrechnung [Arithmetik: Brüche und Dezimalzahlen]	Aufgabenbeispiel zur Geometrie [Geometrie/Messen: Punkte, Geraden und Winkel]
Welcher dieser Brüche ist größer als $\frac{1}{2}$? <input checked="" type="radio"/> A $\frac{3}{5}$ <input type="radio"/> B $\frac{3}{6}$ <input type="radio"/> C $\frac{3}{8}$ <input type="radio"/> D $\frac{3}{10}$ (.46/.33)	Zeichne unten einen Winkel, der größer als 90 Grad, aber kleiner als 180 Grad ist.  (.31/.13)

Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an.

Als ein weiteres Ergebnis der TCMA lässt sich die Anzahl der Aufgaben, die nach Einschätzung der Expertengruppe nicht explizit Erwähnung in den curricularen Vorgaben finden, als Anteil an der Gesamtanzahl aller Mathematikaufgaben beschreiben. Insgesamt zeigt sich, dass mit 21 Prozent ein gutes Fünftel der Testaufgaben als curricular nicht valide eingestuft worden ist.

Tabelle 3.7 zeigt, wie die als curricular nicht valide eingeschätzten Testaufgaben über die mathematischen Inhaltsbereiche verteilt sind und welche Anteile sie an der Gesamtanzahl an Testaufgaben ausmachen. Es wird deutlich, dass 24 Prozent der Aufgaben aus dem Bereich *Arithmetik* als curricular nicht valide eingeschätzt wurden, während dieses für den Bereich *Geometrie/Messen* lediglich bei 18 Prozent und für den Bereich *Umgang mit Daten* bei nur 15 Prozent der Aufgaben der Fall war. Der relativ hohe Anteil an curricular nicht validen Aufgaben im Inhaltsbereich *Arithmetik* erklärt sich durch entsprechend hohe Aufgabenanteile im Themenfeld *Brüche und Dezimalzahlen*.

Tabelle 3.7: Anteile curricular nicht valider Testaufgaben nach mathematischen Inhaltsbereichen

Mathematische Inhaltsbereiche	Testaufgaben TIMSS 2011	curricular nicht valide Testaufgaben	
	Anzahl	Anzahl	Prozent
Arithmetik	90	22	24
Geometrie/Messen	61	11	18
Umgang mit Daten	26	4	15
Gesamt	177	37	21

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

Tabelle 3.8 zeigt, wie die als curricular nicht valide eingeschätzten Testaufgaben über die *kognitiven Anforderungsbereiche* verteilt sind und welche Anteile sie an der Gesamtanzahl an Testaufgaben ausmachen. Es wird deutlich, dass 30 Prozent der Aufgaben aus dem Bereich *Reproduzieren* als curricular nicht valide eingeschätzt wurden, während es für den Bereich *Anwenden* lediglich 18 Prozent und für den Bereich *Problemlösen* nur 9 Prozent der Aufgaben sind. Der relativ hohe Anteil an curricular nicht validen Aufgaben im Bereich *Reproduzieren* erklärt sich auch hier durch entsprechend hohe Aufgabenanteile im Themenfeld *Brüche und Dezimalzahlen*.

Tabelle 3.8: Anteile curricular nicht valider Testaufgaben nach kognitiven Anforderungsbereichen

Kognitive Anforderungsbereiche	Testaufgaben TIMSS 2011	curricular nicht valide Testaufgaben	
	Anzahl	Anzahl	Prozent
Reproduzieren	71	21	30
Anwenden	71	13	18
Problemlösen	35	3	9
Gesamt	177	37	21

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Werden die curricular validen TIMSS-Items nach den mathematischen Leitideen der Bildungsstandards im Fach Mathematik für den Primarbereich (KMK, 2005) klassifiziert, ergibt sich die aus der Tabelle 3.9 ersichtliche Verteilung.

Tabelle 3.9: Klassifikation curricular valider TIMSS-Mathematikaufgaben gemäß mathematischer Leitideen der Bildungsstandards

Mathematische Leitidee	Prozent
Zahlen und Operationen	26
Raum und Form	33
Muster und Strukturen	14
Größen und Messen	12
Daten, Häufigkeit und Wahrscheinlichkeit	15

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Die Anteile der Leitideen *Raum und Form* und *Daten, Häufigkeit und Wahrscheinlichkeit* stimmen dabei nahezu mit den beiden TIMSS-Inhaltsbereichen *Geometrie/Messen* und *Umgang mit Daten* überein. Die Anteile der Leitideen *Zahlen und Operationen* und *Muster und Strukturen* haben ihre Entsprechung im TIMSS-Inhaltsbereich *Arithmetik*. Auch die Leitidee *Größen und Messen* geht größtenteils im TIMSS-Inhaltsbereich *Arithmetik* auf, insofern die Aufgaben nicht dem Größenbereich *Längen* entsprechen.

Insgesamt lässt sich also festhalten, dass die curriculare Validität der in TIMSS 2011 für den Mathematiktest genutzten Aufgaben zufriedenstellend ist. Insgesamt wurden 21 Prozent der verwendeten Testaufgaben als curricular nicht valide eingeschätzt. Für TIMSS 2007 lag der Anteil mit 26 Prozent etwas darüber (vgl. Walther et al., 2008). Der Anspruch, mit der TIMSS-Rahmenkonzeption Kompetenzen zu erfassen, die in der Schule aufgrund von curricularen Vorgaben vermittelt werden, wurde damit auch in TIMSS 2011 für Deutschland nur mit gewisser Einschränkung erfüllt. Aus diesem Grund wurde gezielt geprüft, ob diese Tatsache einen Einfluss auf die Ergebnisse der Schülerinnen und Schüler im Bereich Mathematik in Deutschland hat. Darauf wird im Ergebnisteil (Abschnitt 5.1) noch näher eingegangen.

4.3 Kompetenzstufen in Mathematik

In TIMSS erfolgt die Darstellung der Leistungsstände von Schülerinnen und Schülern durch die Angabe bestimmter statistischer Kennwerte, die zwar Auskunft über relative Unterschiede in Leistungsständen geben, jedoch zunächst nichts Inhaltliches über die mathematischen Kompetenzen von Viertklässlerinnen und Viertklässlern aussagen. Um zu verdeutlichen, was die getesteten Schülerinnen und Schüler am Ende der vierten Jahrgangsstufe können und über welche spezifischen mathematischen Kompetenzen sie verfügen, wird die TIMSS-2011-Leistungsskala durch Setzung von vier *Benchmarks* (vgl. Martin & Mullis, 2011) in fünf unterschiedliche Abschnitte eingeteilt.

Für diese Abschnitte ist es möglich, die unterschiedlichen mathematischen und kognitiven Anforderungen, die bewältigt werden müssen, um die einem entsprechenden Abschnitt zugeordneten Aufgaben mit hinreichender Sicherheit zu lösen, als aufeinander aufbauende *Kompetenzstufen* zu beschreiben. Jede der fünf Kompetenzstufen gibt dann an, über welches mathematische Wissen und über welche mathematischen Fähigkeiten und Fertigkeiten diejenigen Schülerinnen und Schüler im Allgemeinen verfügen, die aufgrund ihres Testergebnisses diesem bestimmten Abschnitt auf der Leistungsskala und damit dieser bestimmten Kompetenzstufe zugeordnet werden.

Dabei wird davon ausgegangen, dass die Kompetenzstufen hierarchisch geordnet sind. Das heißt, dass Schülerinnen und Schüler, deren gezeigte mathematische Leistungen auf dem Niveau der Kompetenzstufe III liegen, Testaufgaben, die typisch für diese Stufe sind, mit einer geringeren Wahrscheinlichkeit lösen als Testaufgaben der Stufen II oder I, aber mit einer noch geringeren Wahrscheinlichkeit Testaufgaben der Stufen IV und V (vgl. Kapitel 2).

Eine ausführliche inhaltliche Beschreibung der Kompetenzstufen findet sich im Anhang B. Tabelle 3.10 ist die Zuordnung der Leistungsniveaus zu Abschnitten auf der Leistungsskala sowie eine komprimierte Skizze der fünf Kompetenzstufen zu entnehmen. Dabei ist festzuhalten, dass sich die Beschreibungen der Kompetenzstufen von denen der fünf TIMSS-2007-Kom-

petenzstufen (vgl. Walther et al., 2008, S. 68) in wenigen Details in Bezug auf die Formulierungen unterscheiden. Das heißt, dass die inhaltliche Umschreibung der Stufen für jeden Untersuchungszyklus auf der Grundlage der von den Schülerinnen und Schülern erzielten Resultate und der eingesetzten Aufgaben neu justiert wird. Details zu diesem Konstruktionsprozess werden im Kapitel 2 beschrieben.

Tabelle 3.10: Beschreibung der fünf Kompetenzstufen für die Gesamtskala Mathematik

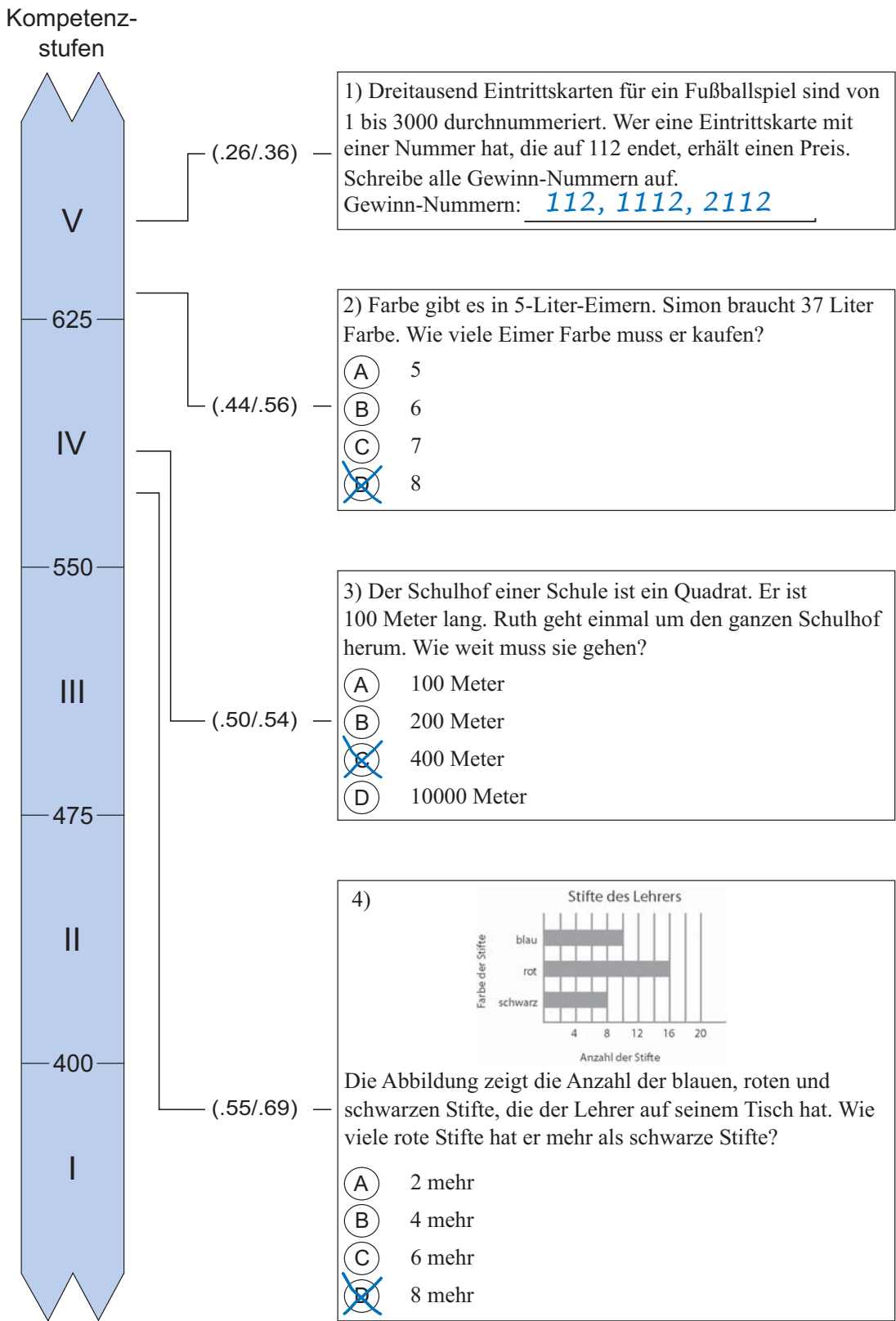
Kompetenzstufe V (ab 625): fortgeschritten
<i>Die Schülerinnen und Schüler können ihre mathematischen Fertigkeiten und Fähigkeiten verständlich beim Lösen verhältnismäßig komplexer Probleme anwenden und ihr Vorgehen erläutern. Sie können mehrschrittige Sachaufgaben in vielfältigen Kontexten mit natürlichen Zahlen lösen, auch durch Nutzen proportionaler Beziehungen. Sie lassen ein fortschreitendes Verständnis für Brüche und Dezimalbrüche erkennen. Die Schülerinnen und Schüler können ihr geometrisches Wissen und Können im Bereich zwei- und dreidimensionaler Figuren in vielfältigen Situationen anwenden. Sie können Schlussfolgerungen aus Daten in Tabellen ziehen und diese begründen.</i>
Kompetenzstufe IV (550–624): hoch
<i>Die Schülerinnen und Schüler können ihre mathematischen Fertigkeiten und Fähigkeiten für das Lösen von Problemen anwenden. Sie können mit Hilfe der Grundrechenarten Textaufgaben lösen und dabei auch die Division sowie ihr Verständnis des Stellenwertprinzips in verschiedenen Situationen nutzen. Sie können Zahlenmuster fortsetzen und dabei auch weiter entfernte Werte bestimmen. Die Schülerinnen und Schüler zeigen Verständnis der Achsensymmetrie und geometrischer Eigenschaften. Sie können Daten aus Diagrammen und Tabellen entnehmen und interpretieren, um Probleme zu lösen. Sie können Informationen aus Piktogrammen und Strichlisten nutzen, um Säulendiagramme zu vervollständigen.</i>
Kompetenzstufe III (475–549): durchschnittlich
<i>Die Schülerinnen und Schüler können elementares mathematisches Wissen sowie elementare mathematische Fertigkeiten und Fähigkeiten in einfachen Situationen anwenden. Die Schülerinnen und Schüler auf diesem Niveau zeigen Grundwissen von natürlichen Zahlen und anfängliches Wissen von Brüchen. Sie können sich dreidimensionale Figuren ausgehend von zweidimensionalen Darstellungen vorstellen. Sie entnehmen Säulendiagrammen, Piktogrammen und Tabellen Informationen, um einfache Probleme zu lösen.</i>
Kompetenzstufe II (400–474): niedrig
<i>Die Schülerinnen und Schüler verfügen über elementares mathematisches Wissen sowie über elementare mathematische Fertigkeiten und Fähigkeiten. Sie können einfache Additions- und Subtraktionsaufgaben lösen. Sie verfügen über anfängliches Wissen zu zueinander parallelen und rechtwinkligen Geraden, geometrischen Grundformen und Koordinatensystemen. Sie können einfachen Balkendiagrammen und Tabellen Informationen entnehmen und diese vervollständigen.</i>
Kompetenzstufe I (unter 400): rudimentär
<i>Die Schülerinnen und Schüler verfügen über rudimentäres schulisches Anfangswissen. Auf diesem Niveau können viele Kinder einfache Routineaufgaben mit Grundrechenarten ausführen, jedoch selbst einfache Aufgaben der Kompetenzstufe II nur gelegentlich oder ansatzweise lösen.</i>

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

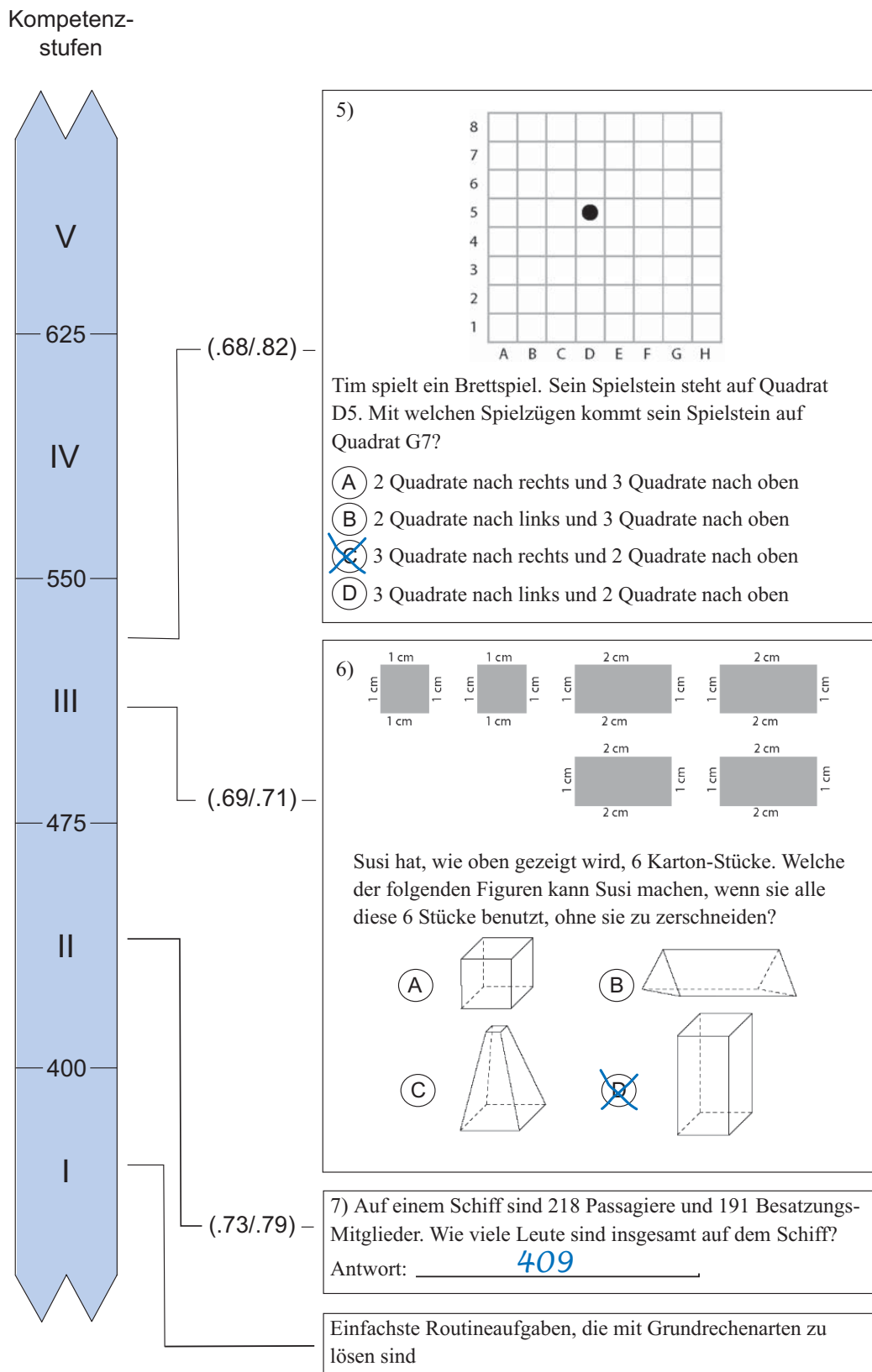
Eine Illustration der Kompetenzstufen anhand einiger Beispielaufgaben findet sich in den Abbildungen 3.3 und 3.4.

Abbildung 3.3: Kompetenzstufen und Beispielaufgaben (Gesamtskala Mathematik) I



Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an.

Abbildung 3.4: Kompetenzstufen und Beispielaufgaben (Gesamtskala Mathematik) II



Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an.

Die in den beiden Abbildungen angegebenen Werte in Klammern repräsentieren die internationalen und nationalen Lösungshäufigkeiten. Die erste Aufgabe beispielsweise wurde von 26 Prozent der Schülerinnen und Schüler im internationalen Durchschnitt und von 36 Prozent der Schülerinnen und Schüler in Deutschland gelöst. Die Zuordnung der Aufgaben zu den Kompetenzstufen erfolgt auf Grundlage internationaler Leistungsdaten. Somit kann es auch sein, dass in einzelnen Teilnehmerstaaten und Regionen eine Aufgabe, die für eine bestimmte Kompetenzstufe charakteristisch ist, eine bedeutsam höhere Lösungshäufigkeit aufweist als eine Aufgabe, die auf Grundlage der internationalen Lösungshäufigkeiten eigentlich einer niedrigeren Kompetenzstufe zugeordnet wurde. Dies zeigt sich beispielsweise für Deutschland im Vergleich der Aufgaben 5 auf Kompetenzstufe III (82%) und 7 auf Kompetenzstufe II (79%). Im Weiteren wird für die einzelnen Aufgaben ausgeführt, über welche mathematischen Kompetenzen Schülerinnen und Schüler jeweils verfügen müssen, um die Aufgaben erfolgreich zu lösen.

Kompetenzstufe V. Bei Aufgabe 1 sollen die Schülerinnen und Schüler alle Zahlen kleiner 3000 angeben, welche auf ‚112‘ enden. Um die Aufgabe korrekt zu lösen, müssen die Kinder das Stellenwertsystem verstanden haben. Sie müssen erkennen, dass die Ziffern in ‚112‘ die Hunderter, Zehner und Einer in den gesuchten Zahlen repräsentieren.

In Aufgabe 2 sollen die Kinder eine Textaufgabe zur Division mit Rest lösen. Um die korrekte Antwort zu erhalten, reicht es nicht, die Division durchzuführen, denn anschließend muss das Ergebnis auf die Ausgangssituation zurückbezogen und die nächstgrößere natürliche Zahl angegeben werden.

Kompetenzstufe IV. Bei Aufgabe 3 sollen die Schülerinnen und Schüler den Umfang eines Quadrats bestimmen, welches 100 m lang ist. Dazu müssen die Schülerinnen und Schüler verstanden haben, dass alle vier Seiten 100 m lang sind, und diese Werte addieren.

Bei Aufgabe 4 sollen die Kinder den Unterschied zwischen zwei Anzahlen angeben, welche mittels eines Balkendiagramms dargestellt werden. Dazu müssen sie dem Balkendiagramm die erforderlichen Informationen entnehmen und anschließend die Differenz bestimmen.

Kompetenzstufe III. Bei Aufgabe 5 sollen die Schülerinnen und Schüler einen Weg zwischen zwei Punkten in einem Koordinatensystem angeben. Um den Zielpunkt zu finden und die verschiedenen angegebenen Wege hinsichtlich der Passung zwischen Start- und Zielpunkt zu prüfen, müssen sie sich im Koordinatensystem orientieren können.

Bei Aufgabe 6 sollen die Schülerinnen und Schüler aus zwei quadratischen und vier rechteckigen Kartonstücken in der Vorstellung entscheiden, welche von vier vorgegebenen Figuren sie aus diesen sechs Stücken zusammensetzen können. Hierzu müssen sie Eigenschaften von einfachen Körpern kennen und über Raumvorstellungsvermögen verfügen.

Kompetenzstufe II. Bei Aufgabe 7 handelt es sich um eine Textaufgabe zur Addition mit Übertrag im Zahlenraum bis 1000. Zur Lösung dieser Aufgaben müssen die Kinder lediglich elementare mathematische Kompetenzen zeigen.

Kompetenzstufe I. Die Schülerinnen und Schüler, die auf dieser Kompetenzstufe zu verorten sind, lassen im TIMSS-Test nur basale Kompetenzen erkennen. Es ist davon auszugehen, dass Schülerinnen und Schüler, die dieser Kompetenzstufe zugeordnet sind, jedoch einfache Routineaufgaben mit Grundrechenarten ausführen können.

5 Ergebnisse

In diesem Abschnitt werden die Ergebnisse von TIMSS 2011 entlang der in Abschnitt 1 angeführten Forschungsfragen berichtet und mit den Resultaten von TIMSS 2007 verglichen. Abschnitt 5.1 befasst sich mit den Ergebnissen, die die Schülerinnen und Schüler auf der Gesamtskala Mathematik im internationalen Vergleich erzielt haben. In Abschnitt 5.2 wird aufgezeigt, wie sich die Schülerinnen und Schüler in den Teilnehmerstaaten und Regionen auf die fünf Kompetenzstufen verteilen und wie groß jeweils die Gruppen der auffällig leistungsschwachen und leistungsstarken Kinder sind.

Anschließend werden in den Abschnitten 5.3 und 5.4 die Ergebnisse zu Teilkompetenzen – aufgeschlüsselt nach Inhaltsbereichen und nach kognitiven Anforderungsbereichen – dargestellt. Neben den Ergebnissen aus 2011 wird auch in diesen Abschnitten beschrieben, inwieweit sich die Ergebnisse von TIMSS 2007 und TIMSS 2011 unterscheiden. Abschnitt 5.5 ist nationalen Ergebnissen zum Selbstkonzept und zu Einstellungen der Schülerinnen und Schüler im Bereich Mathematik gewidmet. In Bezug auf Unterschiede in der mathematischen Kompetenz in Abhängigkeit von Geschlecht, sozialer Herkunft oder Migrationshintergrund sei auf die Kapitel 6, 7 und 8 in diesem Band verwiesen.

An TIMSS 2007 nahmen 36 Staaten und Regionen teil plus 7 sogenannte Benchmark-Teilnehmer, also einzelne Regionen wie Florida (USA) oder Alberta (Kanada). Im Jahr 2011 waren 50 Staaten und Regionen beteiligt, darunter 21 EU- beziehungsweise 26 OECD-Staaten; außerdem drei Staaten, die mit der Jahrgangsstufe 6 teilnahmen, sowie sieben Benchmark-Teilnehmer (vgl. Kapitel 2).

Für alle weiteren Darstellungen wird eine Auswahl an Staaten anhand folgender Kriterien getroffen: (1) Teilnehmer, die Mitglieder der EU sind, (2) Teilnehmer, die der OECD angehören, (3) Teilnehmer, die auf der Gesamtskala Mathematik signifikant bessere Ergebnisse erzielt haben als Deutschland. Eine vollständige Darstellung der Ergebnisse für alle Teilnehmerstaaten und Regionen ist dem Ergebnisbericht der internationalen Studienleitung zu entnehmen (vgl. Mullis, Martin, Foy & Arora, 2012).

Damit die Ergebnisse zur mathematischen Kompetenz der Schülerinnen und Schüler in Deutschland eingeordnet werden können, werden in diesem Bericht neben dem Vergleich mit einzelnen anderen Staaten drei weitere Vergleichsmöglichkeiten angeboten, für die die Werte unterschiedlicher Staaten zu sogenannten Vergleichsgruppen (VG) zusammengefasst werden: Die Ergebnisse in Deutschland können mit dem Mittelwert aller Staaten (ohne Benchmark-Teilnehmer und Teilnehmer mit Jahrgangsstufe 6), mit dem Mittelwert der beteiligten OECD-Staaten (VG_{OECD}) und mit dem Mittelwert der beteiligten EU-Staaten (VG_{EU}) verglichen werden. Im Unterschied zu den Abbildungen, in denen die Werte aus 2011 dargestellt sind, wird bei allen Abbildungen, in denen Unterschiede zwischen TIMSS 2007 und TIMSS 2011 berichtet werden, auf die Angabe des internationalen Mittelwerts sowie des Mittelwerts der VG_{EU} und der VG_{OECD} verzichtet, da es in den beiden Untersuchungen unterschiedliche

Teilnehmer und damit auch unterschiedliche Zusammensetzungen dieser Gruppen gab und somit Veränderungen schwer zu interpretieren sind.

5.1 Kompetenzen im internationalen Vergleich

In diesem Abschnitt wird die Mathematikleistung von Schülerinnen und Schülern im Vergleich zu den Leistungsergebnissen in den übrigen Teilnehmerstaaten berichtet. Dabei geht es zunächst um die Mittelwerte, die Streuung der Leistungsergebnisse und die Positionierung des Ergebnisses für Deutschland im internationalen Vergleich.

Die Einordnung der von Grundschulkindern in Deutschland erzielten Leistungsergebnisse in die nach Leistungsstärke gestufte internationale Rangreihe sowie der internationale Vergleich von Leistungsergebnissen erfolgen zunächst formal. Sie stützen sich ausschließlich auf die Leistungswerte auf der internationalen TIMSS-2011-Leistungsskala, ohne inhaltliche, auf mathematische Inhalts- oder Anforderungsbereiche bezogene Interpretationen zu geben. Dies erfolgt in den nachfolgenden Abschnitten.

Leistungsmittelwerte der Staaten: In Abbildung 3.5 sind für alle Teilnehmerstaaten und Regionen die jeweiligen Leistungsmittelwerte der Schülerinnen und Schüler eines Teilnehmerstaates (M), ihre Standardabweichungen (SD) sowie die Standardfehler (SE) dieser beiden Werte angegeben. Des Weiteren ist der Mittelwert für alle Teilnehmerstaaten (Internationaler Mittelwert) sowie die Mittelwerte für die VG_{OECD} und VG_{EU} aufgeführt. Teilnehmer mit einem besonderen Teilnahmestatus (Benchmark-Teilnehmer und Teilnahme mit der 6. Jahrgangsstufe) sind gesondert dargestellt. Die Leistungsmittelwerte der Teilnehmer sind innerhalb dieser Gruppen in einer Rangreihe angeordnet.

Das Format der Bezeichnungen der Teilnehmer sowie die beige gestellten Fußnoten verweisen auf Besonderheiten bezüglich der nationalen Untersuchungspopulationen und -stichproben. Die Bedeutung der einzelnen Fußnoten wird ausführlich in Kapitel 2 erläutert. Um eine internationale Vergleichbarkeit der Ergebnisse zu gewährleisten, verlangt die internationale Studienleitung die Einhaltung strenger methodischer Standards. Teilnehmer, die aus unterschiedlichen Gründen diese Vorgaben nicht eingehalten haben, sind durch eine Kursivsetzung ihres Namens gekennzeichnet. Dies trifft auf 26 der 59 Teilnehmerstaaten und Regionen zu. Während für die anderen 33 Teilnehmer – zu denen auch Deutschland zählt – eine internationale Vergleichbarkeit der Ergebnisse gegeben ist, gilt für die kursiv gesetzten Teilnehmer, dass auf eine Repräsentativität der Stichprobe für alle Viertklässlerinnen und Viertklässler des jeweiligen Teilnehmerstaates beziehungsweise der Region nicht oder nur mit Einschränkung geschlossen werden kann. Für die vergleichende Interpretation der Ergebnisse Deutschlands mit kursiv gesetzten Teilnehmerstaaten und Regionen ist unter Berücksichtigung der Tabellen im Anhang A ‚Weiterführende Informationen zu den Teilnehmerstaaten‘ für den Einzelfall zu prüfen, inwiefern ein differentes Abschneiden durch Unterschiede in den Teilnahmebedingungen bedingt sein könnte.

Die Schülerinnen und Schüler in Deutschland erreichen in Mathematik einen Leistungsmittelwert von 528 Punkten und befinden sich damit noch im oberen Drittel der Rangreihe. Damit liegt Deutschland signifikant sowohl über dem Skalenmittelwert von 500, der in TIMSS 1995 festgelegt wurde, als auch über

dem internationalen Mittelwert von TIMSS 2011, der 491 beträgt. Insgesamt reicht das Spektrum der Leistungsmittelwerte auf der Gesamtskala Mathematik von über 600 Punkten bis knapp unter 250 Punkten. Am oberen Ende der Leistungsskala lassen sich Singapur (606), Republik Korea (Südkorea) (605) und Hongkong (602) verorten. Am unteren Ende der Leistungsskala lassen sich Teilnehmer wie Marokko (335) oder Jemen (248) verorten.

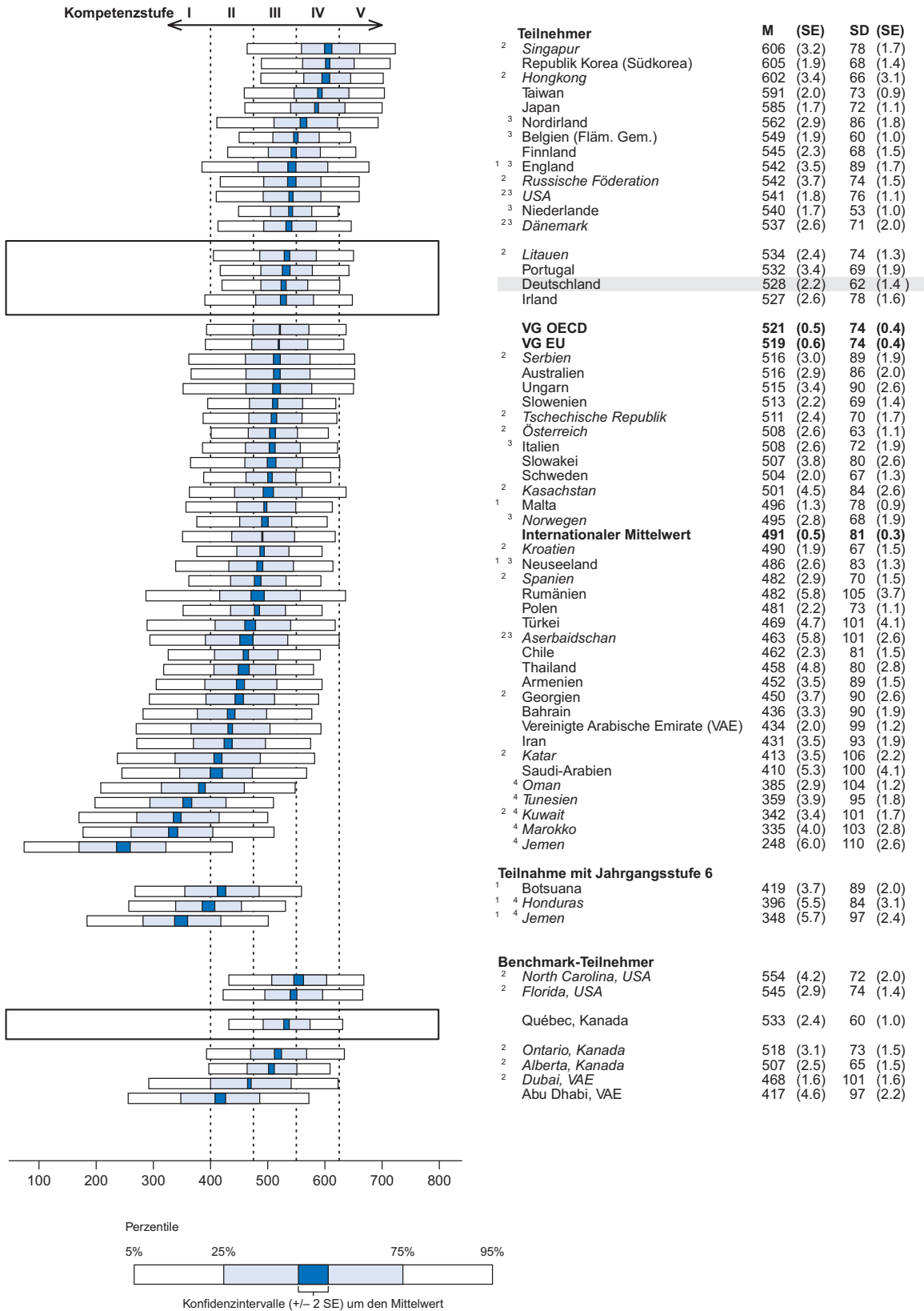
Signifikante Leistungsunterschiede: Bei der Interpretation der Ergebnisse ist zu beachten, dass die Anordnung der Staaten nach Rangplätzen nicht notwendigerweise Schlüsse auf Unterschiede in der mathematischen Kompetenz zwischen den Staaten zulässt. Die deskriptiven, numerischen Unterschiede können zufällige Unterschiede darstellen, die statistisch nicht bedeutsam sind. Deshalb sind in der Abbildung 3.5 die Perzentilbänder jener Teilnehmer mit einer schwarzen Umrandung gekennzeichnet, deren Leistungsmittelwerte sich trotz deskriptiver Unterschiede nicht signifikant vom Leistungsmittelwert für Deutschland (528) unterscheiden. Dieser Staatengruppe gehören Litauen (534), Portugal (532) und Irland (527) an. Signifikant bessere Leistungsmittelwerte als die Viertklässlerinnen und Viertklässler in Deutschland erzielen hingegen die Schülerinnen und Schüler aus Singapur (606), Republik Korea (Südkorea) (605), Hongkong (602), Taiwan (591), Japan (585), Nordirland (562), Belgien (549), Finnland (545), England (542), der Russischen Föderation (542), den USA (541), den Niederlanden (540) und Dänemark (537). Signifikant schlechtere Leistungsmittelwerte als die Viertklässlerinnen und Viertklässler in Deutschland erzielen die Schülerinnen und Schüler aus weiteren 33 Teilnehmerstaaten, zu denen unter anderem Ungarn (515), Slowenien (513), die Tschechische Republik (511), Italien (508), Österreich (508), die Slowakei (507), Schweden (504), Norwegen (495) und Polen (481) zählen. Der Leistungsmittelwert der Viertklässlerinnen und Viertklässler aus Deutschland liegt zudem signifikant über dem Mittelwert aller teilnehmenden EU-Staaten (VG_{EU} , 519) sowie aller teilnehmenden OECD-Staaten (VG_{OECD} , 521).

Leistungsstreuung: Die im vorigen Abschnitt dargestellten Ergebnisse zeigen, inwieweit sich die Leistungsmittelwerte der einzelnen Teilnehmerstaaten voneinander unterscheiden. Zu analysieren ist allerdings auch, wie unterschiedlich die mathematischen Kompetenzen von Schülerinnen und Schülern innerhalb der einzelnen Staaten verteilt sind. Auf der linken Seite der Abbildung 3.5 ist die Leistungsstreuung durch Perzentilbänder illustriert, die jeweils das 5., 25., 75. und 95. Perzentil angeben (vgl. Kapitel 2). Der Leistungswert etwa, der dem 75. Perzentil entspricht, sagt aus, dass 75 Prozent der Schülerinnen und Schüler einen niedrigeren oder gleich großen und damit 25 Prozent der Schülerinnen und Schüler einen größeren Leistungswert erzielt haben.

Die genauen Leistungswerte, die die Schülerinnen und Schüler der einzelnen Teilnehmer auf den jeweiligen Perzentilen erzielen, sind für die Teilnehmer, die an TIMSS 2007 und TIMSS 2011 teilgenommen haben, in Tabelle 3.12 dargestellt.

Betrachtet man die in Abbildung 3.5 dargestellten Perzentilbänder, so zeigt sich für Deutschland eine vergleichsweise geringe Leistungsstreuung. Der Leistungsunterschied zwischen den jeweils 5 Prozent leistungsschwächsten und leistungsstärksten Kindern in Deutschland ist im internationalen Vergleich kleiner und damit die Verteilung der Mathematikleistung auf die „mittleren“ 90 Prozent der Kinder homogener als in Staaten mit breiteren Perzentil-

Abbildung 3.5: Testleistung der Schülerinnen und Schüler im internationalen Vergleich – Gesamtskala Mathematik



□ Nicht statistisch signifikant vom deutschen Mittelwert abweichende Staaten ($p > .05$).
 Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 4 = Sehr hoher Anteil an Schülerinnen und Schülern mit nicht skalierbaren Leistungswerten.

bändern. In Deutschland beträgt der Leistungsunterschied zwischen den beiden ‚Extremgruppen‘ 206 Punkte; in der Republik Korea (Südkorea), einem der signifikant besseren Teilnehmer, sind es 225 Punkte. Auch im europäischen Vergleich erweist sich die Leistungsstreuung Deutschlands als verhältnismäßig gering, lediglich in den Niederlanden mit einem Leistungsunterschied von 174 Punkten sowie in Belgien (195 Punkte) und Österreich (205 Punkte) fällt die Streuung geringer und damit noch homogener aus. Die höchsten Unterschiede lassen sich in Rumänien (349), Ungarn (298), England (292) und Nordirland (282) finden.

Die Leistungsstreuung wird auch in der Standardabweichung (*SD*) ausgewiesen (vgl. Kapitel 2). Wie in Abbildung 3.5 ersichtlich, ergibt sich für Deutschland ein Wert von 62 Punkten. Dieser Wert ist dahingehend zu interpretieren, dass gut zwei Drittel der Viertklässlerinnen und Viertklässler in Deutschland Leistungswerte im Skalenbereich von 466 (528–62) bis 590 (528+62) Punkten erzielen. Im europäischen wie auch internationalen Vergleich ist Deutschland damit nach den Niederlanden (53) und der Flämischen Gemeinschaft in Belgien (60) der Teilnehmer mit der geringsten Streuung. Im europäischen Vergleich weisen Rumänien (105), Ungarn (90) und England (89) die höchsten Standardabweichungen auf. Im Vergleich mit England bedeutet dies beispielweise, dass England mit 542 Punkten zwar einen signifikant höheren Mittelwert als Deutschland (528) erzielt, die Leistungswerte der Schülerinnen und Schüler in England aber wesentlich heterogener ausfallen.

Bedeutung der curricular nicht validen Aufgaben: In Abschnitt 4.2 wurde die *Test-Curriculum-Matching Analyse* (TCMA) erörtert, mit der die in TIMSS 2011 eingesetzten Aufgaben auf ihre Repräsentativität für die Curricula der einzelnen Teilnehmer eingeschätzt wurden. Für Deutschland zeigte sich, dass 21 Prozent der eingesetzten Aufgaben als curricular nicht valide eingeschätzt wurden, was bedeutet, dass die in den Aufgaben thematisierten Inhalte weitestgehend nicht in den von Grundschulen in Deutschland geltenden Mathematik-Lehrplänen vorgesehen sind. Um zu prüfen, inwieweit sich dies auf die Ergebnisse im internationalen Vergleich auswirkt, wurde die Gesamtskala Mathematik in einer modifizierten Form erneut berechnet, wobei der Berechnung der Mittelwerte aller Staaten lediglich jene 138 Testaufgaben zugrunde gelegt wurden, die für Deutschland als curricular valide zu bewerten sind.

Im linken Teil der Tabelle 3.11 sind die auch in Abbildung 3.5 dargestellten Mittelwerte aller Teilnehmer unter Berücksichtigung von 175 Testaufgaben angegeben. Zwei der 177 eingesetzten Testaufgaben wurden wegen ungenügender Messeigenschaften nicht für die Ermittlung der Leistungsmittelwerte berücksichtigt. Im mittleren Teil der Tabelle sind Leistungsmittelwerte der Teilnehmer aufgeführt, die sich als Ergebnis der Reskalierung der internationalen Datensätze unter Annahme eines einparametrischen Skalierungsmodells ohne Hintergrundmodell (vgl. Kapitel 2) ergeben haben. Betrachtet man die Rangplatzverschiebung von Modell I zu Modell II zeigt sich, dass die Unterschiede eher marginal sind. Im rechten Teil von Tabelle 3.11 sind nun diejenigen Ergebnisse aufgeführt, die ermittelt wurden, nachdem die für die deutschen Schülerinnen und Schüler curricular nicht validen Aufgaben aus den Berechnungen eliminiert wurden. Betrachtet man nun die Rangplatzverschiebung von Modell II zu Modell III ergibt sich für Deutschland eine nominelle Verbesserung um vier Rangplätze. Die Veränderung des Leistungsmittelwerts ist jedoch nicht signifikant. Die Schülerinnen und Schüler in den USA, den Niederlanden und Dänemark, die im Modell I signifikant besser abschneiden, sowie die Schülerinnen und Schüler aus Litauen, die

Tabelle 3.11: Testwerte und Rangplätze Mathematik – Vergleich der Skalierungsmodelle mit und ohne für Deutschland curricular validen Testaufgaben

Teilnehmer	Modell I (int. Skalierungsmodell [n = 175 Testaufgaben])					Modell II (Replikation Modell I; Modifikation: 1PL; ohne Hintergrundmodell)					Δ Rang _(INT-1PL)	Modell III (Replikation Modell II; Modifikation: nur mit für Deutschland curricular validen Testaufgaben [n = 138])					Δ Rang _(1PL-CV)
	M (SE)	SD (SE)	Rang _{INT}	M (SE)	SD (SE)	Rang _{1PL}	M (SE)	SD (SE)	Rang _{CV}								
² Singapur	606 (3.2)	78 (1.7)	1	613 (2.9)	81 (1.7)	1	0	601 (2.8)	81 (1.6)	3	-2						
Republik Korea (Südkorea)	605 (1.9)	68 (1.4)	2	612 (1.5)	75 (1.1)	2	0	607 (1.5)	74 (1.0)	1	1						
² Hongkong	602 (3.4)	66 (3.1)	3	612 (3.0)	74 (2.6)	3	0	602 (2.8)	73 (2.3)	2	1						
Taiwan	591 (2.0)	73 (0.9)	4	598 (1.7)	78 (1.6)	4	0	593 (1.6)	77 (1.0)	4	0						
Japan	585 (1.7)	72 (1.1)	5	595 (1.7)	78 (1.0)	5	0	591 (1.7)	77 (1.1)	5	0						
³ Nordirland	562 (2.9)	86 (1.8)	6	577 (2.5)	88 (1.7)	6	0	568 (2.5)	86 (1.4)	6	0						
³ Belgien (Fläm. Gem.)	549 (1.9)	60 (1.0)	7	564 (2.1)	70 (1.2)	7	0	559 (1.8)	70 (0.8)	8	-1						
Finnland	545 (2.3)	68 (1.5)	8	558 (2.1)	76 (1.2)	11	-3	556 (2.4)	77 (1.5)	11	0						
¹ ³ England	542 (3.5)	89 (1.7)	9	560 (3.2)	90 (1.5)	9	0	557 (3.0)	89 (1.8)	10	-1						
² Russische Föderation	542 (3.7)	74 (1.5)	10	561 (3.3)	80 (1.5)	8	2	563 (3.0)	80 (1.2)	7	1						
² ³ USA	541 (1.8)	76 (1.1)	11	559 (1.8)	81 (1.0)	10	1	553 (1.7)	79 (1.2)	13	-3						
³ Niederlande	540 (1.7)	53 (1.0)	12	552 (1.7)	67 (1.4)	15	-3	552 (1.7)	67 (1.3)	15	0						
² ³ Dänemark	537 (2.6)	71 (2.0)	13	553 (2.3)	78 (1.4)	12	1	551 (2.4)	78 (1.4)	16	-4						
² Litauen	534 (2.4)	74 (1.3)	14	552 (2.4)	79 (1.7)	14	0	552 (2.3)	79 (1.5)	14	0						
Portugal	532 (3.4)	69 (1.9)	15	552 (3.0)	77 (1.8)	13	2	558 (2.7)	79 (1.6)	9	4						
Deutschland	528 (2.2)	62 (1.4)	16	547 (2.1)	75 (1.3)	16	0	555 (2.3)	78 (1.3)	12	4						
Irland	527 (2.6)	78 (1.6)	17	545 (2.4)	81 (1.6)	17	0	541 (2.6)	82 (1.9)	18	-1						
² Serbien	516 (3.0)	89 (1.9)	18	537 (2.7)	85 (1.8)	20	-2	539 (2.6)	86 (1.9)	21	-1						
Australien	516 (2.9)	86 (2.0)	19	538 (2.4)	86 (1.4)	19	0	536 (2.3)	84 (1.8)	23	-4						
Ungarn	515 (3.4)	90 (2.6)	20	540 (2.9)	88 (1.3)	18	2	542 (2.7)	89 (1.5)	17	1						
Slowenien	513 (2.2)	69 (1.4)	21	535 (2.1)	76 (1.4)	21	0	540 (2.1)	80 (1.4)	19	2						
² Tschechische Republik	511 (2.4)	70 (1.7)	22	531 (2.4)	77 (1.7)	22	0	539 (2.5)	80 (1.4)	20	2						
² Österreich	508 (2.6)	63 (1.1)	23	529 (2.6)	72 (1.3)	25	-2	538 (2.5)	75 (1.3)	22	3						
³ Italien	508 (2.6)	72 (1.9)	24	530 (2.5)	77 (1.6)	23	1	531 (2.4)	78 (1.5)	25	-2						
Slowakei	507 (3.8)	80 (2.6)	25	529 (3.1)	81 (1.6)	24	1	532 (3.0)	83 (1.7)	24	0						
Schweden	504 (2.0)	67 (1.3)	26	522 (2.1)	74 (1.3)	27	-1	528 (2.2)	76 (1.6)	27	0						
² Kasachstan	501 (4.5)	84 (2.6)	27	529 (4.0)	84 (2.0)	26	1	529 (3.9)	85 (1.7)	26	0						
¹ Malta	496 (1.3)	78 (0.9)	28	519 (1.4)	79 (1.0)	28	0	516 (1.4)	79 (1.2)	30	-2						
³ Norwegen	495 (2.8)	68 (1.9)	29	517 (2.6)	73 (1.5)	29	0	521 (2.8)	76 (1.7)	29	0						
² Kroatien	490 (1.9)	67 (1.5)	30	514 (1.8)	73 (1.2)	30	0	523 (1.9)	75 (1.3)	28	2						
¹ ³ Neuseeland	486 (2.6)	83 (1.3)	31	509 (2.1)	81 (1.5)	32	-1	509 (1.9)	80 (1.0)	33	-1						
² Spanien	482 (2.9)	70 (1.5)	32	506 (2.5)	72 (1.0)	33	-1	508 (2.2)	75 (1.5)	34	-1						
Rumänien	482 (5.8)	105 (3.7)	33	513 (4.4)	93 (2.1)	31	2	513 (4.3)	93 (2.1)	31	0						
Polen	481 (2.2)	73 (1.1)	34	505 (2.0)	75 (1.4)	34	0	513 (1.8)	78 (1.5)	32	2						
Türkei	469 (4.7)	101 (4.1)	35	505 (3.0)	84 (1.7)	35	0	506 (3.0)	86 (1.5)	35	0						
² ³ Aserbaidschan	463 (5.8)	101 (2.6)	36	499 (4.7)	88 (2.2)	36	0	498 (4.5)	87 (1.8)	36	0						
Chile	462 (2.3)	81 (1.5)	37	492 (1.8)	78 (1.2)	37	0	496 (2.1)	79 (1.6)	37	0						
Thailand	458 (4.8)	80 (2.8)	38	489 (3.7)	75 (2.0)	38	0	491 (4.0)	77 (1.9)	38	0						
Armenien	452 (3.5)	89 (1.5)	39	487 (2.9)	80 (1.2)	39	0	486 (2.7)	81 (1.2)	40	-1						
² Georgien	450 (3.7)	90 (2.6)	40	486 (2.5)	79 (1.5)	40	0	490 (2.6)	82 (1.9)	39	1						
Bahrain	436 (3.3)	90 (1.9)	41	476 (2.2)	78 (1.5)	41	0	474 (2.2)	79 (1.3)	43	-2						
Vereinigte Arabische Emirate (VAE)	434 (2.0)	99 (1.2)	42	476 (1.4)	83 (1.1)	42	0	475 (1.6)	84 (1.2)	42	0						
Iran	431 (3.5)	93 (1.9)	43	474 (2.5)	78 (1.2)	43	0	476 (2.5)	79 (1.3)	41	2						
² Katar	413 (3.5)	106 (2.2)	44	465 (2.5)	83 (1.9)	44	0	463 (2.4)	85 (1.4)	44	0						
Saudi-Arabien	410 (5.3)	100 (4.1)	45	461 (3.5)	79 (3.2)	45	0	458 (3.4)	80 (2.7)	45	0						
⁴ Oman	385 (2.9)	104 (1.2)	46	448 (1.8)	77 (0.9)	46	0	446 (2.0)	77 (1.0)	46	0						
⁴ Tunesien	359 (3.9)	95 (1.8)	47	429 (2.4)	72 (1.0)	47	0	432 (2.5)	74 (1.3)	47	0						
² ⁴ Kuwait	342 (3.4)	101 (1.7)	48	423 (2.0)	69 (1.6)	49	-1	422 (1.8)	70 (1.0)	49	0						
⁴ Marokko	335 (4.0)	103 (2.8)	49	423 (2.5)	72 (2.0)	48	1	426 (2.4)	76 (2.2)	48	0						
⁴ Jemen	248 (6.0)	110 (2.6)	50	392 (2.3)	64 (1.4)	50	0	392 (2.2)	64 (1.4)	50	0						

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

4 = Sehr hoher Anteil an Schülerinnen und Schülern mit nicht skalierbaren Leistungswerten.

mit Deutschland in einer Leistungsgruppe waren und besser abschnitten, erzielen in Modell III leicht schlechtere Punktwerte als die Schülerinnen und Schüler in Deutschland. Insgesamt lässt sich zur Frage der curricularen Validität festhalten, dass das Abschneiden Deutschlands auf der TIMSS-Leistungsskala nicht auf das Testdesign zurückzuführen ist. Eine Modifikation des Testdesigns würde vermutlich nur zu einer geringfügigen Veränderung der Rangfolge führen, zumal unklar bleibt, ob solche Rangplatzverschiebungen sich auch für andere Teilnehmer ergeben hätten, wenn man die jeweils für diesen Staat curricular nicht validen Aufgaben eliminieren würde.

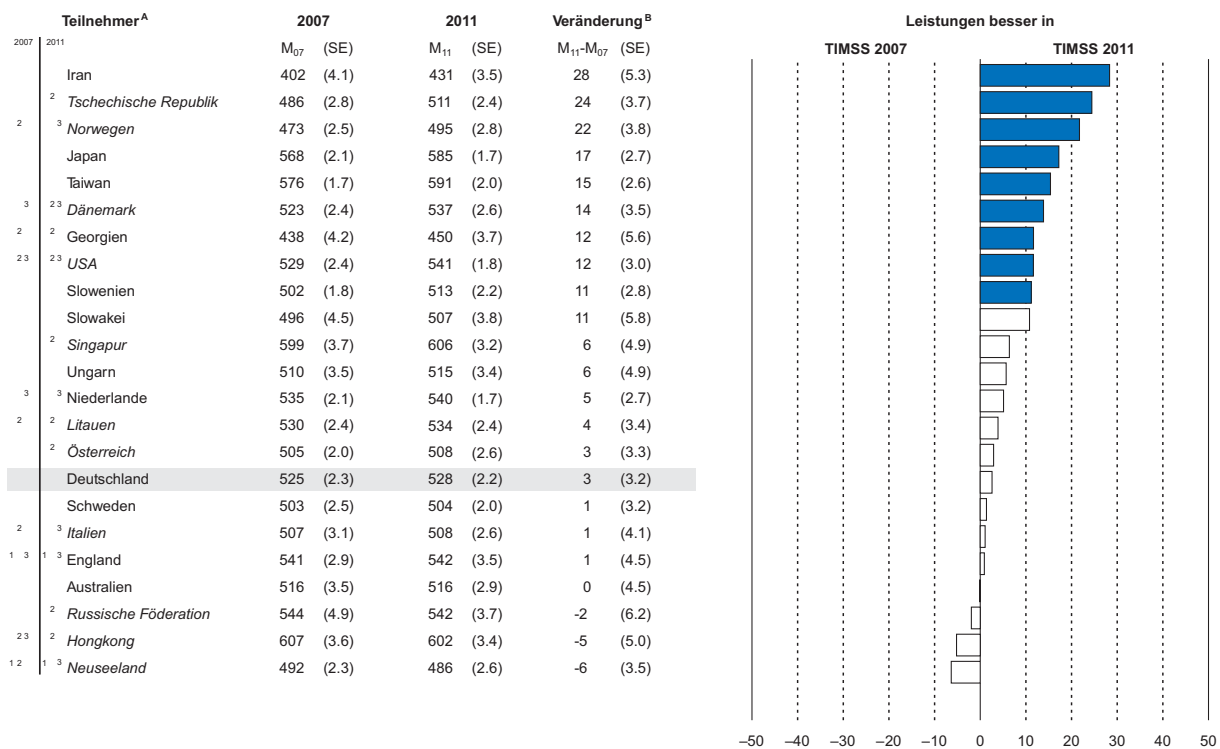
Vergleich der Testleistungen 2007 und 2011 auf der Gesamtskala: Da Deutschland im Jahr 2011 das zweite Mal an TIMSS-Grundschule teilgenommen hat, können erstmalig Unterschiede in der mathematischen Kompetenz von Schülerinnen und Schülern am Ende der Grundschulzeit berichtet werden. In Abbildung 3.6 sind für 23 Teilnehmer der insgesamt 30 Teilnehmerstaaten, die sowohl an TIMSS 2007 als auch an TIMSS 2011 teilnahmen, jeweils die nationalen Mittelwerte und die dazugehörigen Standardfehler der mathematischen Kompetenz im Vergleich gegenübergestellt. Außerdem ist für jeden Teilnehmerstaat die Differenz der Mittelwerte aus 2007 und 2011 mit dem dazugehörigen Standardfehler dargestellt.

Zunächst kann festgestellt werden, dass Deutschland zu den 14 Teilnehmern gehört, für die im Vergleich der Ergebnisse von 2011 und 2007 keine signifikanten Veränderungen festzustellen sind. In der Abbildung 3.6 ist für diese Teilnehmer der Differenzbalken in weiß dargestellt. Blau gefärbt sind dagegen die Differenzbalken der Staaten, in denen signifikante Veränderungen der Leistungsmittelwerte zu verzeichnen sind. Hier zeigt sich, dass kein Teilnehmer 2011 signifikant schlechtere Ergebnisse erzielte als 2007. Die größten Steigerungen in den Leistungsmittelwerten konnten Iran (+28), die Tschechische Republik (+24), Norwegen (+22) und Japan (+17) erzielen.

Extremgruppen: In Tabelle 3.12 sind die Mathematikleistungen nach Perzentilen aufgeführt. Um auch hier die Veränderungen zwischen 2007 und 2011 einschätzen zu können, werden die Werte im Vergleich dargestellt und entsprechend nur Ergebnisse für diejenigen Teilnehmer aufgenommen, die an beiden Erhebungen beteiligt waren. Die Staaten sind nach ihren Mittelwerten in der Gesamtskala Mathematik (vgl. Abbildung 3.5) geordnet. Anhand der Leistungswerte lassen sich insbesondere Aussagen über die Anteile von besonders leistungsschwachen und besonders leistungsstarken Schülerinnen und Schülern treffen.

Hierzu werden zunächst die 5 Prozent aller Kinder in den Blick genommen, die die schwächsten Mathematikleistungen erzielt haben (Leistungswerte zu Perzentil 5). Aus Tabelle 3.12 wird deutlich, dass 2011 die leistungsschwächsten 5 Prozent der Viertklässlerinnen und Viertklässler in Deutschland Leistungsmittelwerte erzielen, die höchstens 420 Punkte betragen. Während sich die Werte zu Perzentil 5 für die Russische Föderation (417), Dänemark (413), die USA (410) und Litauen (405) nicht signifikant von Deutschland unterscheiden, zeigt sich im internationalen Vergleich, dass die leistungsschwächsten Schülerinnen und Schüler in Hongkong (488), Singapur (464), Japan (460), Taiwan (459) und den Niederlanden (449) signifikant höhere Leistungsmittelwerte erreichen als die leistungsschwächsten in Deutschland. Signifikant niedrigere Leistungsmittelwerte zu Perzentil 5 als in Deutschland erzielen die leistungs-

Abbildung 3.6: Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 – Gesamtskala Mathematik



■ Statistisch signifikante Unterschiede (p < .05).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerbene erreichen nicht die internationalen Vorgaben.

A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen

Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

schwächsten Schülerinnen und Schüler in 14 Teilnehmerstaaten, zu denen beispielsweise auch England (385) und Ungarn (352) gehören.

Die 5 Prozent leistungsstärksten Schülerinnen und Schüler in Deutschland erreichen Leistungswerte, die mindestens 626 Punkte betragen. Während sich für Perzentil 95 die Niederlande (623), die Tschechische Republik (621), Slowenien (619), die Slowakei (626) und Italien (622) nicht signifikant von Deutschland unterscheiden, zeigt sich im internationalen Vergleich jedoch, dass die 5 Prozent leistungsstärksten Schülerinnen und Schüler von 11 Teilnehmern signifikant höhere Leistungsmittelwerte erreichen als die entsprechende Vergleichsgruppe in Deutschland. Zu diesen Teilnehmern gehören neben den Staaten, die auch auf der Gesamtskala Mathematik signifikante höhere Leistungswerte als Deutschland erzielen, auch Australien (652), Litauen (650) und Ungarn (650). Das ist ein Fingerzeig darauf, dass in Deutschland die Förderung der schwachen Schülerinnen und Schüler im internationalen Vergleich etwas besser zu gelingen scheint als die Förderung der leistungsstarken Viertklässlerinnen und Viertklässler.

Zudem können der Tabelle 3.12 die Veränderungen der Mathematikleistungen nach den Perzentilen zwischen TIMSS 2007 und TIMSS 2011 entnommen werden. Signifikant höhere beziehungsweise niedrigere Ergebnisse sind durch

ein Plus- beziehungsweise Minuszeichen markiert. Vergleicht man die Werte von 2011 und 2007 zunächst für Deutschland, so ergibt sich kein signifikanter Unterschied. Bei Betrachtung der anderen Teilnehmer zeigt sich, dass es in nahezu der Hälfte aller Teilnehmerstaaten – nämlich 10 von 24 – keine signifikanten Veränderungen auf allen vier Perzentilen gegeben hat. Wenn es zu bedeutsamen Veränderungen gekommen ist, dann waren diese in der Regel Steigerungen der Punktwerte. Lediglich in Neuseeland lagen die Werte für die Perzentile 75 und 95 sowie in den Niederlanden für das Perzentil 95 von 2011 signifikant unter denen von 2007.

Tabelle 3.12: Mathematikleistungen nach Perzentilen im Vergleich: TIMSS 2007 und TIMSS 2011 – Gesamtskala Mathematik

Teilnehmer ^A	TIMSS 2007				TIMSS 2011			
	Perzentile				Perzentile			
2007 2011	5 (SE)	25 (SE)	75 (SE)	95 (SE)	5 (SE)	25 (SE)	75 (SE)	95 (SE)
² Singapur	447 (6.5)	548 (5.1)	659 (4.0)	725 (4.1)	464 (7.4)	559 (5.3)	661 (4.2)	723 (3.2)
^{2,3} Hongkong	493 (9.1)	564 (4.2)	653 (4.0)	712 (5.3)	488 (11.0)	563 (3.3)	645 (3.4)	702 (2.7)
Taiwan	457 (4.1)	532 (2.1)	623 (2.2)	686 (2.1)	459 (6.4)	546+ (3.6)	642+ (2.3)	704+ (2.5)
Japan	438 (2.6)	520 (2.1)	620 (2.1)	688 (3.8)	460+ (6.9)	540+ (2.1)	635+ (2.4)	700+ (4.7)
^{1,3} England	392 (4.2)	487 (3.5)	600 (3.6)	676 (4.3)	385 (5.9)	483 (6.4)	605 (2.5)	677 (5.7)
² Russische Föderation	400 (4.1)	492 (5.4)	599 (5.1)	677 (9.8)	417+ (6.7)	493 (4.0)	593 (4.5)	660 (6.5)
^{2,3} USA	401 (3.8)	479 (2.3)	581 (3.0)	650 (5.2)	410 (3.3)	492+ (2.2)	593+ (2.3)	660 (1.5)
³ Niederlande	429 (7.0)	495 (2.9)	577 (2.8)	632 (2.3)	449+ (3.2)	505+ (2.6)	577 (1.9)	623- (2.7)
³ Dänemark	403 (9.9)	478 (3.9)	571 (2.4)	634 (4.8)	413 (5.3)	493+ (2.8)	585+ (3.0)	646 (4.5)
² Litauen	396 (3.7)	482 (3.5)	583 (2.9)	645 (5.6)	405 (4.5)	486 (2.9)	585 (2.7)	650 (2.1)
Deutschland	409 (10.3)	483 (2.6)	572 (2.2)	629 (2.6)	420 (7.8)	488 (2.9)	570 (2.3)	626 (2.1)
Australien	373 (8.2)	463 (4.1)	573 (4.2)	647 (3.9)	366 (6.2)	462 (3.1)	574 (4.0)	652 (6.0)
Ungarn	347 (12.4)	452 (6.6)	574 (3.7)	647 (4.6)	352 (8.3)	462 (3.8)	577 (2.8)	650 (2.8)
Slowenien	376 (4.0)	457 (2.5)	550 (2.3)	613 (2.8)	395+ (4.4)	468+ (2.3)	561+ (3.6)	619 (3.3)
² Tschechische Republik	361 (6.6)	440 (4.9)	536 (2.9)	597 (2.9)	387+ (6.0)	467+ (2.8)	560+ (2.3)	621+ (3.7)
² Österreich	386 (3.1)	462 (3.6)	552 (3.0)	612 (1.5)	401+ (4.3)	466 (3.7)	552 (2.9)	606 (4.4)
² Italien	374 (6.1)	457 (3.7)	558 (3.3)	629 (12.2)	386 (6.2)	461 (4.6)	557 (3.3)	622 (3.6)
Slowakei	350 (9.8)	446 (4.2)	553 (3.8)	623 (5.2)	365 (8.4)	460+ (3.6)	561 (2.2)	626 (5.2)
Schweden	388 (4.5)	459 (3.3)	548 (3.0)	608 (2.7)	388 (6.9)	462 (2.8)	549 (2.4)	610 (4.8)
² Norwegen	341 (7.0)	424 (4.6)	526 (2.8)	591 (5.6)	376+ (5.9)	451+ (4.0)	542+ (5.5)	604 (6.5)
^{1,2} Neuseeland	341 (7.3)	436 (2.6)	553 (2.5)	626 (3.6)	339 (6.7)	432 (2.3)	545- (2.7)	614- (3.5)
² Georgien	289 (6.3)	378 (6.8)	501 (5.8)	582 (6.4)	293 (8.4)	392 (6.0)	512 (2.7)	589 (6.0)
Iran	260 (5.6)	346 (4.9)	461 (4.2)	534 (6.4)	271 (6.4)	370+ (3.3)	496+ (3.9)	575+ (2.9)

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen

Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

+ = Mittelwert für 2011 signifikant höher als 2007 ($p < .05$).

- = Mittelwert für 2011 signifikant niedriger als 2007 ($p < .05$).

5.2 Kompetenzstufen

In Abschnitt 4.3 wurden fünf Stufen mathematischer Kompetenz unterschieden. Im Folgenden wird zunächst die Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen betrachtet und dabei insbesondere deren Anteil auf den Kompetenzstufen V und I in den Blick genommen. Anschließend wird ein Vergleich von TIMSS 2007 und 2011 gezogen.

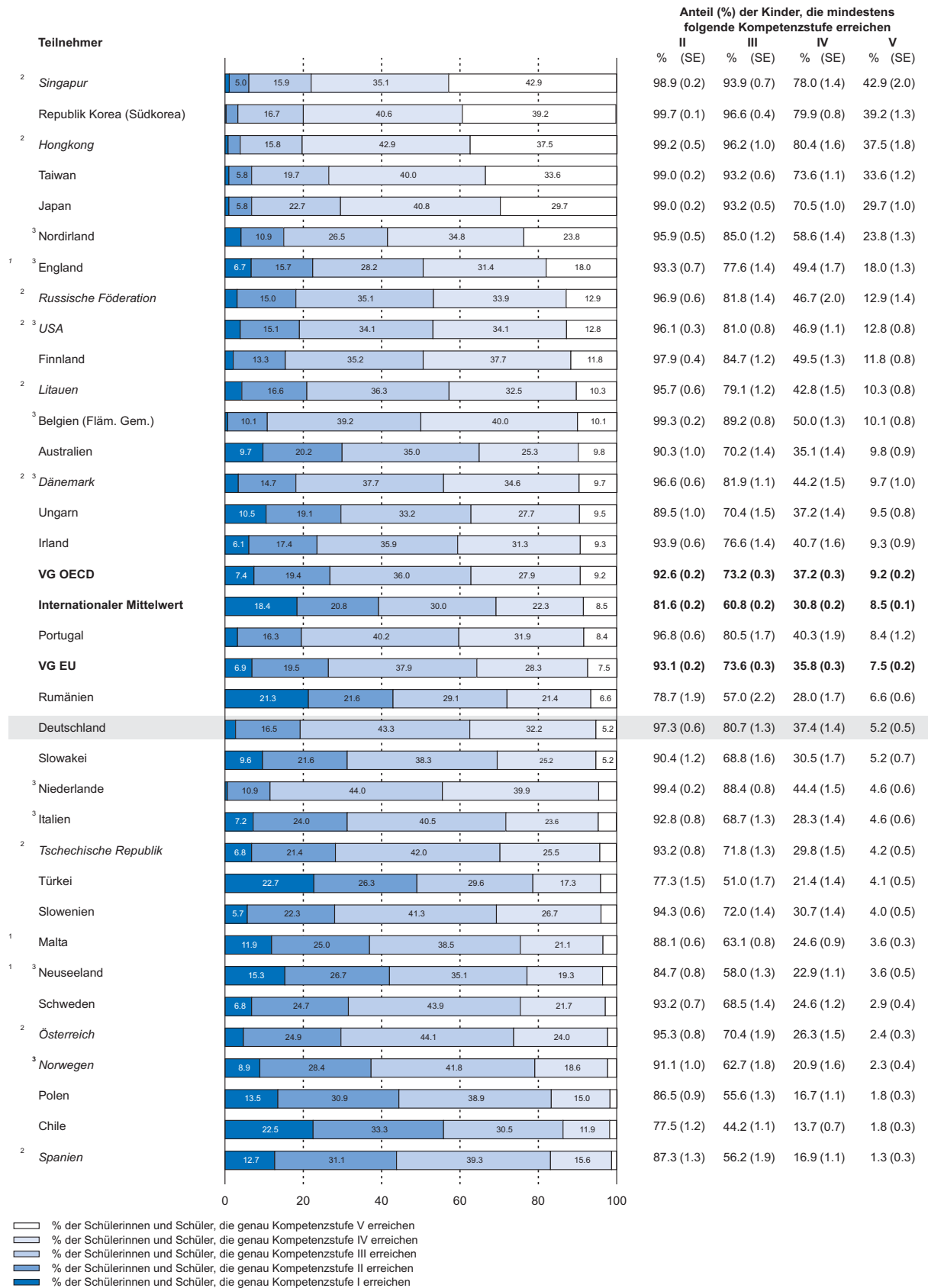
Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen: In Abbildung 3.7 ist die Verteilung der Schülerinnen und Schüler auf die fünf Kompetenzstufen im internationalen Vergleich dargestellt. Der linke Teil der Abbildung zeigt die Anteile der Schülerinnen und Schüler, die jeweils genau eine bestimmte Kompetenzstufe erreicht haben. Im rechten Teil der Abbildung sind zusätzlich die kumulierten Anteile angegeben, das heißt die Anteile der Schülerinnen und Schüler, die jeweils mindestens eine bestimmte Kompetenzstufe erreicht haben. Es zeigt sich, dass in Deutschland etwa 97 Prozent der Schülerinnen und Schüler mindestens Kompetenzstufe II, 81 Prozent mindestens Kompetenzstufe III, 37 Prozent mindestens Kompetenzstufe IV und lediglich 5 Prozent die höchste Kompetenzstufe erreichen.

Im internationalen Vergleich wird deutlich, dass in den Staaten Litauen (10%), Irland (9%) und Portugal (8%), die auf der Gesamtskala Mathematik in der gleichen Leistungsgruppe wie Deutschland liegen, der Anteil von Schülerinnen und Schülern auf der höchsten Kompetenzstufe vergleichsweise fast doppelt so hoch ist. Auch in den Vergleichsgruppen VG_{EU} und VG_{OECD} sind die entsprechenden Anteile mit 8 Prozent beziehungsweise 9 Prozent höher als in Deutschland. Anteile von mehr als 30 Prozent der Schülerinnen und Schüler auf Kompetenzstufe V erreichen insbesondere die leistungsstarken Teilnehmerstaaten wie Singapur (43%), Republik Korea (Südkorea) (39%), Hongkong (38%) und Taiwan (34%).

Am anderen Ende des Kompetenzspektrums zeigt sich folgendes Bild: In Deutschland beträgt der Anteil an Schülerinnen und Schülern auf der untersten Kompetenzstufe 3 Prozent. Allerdings erreichen zehn Teilnehmerstaaten, zu denen bis auf England (7%) alle Teilnehmer zählen, die auf der Gesamtskala Mathematik signifikant besser abschneiden, vergleichbare oder niedrigere Werte. Die mittleren Anteile der VG_{EU} und VG_{OECD} liegen mit 7 Prozent ebenfalls über dem deutschen Wert.

In Deutschland erreichen 81 Prozent der Schülerinnen und Schüler mindestens Kompetenzstufe III. Dieser Anteil wird knapp auch in den Vergleichsstaaten Litauen (79%), Portugal (81%) und Irland (77%) erreicht. Allerdings ist der Anteil von Viertklässlern, die mindestens Kompetenzstufe II erreichen, in den besonders leistungsstarken Teilnehmerstaaten mit mehr als 90 Prozent deutlich höher. In Deutschland erreicht etwa ein Fünftel der Schülerinnen und Schüler am Ende der Grundschulzeit mathematische Kompetenzen, die unterhalb der Kompetenzstufe III liegen. Somit muss davon ausgegangen werden, dass diese Schülergruppe mit erheblichen Schwierigkeiten beim Lernen in der Sekundarstufe I konfrontiert sein wird. Dieser Befund stellt eine beträchtliche Herausforderung an das Bildungssystem dar.

Abbildung 3.7: Prozentuale Verteilung der Schülerinnen und Schüler auf die fünf Kompetenzstufen (Mathematik) im internationalen Vergleich



Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

Vergleich der Kompetenzstufen 2007 und 2011: Die Tabelle 3.13 gibt die Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen im Vergleich von TIMSS 2007 und TIMSS 2011 für diejenigen Staaten wieder, die an beiden Untersuchungen teilgenommen haben. Diese sind von oben nach unten gemäß ihrer Leistungsmittelwerte auf der Gesamtskala Mathematik angeordnet. Es zeigt sich, dass sich für Deutschland ebenso wie für acht weitere Teilnehmer trotz geringfügiger nomineller Verschiebungen keine signifikanten Veränderungen der Verteilung der Viertklässlerinnen und Viertklässler auf die Kompetenzstufen ergeben. Zwölf Staaten gelingt es, die Anteile der Schülerinnen und Schüler auf Kompetenzstufe I beziehungsweise II signifikant zu reduzieren: Elf Teilnehmer zeigen signifikante Zuwächse auf den Stufen IV beziehungsweise V. Signifikante Zuwächse auf den Stufen I beziehungsweise II sind für keinen Teilnehmer zu beobachten. Im oberen Bereich sind es lediglich Neuseeland und die Niederlande, die auf Stufe V eine signifikante Abnahme des Anteils an leistungsstarken Schülerinnen und Schülern zu verzeichnen haben.

Tabelle 3.13: Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007 und in TIMSS 2011 (Angaben in Prozent)

Teilnehmer ^A	TIMSS 2007					TIMSS 2011				
	Kompetenzstufen					Kompetenzstufen				
	I	II	III	IV	V	I	II	III	IV	V
2007 2011	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)
Australien	8.5 (1.0)	21.0 (1.2)	35.6 (1.2)	26.1 (1.7)	8.8 (0.8)	9.7 (1.0)	20.2 (0.9)	35.0 (1.0)	25.3 (1.0)	9.8 (0.9)
³ ²³ Dänemark	4.7 (0.8)	19.0 (0.9)	40.3 (1.4)	29.3 (1.3)	6.6 (0.7)	3.4 (0.6)	14.7 – (0.8)	37.7 (1.2)	34.6 + (1.5)	9.7 + (1.0)
Deutschland	4.1 (0.5)	17.5 (0.9)	41.3 (0.9)	31.6 (1.1)	5.6 (0.5)	2.7 (0.6)	16.5 (1.1)	43.3 (1.3)	32.2 (1.1)	5.2 (0.5)
¹ ³ England	5.8 (0.7)	15.3 (0.9)	31.0 (1.1)	31.7 (0.9)	16.2 (1.2)	6.7 (0.7)	15.7 (1.1)	28.2 (1.1)	31.4 (1.4)	18.0 (1.3)
² Georgien	33.2 (2.0)	31.6 (1.2)	25.4 (1.2)	8.4 (0.9)	1.3 (0.4)	27.6 – (1.7)	31.2 (1.3)	28.9 (1.5)	10.6 (1.0)	1.8 (0.5)
²³ Hongkong	0.4 (0.1)	2.8 (0.4)	16.1 (1.3)	40.3 (1.3)	40.4 (2.2)	0.8 (0.5)	3.1 (0.7)	15.8 (1.1)	42.9 (1.2)	37.5 (1.8)
Iran	47.3 (2.0)	32.7 (1.4)	16.9 (1.2)	2.9 (0.5)	0.2 (0.1)	35.6 – (1.5)	31.2 (0.9)	23.8 + (0.9)	8.3 + (0.7)	1.1 + (0.2)
² ³ Italien	8.9 (1.0)	23.9 (1.1)	38.2 (1.6)	23.5 (1.3)	5.5 (0.7)	7.2 (0.8)	24.0 (1.0)	40.5 (1.2)	23.6 (1.2)	4.6 (0.6)
Japan	2.2 (0.4)	8.7 (0.6)	27.7 (0.9)	38.7 (1.3)	22.8 (1.2)	1.0 – (0.2)	5.8 – (0.5)	22.7 – (0.9)	40.8 (1.0)	29.7 + (1.0)
² ² Litauen	5.6 (0.7)	17.1 (1.3)	35.4 (1.1)	32.2 (1.2)	9.7 (0.7)	4.3 (0.6)	16.6 (1.1)	36.3 (1.2)	32.5 (1.2)	10.3 (0.8)
¹² ¹ ³ Neuseeland	14.7 (1.0)	24.8 (0.9)	34.1 (0.9)	21.3 (1.0)	5.1 (0.5)	15.3 (0.8)	26.7 (1.0)	35.1 (0.8)	19.3 (0.9)	3.6 – (0.5)
³ ³ Niederlande	2.0 (0.4)	14.2 (1.2)	42.1 (1.2)	35.1 (1.5)	6.6 (0.7)	0.6 – (0.2)	10.9 – (0.8)	44.0 (1.3)	39.9 + (1.2)	4.6 – (0.6)
² ³ Norwegen	17.1 (1.1)	31.0 (1.3)	36.9 (1.4)	13.3 (1.0)	1.7 (0.3)	8.9 – (1.0)	28.4 (1.4)	41.8 + (1.7)	18.6 + (1.4)	2.3 (0.4)
² Österreich	7.1 (0.8)	24.0 (1.1)	42.8 (1.3)	23.0 (0.9)	3.1 (0.3)	4.7 – (0.8)	24.9 (1.5)	44.1 (1.3)	24.0 (1.4)	2.4 (0.3)
² Russische Föderation	5.0 (0.7)	14.5 (1.3)	32.3 (1.4)	32.3 (1.2)	15.9 (1.8)	3.1 – (0.6)	15.0 (1.0)	35.1 (1.1)	33.9 (1.1)	12.9 (1.4)
Schweden	6.8 (0.7)	25.6 (1.0)	43.7 (1.0)	21.2 (1.3)	2.7 (0.3)	6.8 (0.7)	24.7 (1.1)	43.9 (1.4)	21.7 (1.1)	2.9 (0.4)
² Singapur	1.8 (0.3)	6.4 (0.8)	17.3 (1.1)	33.7 (1.3)	40.7 (2.1)	1.1 (0.2)	5.0 (0.6)	15.9 (1.1)	35.1 (1.4)	42.9 (2.0)
Slowakei	11.9 (1.5)	24.9 (1.1)	36.8 (1.0)	21.6 (1.2)	4.8 (0.7)	9.6 (1.2)	21.6 – (1.0)	38.3 (1.1)	25.2 + (1.3)	5.2 (0.7)
Slowenien	8.4 (0.6)	24.9 (1.0)	41.5 (0.9)	21.9 (0.9)	3.3 (0.4)	5.7 – (0.6)	22.3 (1.3)	41.3 (1.2)	26.7 + (1.1)	4.0 (0.5)
Taiwan	1.1 (0.2)	6.5 (0.5)	26.4 (1.2)	41.9 (1.4)	24.1 (1.2)	1.0 (0.2)	5.8 (0.5)	19.7 – (1.0)	40.0 (1.1)	33.6 + (1.2)
² Tschechische Republik	11.8 (1.1)	29.6 (1.2)	39.9 (1.3)	16.9 (1.5)	1.9 (0.4)	6.8 – (0.8)	21.4 – (1.1)	42.0 (1.4)	25.5 + (1.4)	4.2 + (0.5)
Ungarn	11.8 (1.2)	21.1 (1.1)	31.9 (1.2)	26.2 (1.2)	9.0 (0.8)	10.5 (1.0)	19.1 (1.0)	33.2 (0.9)	27.7 (1.0)	9.5 (0.8)
²³ USA	4.8 (0.5)	18.4 (0.9)	36.9 (0.9)	29.8 (0.9)	10.0 (0.8)	3.9 (0.3)	15.1 – (0.6)	34.1 – (0.9)	34.1 + (0.8)	12.8 + (0.8)

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülererebene erreichen nicht die internationalen Vorgaben.

A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen

Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

+ = Mittelwert für 2011 signifikant höher als 2007 ($p < .05$).

– = Mittelwert für 2011 signifikant niedriger als 2007 ($p < .05$).

5.3 Inhaltsbereiche

Die TIMSS-Rahmenkonzeption zur Erfassung mathematischer Kompetenzen ermöglicht es, Teilbereiche der mathematischen Kompetenz darzustellen und auf diese Weise ein differenzierteres Bild über Stärken und Schwächen der Schülerinnen und Schüler zu zeichnen. In diesem Abschnitt wird untersucht, ob sich im internationalen Vergleich unterschiedliche Stärken und Schwächen bei den Leistungen der Schülerinnen und Schüler in den Inhaltsbereichen (*Arithmetik, Geometrie/Messen, Umgang mit Daten*) zeigen. In Abschnitt 5.4 wird dann auf Ergebnisse zu den drei kognitiven Anforderungsbereichen (*Reproduzieren, Anwenden, Problemlösen*) eingegangen.

Kompetenzen in den Inhaltsbereichen im internationalen Vergleich: In der Abbildung 3.9 sind die erreichten Leistungsmittelwerte in den Inhaltsbereichen *Arithmetik, Geometrie/Messen* und *Umgang mit Daten* für die OECD- und EU-Staaten sowie für diejenigen Teilnehmer angeführt, die auf der Gesamtskala Mathematik signifikant bessere Ergebnisse erreichten als Deutschland. Auch hier sind die Mittelwerte für alle Teilnehmerstaaten (Internationaler Mittelwert) sowie die Mittelwerte für die VG_{OECD} und VG_{EU} aufgeführt. Die für die folgende Abbildung gewählte Darstellungsweise ist der Abbildung 3.5 entlehnt, allerdings werden in Abbildung 3.8 die Ergebnisse nach Inhaltsbereich getrennt dargestellt, wobei die Ergebnisse spaltenweise zu lesen sind. Für jeden Teilnehmer sind in der oben dargestellten Tabelle nach Inhaltsbereichen differenziert die jeweiligen Leistungsmittelwerte (M), ihre Standardabweichungen (SD) sowie die Standardfehler (SE) dieser beiden Werte angegeben. Zudem ist auch im unteren Teil die Streuung der Leistungsmittelwerte nach den Inhaltsbereichen getrennt durch die Darstellung der Perzentilbänder illustriert. Signifikante Unterschiede der Leistungsmittelwerte der Teilnehmer zu Deutschland sind jeweils durch + beziehungsweise – gekennzeichnet. Teilnehmerstaaten, in denen Leistungsmittelwerte der Schülerinnen und Schüler in allen drei Bereichen signifikant besser oder schlechter ausfallen als die Leistungsmittelwerte der Viertklässlerinnen und Viertklässler in Deutschland, sind außerdem durch eine Einrahmung hervorgehoben.

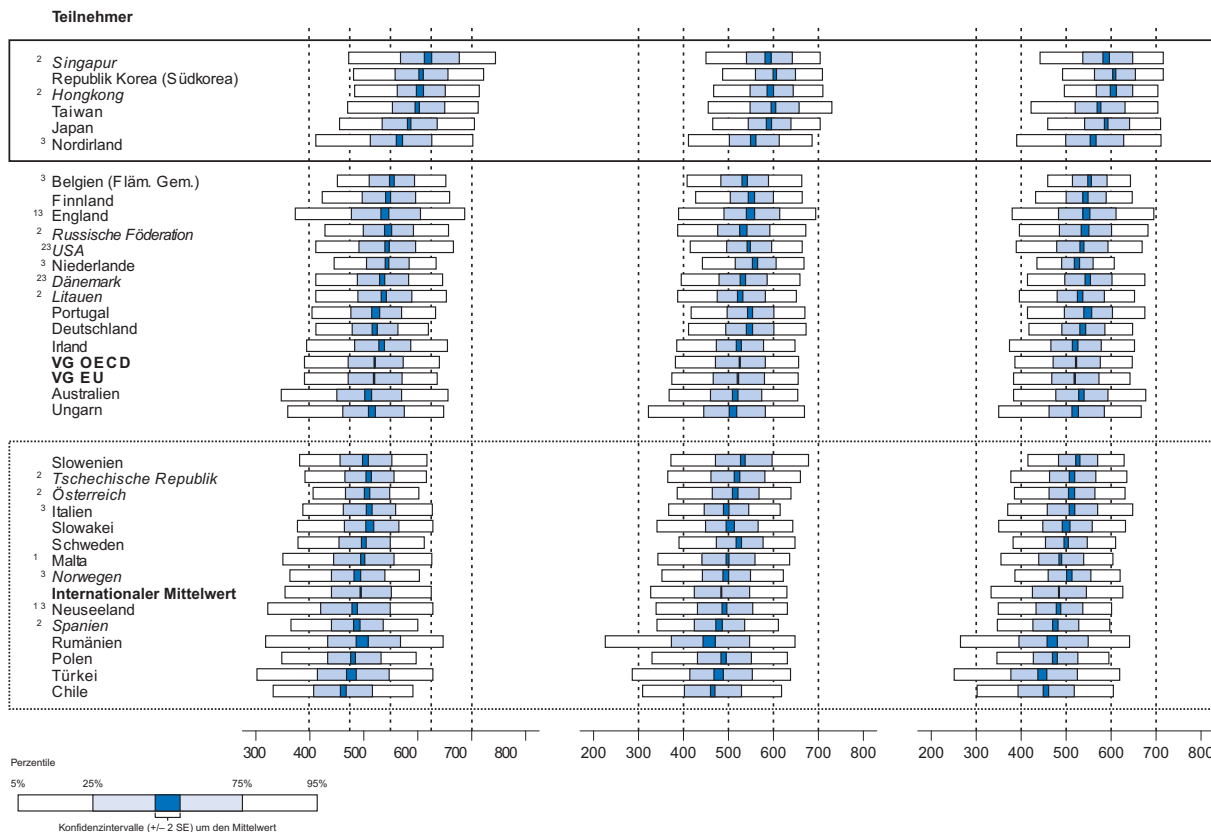
Für die Betrachtung des Ergebnisses zeigt sich, dass die Grundschülerinnen und Grundschüler in Deutschland im Bereich *Arithmetik* 520 Punkte, im Bereich *Geometrie/Messen* 536 Punkte und im Bereich *Umgang mit Daten* 546 Punkte erreichen. Die Leistungen im Bereich *Arithmetik* liegen damit unterhalb des Gesamtmittelwerts Mathematik (528), während sich die Leistungen in den Bereichen *Geometrie/Messen* und *Umgang mit Daten* oberhalb des Gesamtmittelwerts verorten lassen.

Vergleicht man diese Differenz mit den Ergebnissen der Teilnehmer, deren Leistungsmittelwert auf der Gesamtskala Mathematik sich nicht signifikant von dem der Viertklässlerinnen und Viertklässler in Deutschland unterscheiden lässt, so zeigt sich, dass diese Differenz zwischen den Leistungsmittelwerten nach Inhaltsbereichen in Portugal ebenfalls 26 Punkte beträgt, während sie in Irland (13) und in Litauen (11) kleiner ist.

Im internationalen Vergleich zeigt sich zudem, dass insgesamt sechs Teilnehmer in allen drei Inhaltsbereichen signifikant bessere Leistungsmittelwerte erreichen als Deutschland und 14 Teilnehmerstaaten in allen drei Inhaltsbereichen signifikant schlechtere Ergebnisse erzielen. Eine nach Inhaltsbereichen differenzierte Betrachtung verdeutlicht, dass insgesamt in *Arithmetik* 15, in *Geometrie/*

Abbildung 3.8: Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den Inhaltsbereichen *Arithmetik, Geometrie/Messen und Umgang mit Daten*

Teilnehmer	Arithmetik			Geometrie/Messen			Umgang mit Daten		
	M (SE)	SD (SE)		M (SE)	SD (SE)		M (SE)	SD (SE)	
² Singapur	619 (3.4) +	81 (2.3)		589 (3.6) +	83 (1.8)		588 (3.4) +	77 (1.7)	
Republik Korea (Südkorea)	606 (2.0) +	73 (1.1)		607 (1.7) +	68 (1.1)		603 (1.9) +	67 (1.3)	
² Hongkong	604 (3.3) +	71 (3.1)		605 (3.4) +	65 (2.9)		593 (3.6) +	76 (3.7)	
Taiwan	599 (2.0) +	73 (1.1)		573 (2.1) +	85 (1.5)		600 (2.6) +	83 (1.9)	
Japan	584 (1.6) +	76 (1.0)		589 (2.0) +	76 (1.5)		590 (2.9) +	73 (1.3)	
³ Nordirland	566 (2.9) +	87 (1.9)		560 (3.3) +	97 (1.6)		555 (3.0) +	84 (2.6)	
³ Belgien (Fläm. Gem.)	552 (2.2) +	62 (1.2)		552 (2.0) +	56 (1.6)		536 (3.0) -	77 (1.7)	
Finnland	545 (2.3) +	73 (1.2)		543 (2.9) +	66 (1.9)		551 (3.5) +	73 (1.5)	
^{1,3} England	539 (3.7) +	96 (2.2)		545 (3.9) +	96 (2.2)		549 (4.6) +	93 (2.0)	
² Russische Föderation	545 (3.3) +	70 (1.2)		542 (4.3) +	87 (1.7)		533 (4.1) -	87 (1.6)	
^{2,3} USA	543 (2.0) +	78 (1.0)		535 (2.2) +	85 (1.7)		545 (1.8) +	75 (1.3)	
³ Niederlande	543 (1.7) +	58 (1.5)		524 (2.9) -	53 (1.5)		559 (2.9) +	69 (1.9)	
^{2,3} Dänemark	534 (2.4) +	71 (1.7)		548 (3.0) +	79 (1.7)		532 (3.0) -	81 (1.7)	
² Litauen	537 (2.4) +	74 (1.1)		531 (3.0) +	78 (1.4)		526 (3.0) -	80 (2.3)	
Portugal	522 (3.7) +	70 (2.2)		548 (4.4) +	80 (2.4)		548 (2.8) +	77 (2.1)	
Deutschland	520 (2.3)	63 (1.4)		536 (2.6)	70 (1.6)		546 (2.8)	79 (1.7)	
Irland	533 (2.6) +	79 (1.6)		520 (3.1) -	84 (1.8)		523 (2.8) -	80 (1.5)	
VG OECD	520 (0.5)	76 (0.3)		522 (0.6)	79 (0.4)		525 (0.6)	83 (0.4)	
VG EU	519 (0.6)	75 (0.4)		519 (0.7)	79 (0.4)		521 (0.8)	85 (0.5)	
Australien	508 (3.2) -	93 (2.4)		534 (3.0) +	89 (2.7)		515 (3.1) -	86 (2.0)	
Ungarn	515 (3.2) -	87 (2.0)		520 (3.6) -	96 (2.7)		510 (4.2) -	105 (2.8)	
Slowenien	503 (2.7) -	72 (1.2)		526 (2.3) -	66 (0.9)		532 (2.6) -	94 (2.0)	
² Tschechische Republik	509 (2.5) -	69 (1.4)		513 (3.0) -	78 (2.4)		519 (3.1) -	90 (2.1)	
² Österreich	506 (2.5) -	59 (1.2)		512 (3.4) -	75 (1.7)		515 (3.1) -	77 (1.5)	
³ Italien	510 (2.7) -	73 (1.5)		513 (3.1) -	84 (2.0)		495 (3.1) -	75 (1.8)	
Slowakei	511 (3.7) -	77 (2.4)		500 (4.3) -	85 (2.8)		504 (4.6) -	91 (2.8)	
Schweden	500 (2.2) -	71 (1.2)		500 (2.4) -	69 (1.2)		523 (3.0) -	78 (1.5)	
¹ Malta	498 (1.9) -	84 (1.4)		487 (1.5) -	76 (1.7)		498 (1.6) -	89 (1.5)	
³ Norwegen	488 (3.1) -	73 (1.6)		507 (3.0) -	72 (1.7)		494 (3.2) -	82 (2.2)	
Internationaler Mittelwert	494 (0.5)	82 (0.3)		484 (0.5)	89 (0.3)		484 (0.5)	92 (0.3)	
^{1,3} Neuseeland	483 (2.5) -	94 (1.7)		483 (2.5) -	77 (1.7)		491 (2.7) -	89 (1.8)	
² Spanien	487 (3.0) -	71 (1.5)		476 (3.0) -	75 (1.9)		479 (3.6) -	83 (2.1)	
Rumänien	497 (5.6) -	100 (3.2)		469 (5.7) -	114 (3.4)		457 (6.8) -	128 (5.4)	
Polen	480 (2.2) -	75 (1.3)		475 (2.7) -	75 (1.2)		489 (2.9) -	91 (1.5)	
Türkei	477 (4.5) -	100 (3.4)		447 (5.0) -	111 (3.4)		478 (5.2) -	108 (4.8)	
Chile	462 (2.7) -	79 (1.4)		455 (3.0) -	92 (2.2)		465 (2.5) -	93 (1.5)	



Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss

☐ In allen drei Bereichen signifikant positiv von den deutschen Mittelwerten abweichende Staaten ($p < .05$).

☐ In allen drei Bereichen signifikant negativ von den deutschen Mittelwerten abweichende Staaten ($p < .05$).

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schulebene erreichen nicht die internationalen Vorgaben.

+ = Mittelwert statistisch signifikant höher als der deutsche Mittelwert ($p < .05$)

- = Mittelwert statistisch signifikant niedriger als der deutsche Mittelwert ($p < .05$)

Messen 17 und im Bereich *Umgang mit Daten* 21 Teilnehmer schlechter abschneiden als Deutschland. Diese Tendenz verstärkt sich, wenn man berücksichtigt, dass in *Arithmetik* immerhin 15, in *Geometrie/Messen* nur 9 und in *Umgang mit Daten* gar nur 7 Teilnehmer signifikant höhere Leistungsmittelwerte aufweisen als Deutschland.

Vergleicht man mit den Leistungsmittelwerten der VG_{EU} und VG_{OECD} zeigt sich ein ähnliches Bild: Während sich keine signifikanten Unterschiede für den Bereich *Arithmetik* ergeben, fallen die Mittelwerte für die beiden anderen Inhaltsbereiche signifikant höher aus. Hingegen zeigen sich beim Vergleich der Ergebnisse zwischen den Bereichen für die VG_{EU} beziehungsweise VG_{OECD} – im Unterschied zu Deutschland – keine nennenswerten Unterschiede.

Dass im nationalen und internationalen Vergleich schlechte Abschneiden der Schülerinnen und Schüler im Inhaltsbereich *Arithmetik* lässt sich unter Umständen zu einem gewissen Grad auch vor dem Hintergrund der vergleichsweise hohen Anteile von nicht curricular validen Aufgaben in diesem Bereich interpretieren.

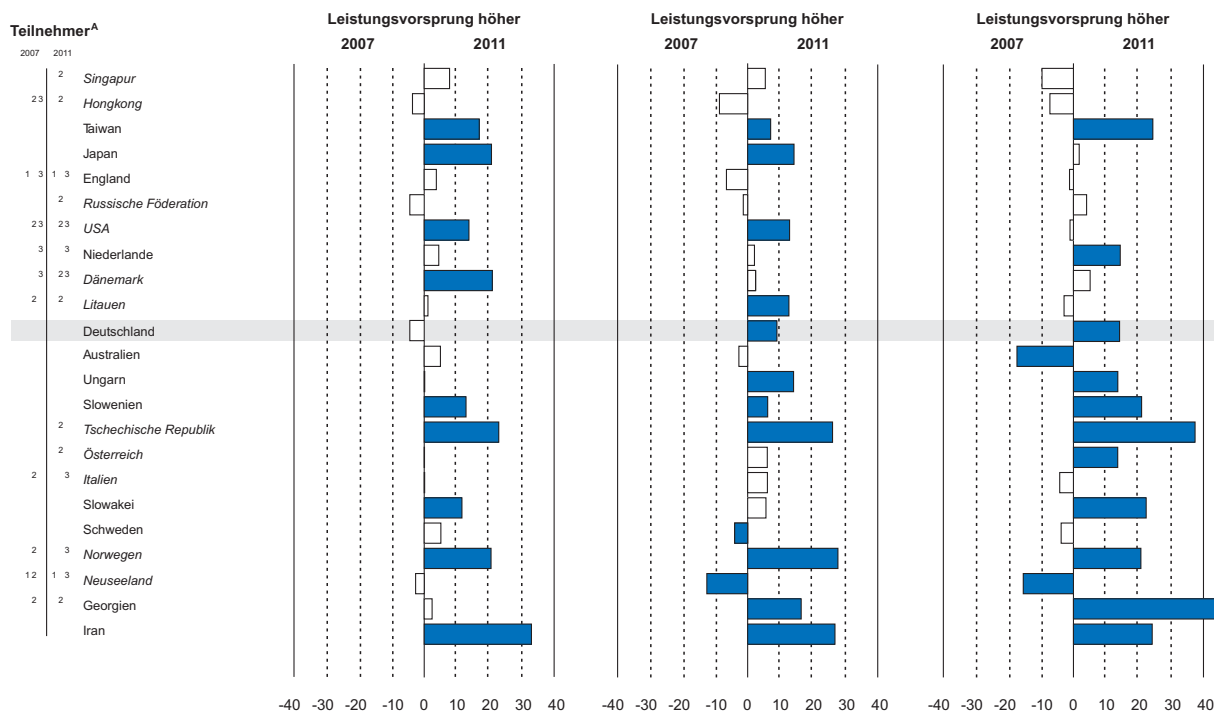
Vergleich der Kompetenzen in den Inhaltsbereichen 2007 und 2011: In Abbildung 3.9 sind Unterschiede in Bezug auf die mathematischen Kompetenzen zwischen den Studienzyklen TIMSS 2007 und TIMSS 2011 getrennt nach den Inhaltsbereichen *Arithmetik*, *Geometrie/Messen* und *Umgang mit Daten* dargestellt. Die Gestaltung der Abbildung sowie die Auswahl der Teilnehmerstaaten ist an Abbildung 3.8 angelehnt, jedoch werden hier die Differenzen zwischen den Leistungsmittelwerten der Schülerinnen und Schüler numerisch in der oben dargestellten Tabelle aufgeführt und durch Differenzbalken im unteren Teil der Abbildung illustriert. Positive Werte in der Spalte Veränderung bedeuten, dass im Jahr 2011 mehr Punkte erzielt wurden als 2007. Handelt es sich um eine signifikante Veränderung der Leistungsmittelwerte zwischen 2007 und 2011, ist dies einer blauen Färbung der Differenzbalken zu entnehmen.

Betrachtet man nun die Ergebnisse in Abbildung 3.9, lässt sich für Deutschland festhalten, dass sich für den Bereich *Arithmetik* kein signifikanter Unterschied ergibt. Dagegen zeigt sich für den Bereich *Geometrie/Messen* eine signifikante positive Differenz von 9 Punkten und im Bereich *Umgang mit Daten* eine signifikante positive Differenz von 14 Punkten. Auch wenn diese erfreulichen Entwicklungen zu verzeichnen sind, reichen sie aufgrund der leichten Verschlechterung im Bereich *Arithmetik* nicht aus, um von einer signifikanten Steigerung auf der Gesamtskala Mathematik sprechen zu können (vgl. Abschnitt 5.1).

Die Abbildung zeigt zudem, dass zwar für die überwiegende Zahl der dargestellten Teilnehmer positive Differenzen zu verzeichnen, jedoch die Muster der Veränderung nicht einheitlich sind. Während sich für einige Teilnehmer signifikant positive Differenzen in allen drei Inhaltsbereichen ergeben, zeigen sich in einigen Fällen – wie auch in Deutschland – lediglich signifikante Veränderungen für einzelne Bereiche. Allerdings lassen sich auch in drei Fällen signifikante Verschlechterungen feststellen: So erreichte beispielsweise Australien im Bereich *Umgang mit Daten* 2011 17 Punkte weniger als 2007.

Abbildung 3.9: Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 in den Inhaltsbereichen Arithmetik, Geometrie/Messen und Umgang mit Daten

Teilnehmer ^A	Arithmetik			Geometrie/Messen			Umgang mit Daten		
	2007	2011	Veränderung ^B	2007	2011	Veränderung ^B	2007	2011	Veränderung ^B
	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)
² Singapur	611 (4.1)	619 (3.4)	8 (5.4)	584 (4.4)	589 (3.6)	5 (5.7)	597 (3.9)	588 (3.4)	-9 (5.2)
²³ ² Hongkong	608 (3.7)	604 (3.3)	-4 (5.0)	613 (3.7)	605 (3.4)	-9 (5.1)	600 (3.4)	593 (3.6)	-7 (5.0)
Taiwan	583 (1.8)	599 (2.0)	17 (2.7)	566 (2.7)	573 (2.1)	7 (3.4)	576 (2.3)	600 (2.6)	24 (3.4)
Japan	564 (2.1)	584 (1.6)	20 (2.7)	575 (2.6)	589 (2.0)	14 (3.3)	588 (3.5)	590 (2.9)	2 (4.6)
¹ ³ ¹ ³ England	535 (3.1)	539 (3.7)	4 (4.8)	552 (3.3)	545 (3.9)	-6 (5.1)	551 (3.1)	549 (4.6)	-1 (5.6)
² Russische Föderation	549 (4.4)	545 (3.3)	-4 (5.4)	543 (6.2)	542 (4.3)	-1 (7.5)	529 (6.2)	533 (4.1)	4 (7.4)
²³ ²³ USA	529 (2.6)	543 (2.0)	13 (3.3)	522 (3.0)	535 (2.2)	13 (3.7)	546 (2.9)	545 (1.8)	-1 (3.4)
³ ³ Niederlande	539 (2.2)	543 (1.7)	4 (2.7)	522 (2.7)	524 (2.9)	2 (4.0)	545 (2.8)	559 (2.9)	14 (4.0)
³ ²³ Dänemark	513 (2.9)	534 (2.4)	21 (3.8)	546 (3.2)	548 (3.0)	2 (4.4)	527 (4.2)	532 (3.0)	5 (5.2)
² ² Litauen	536 (2.2)	537 (2.4)	1 (3.3)	518 (3.0)	531 (3.0)	12 (4.2)	529 (3.6)	526 (3.0)	-3 (4.7)
Deutschland	524 (2.2)	520 (2.3)	-4 (3.2)	527 (2.4)	536 (2.6)	9 (3.6)	532 (3.7)	546 (2.8)	14 (4.6)
Australien	503 (3.5)	508 (3.2)	5 (4.8)	536 (3.7)	534 (3.0)	-3 (4.8)	532 (4.1)	515 (3.1)	-17 (5.2)
Ungarn	515 (3.5)	515 (3.2)	0 (4.8)	507 (3.9)	520 (3.6)	14 (5.3)	497 (4.2)	510 (4.2)	13 (6.0)
Slowenien	490 (1.9)	503 (2.7)	13 (3.2)	520 (2.0)	526 (2.3)	6 (3.0)	512 (2.6)	532 (2.6)	21 (3.7)
² Tschechische Republik	486 (2.9)	509 (2.5)	23 (3.8)	487 (3.3)	513 (3.0)	26 (4.5)	482 (4.2)	519 (3.1)	37 (5.2)
² Österreich	506 (2.2)	506 (2.5)	0 (3.3)	506 (2.8)	512 (3.4)	6 (4.4)	502 (3.4)	515 (3.1)	13 (4.6)
² ³ Italien	510 (3.0)	510 (2.7)	0 (4.0)	507 (3.6)	513 (3.1)	6 (4.8)	499 (4.1)	495 (3.1)	-4 (5.1)
Slowakei	500 (3.9)	511 (3.7)	11 (5.4)	494 (5.3)	500 (4.3)	6 (6.8)	482 (5.4)	504 (4.6)	22 (7.1)
Schweden	495 (2.5)	500 (2.2)	5 (3.3)	503 (2.9)	500 (2.4)	-4 (3.8)	527 (3.2)	523 (3.0)	-4 (4.3)
² ³ Norwegen	468 (2.8)	488 (3.1)	20 (4.2)	479 (3.6)	507 (3.0)	27 (4.7)	474 (2.9)	494 (3.2)	20 (4.4)
¹² ¹ ³ Neuseeland	485 (2.6)	483 (2.5)	-3 (3.6)	495 (2.6)	483 (2.5)	-12 (3.6)	506 (3.0)	491 (2.7)	-15 (4.1)
² ² Georgien	470 (3.7)	473 (3.1)	2 (4.8)	395 (5.9)	411 (4.3)	16 (7.3)	390 (5.4)	433 (4.0)	43 (6.7)
Iran	407 (3.5)	440 (3.3)	32 (4.8)	408 (3.9)	435 (3.9)	26 (5.6)	374 (5.0)	397 (4.3)	24 (6.6)



■ Statistisch signifikante Unterschiede (p < .05).

Kursiv gesetzt sind Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerbene erreichen nicht die internationalen Vorgaben.

A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

5.4 Kognitive Anforderungen

Wie in Abschnitt 3 ausgeführt, werden in TIMSS nicht nur die inhaltsbezogenen Kompetenzen, sondern auch die fachspezifischen kognitiven Anforderungen in den Blick genommen. Analog zu den Ergebnissen des vorangehenden Abschnitts sollen daher in diesem Abschnitt die Ergebnisse entlang der drei kognitiven Anforderungsbereiche berichtet werden.

Kompetenzen in den Anforderungsbereichen im internationalen Vergleich: In der Abbildung 3.10 sind die erreichten Leistungsmittelwerte in den Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* im internationalen Vergleich dargestellt. Die Gestaltung der Abbildung sowie die Auswahl der Teilnehmerstaaten entspricht Abbildung 3.8.

Bei der Betrachtung des Ergebnisses zeigt sich, dass die Viertklässlerinnen und Viertklässler in Deutschland im Bereich *Reproduzieren* 524 Punkte, im Bereich *Anwenden* 528 Punkte und im Bereich *Problemlösen* 532 Punkte erreichen. Die Leistungen in den Bereichen *Reproduzieren* und *Anwenden* liegen damit unterhalb des Gesamtmittelwerts Mathematik (528), während sich die Leistungen im Bereich *Problemlösen* (532) oberhalb des Gesamtmittelwerts verorten lassen. Die Spanne zwischen dem höchsten und dem niedrigsten Leistungswert beträgt in Deutschland also 8 Punkte.

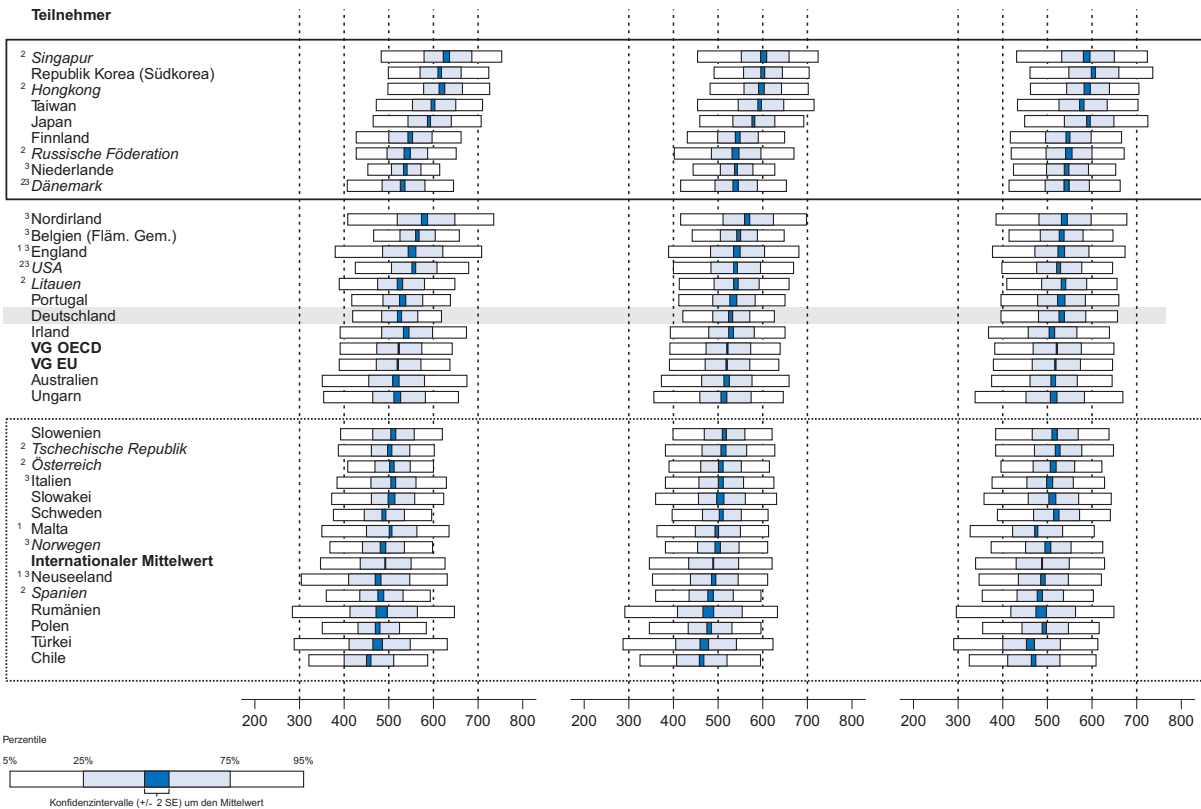
Vergleicht man diese Differenz mit den Ergebnissen der Teilnehmer, deren Leistungsmittelwert sich auf der Gesamtskala Mathematik nicht signifikant von dem der Viertklässlerinnen und Viertklässler in Deutschland unterscheiden lässt, so zeigt sich, dass diese Differenz zwischen den Leistungsmittelwerten nach kognitiven Anforderungsbereichen in Portugal lediglich 3 Punkte beträgt, während sie in Irland (29) und in Litauen (15) wesentlich größer ausfällt.

Im internationalen Vergleich zeigt sich zudem, dass insgesamt neun Teilnehmer in allen drei kognitiven Anforderungsbereichen signifikant bessere Leistungsmittelwerte erzielen als Deutschland und 14 Teilnehmer in allen drei kognitiven Anforderungsbereichen signifikant schlechtere Ergebnisse erreichen. Eine nach kognitiven Anforderungsbereichen differenzierte Betrachtung verdeutlicht, dass insgesamt beim *Reproduzieren* 16, beim *Anwenden* 16 und beim *Problemlösen* 17 Teilnehmer schlechter abschneiden als Deutschland. Vergleicht man auch hier mit den Leistungsmittelwerten der VG_{EU} und VG_{OECD} zeigt sich folgendes Bild: Während sich keine signifikanten Unterschiede für den Bereich *Reproduzieren* ergeben, fallen die Mittelwerte für die beiden anderen Anforderungsbereiche signifikant höher aus. Beim Vergleich der Ergebnisse zwischen den Bereichen für die VG_{EU} beziehungsweise VG_{OECD} zeigen sich hier – im Unterschied zu Deutschland – keine nennenswerten Unterschiede.

Insgesamt zeigt sich für TIMSS 2011, dass im Bereich *Problemlösen* im Vergleich zu Deutschland weniger Staaten signifikant besser und mehr Staaten signifikant schwächer abschneiden als im Bereich *Anwenden* und insbesondere als im Bereich *Reproduzieren*. Ein Grund dafür könnte sein, dass 30 Prozent der Testaufgaben im Bereich *Reproduzieren* als curricular nicht valide eingestuft worden sind, während dieser Prozentsatz für die Anforderungsbereiche *Anwenden* (18%) beziehungsweise *Problemlösen* (9%) zum Teil deutlich niedriger lag. Vielleicht zeigen sich hier aber auch besondere Stärken der Schülerinnen und Schüler in Deutschland.

Abbildung 3.10: Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den kognitiven Anforderungsbereichen *Reproduzieren, Anwenden und Problemlösen*

Teilnehmer	Reproduzieren		Anwenden		Problemlösen	
	M (SE)	SD (SE)	M (SE)	SD (SE)	M (SE)	SD (SE)
² Singapur	629 (3.5) +	81 (2.0)	602 (3.4) +	81 (2.0)	588 (3.7) +	88 (2.0)
Republik Korea (Südkorea)	614 (2.0) +	69 (1.9)	600 (2.2) +	65 (1.0)	603 (2.3) +	84 (1.4)
² Hongkong	619 (3.2) +	69 (3.4)	597 (3.2) +	68 (3.1)	589 (3.4) +	74 (3.2)
Taiwan	599 (2.1) +	72 (1.1)	593 (2.0) +	79 (1.2)	577 (2.5) +	81 (1.8)
Japan	590 (1.7) +	73 (1.7)	579 (1.6) +	71 (1.5)	592 (2.0) +	84 (2.4)
Finnland	548 (2.6) +	72 (2.2)	544 (2.7) +	67 (1.3)	546 (2.2) +	76 (1.7)
² Russische Föderation	541 (3.4) +	67 (1.4)	539 (3.9) +	82 (1.5)	548 (3.6) +	77 (1.4)
³ Niederlande	537 (2.0) +	49 (1.1)	540 (1.6) +	55 (1.1)	543 (2.6) +	70 (1.6)
² Dänemark	531 (2.6) +	72 (1.8)	539 (2.9) +	72 (2.1)	543 (2.7) +	76 (2.1)
³ Nordirland	580 (3.4) +	99 (1.5)	565 (2.9) +	86 (2.0)	538 (3.3)	89 (1.9)
³ Belgien (Fläm. Gem.)	564 (1.9) +	58 (1.3)	546 (2.2) +	62 (1.3)	532 (2.7)	71 (1.2)
¹³ England	552 (4.3) +	101 (1.8)	542 (3.7) +	90 (2.3)	531 (3.7)	91 (1.9)
²³ USA	556 (2.1) +	77 (1.4)	539 (2.1) +	82 (1.2)	525 (2.2)	75 (1.6)
² Litauen	525 (3.0)	78 (1.6)	540 (2.5) +	75 (1.3)	536 (2.5)	75 (1.4)
Portugal	531 (3.5)	67 (2.1)	534 (3.9)	72 (2.2)	531 (4.1)	79 (2.5)
Deutschland	524 (2.3)	61 (1.6)	528 (2.3)	62 (1.6)	532 (3.0)	79 (1.6)
Irland	539 (3.1) +	86 (2.2)	529 (2.7)	78 (1.9)	510 (3.1) -	82 (1.7)
VG OECD	522 (0.6)	76 (0.4)	521 (0.6) -	75 (0.4)	521 (0.6) -	81 (0.4)
VG EU	520 (0.7)	75 (0.4)	519 (0.7) -	75 (0.4)	518 (0.7) -	81 (0.4)
Australien	516 (3.5)	97 (2.7)	519 (3.0)	86 (2.1)	513 (2.6)	81 (2.2)
Ungarn	519 (3.8)	92 (2.7)	513 (3.3) -	88 (2.1)	514 (3.7) -	100 (2.6)
Slowenien	510 (2.8) -	69 (1.3)	514 (2.3) -	68 (1.3)	516 (2.9) -	78 (1.6)
² Tschechische Republik	502 (2.4) -	65 (1.8)	512 (2.9) -	75 (2.2)	523 (2.7)	80 (2.2)
² Österreich	507 (2.5) -	59 (1.3)	506 (2.6) -	68 (1.1)	513 (3.3) -	69 (1.5)
³ Italien	510 (2.7) -	75 (1.6)	506 (2.9) -	74 (2.3)	505 (3.4) -	77 (2.9)
Slowakei	506 (3.8) -	76 (2.7)	505 (4.0) -	82 (2.6)	511 (3.9) -	86 (2.6)
Schweden	489 (2.2) -	67 (1.6)	507 (2.2) -	65 (1.3)	520 (3.0) -	77 (1.8)
Malta	504 (1.5) -	86 (1.7)	497 (2.0) -	76 (1.4)	475 (1.7) -	85 (1.3)
³ Norwegen	487 (3.1) -	70 (1.9)	499 (3.0) -	70 (1.5)	501 (3.3) -	75 (1.6)
Internationaler Mittelwert	492 (0.5) -	85 (0.3)	489 (0.5) -	84 (0.3)	488 (0.5) -	88 (0.3)
¹³ Neuseeland	476 (3.2) -	100 (1.7)	490 (2.4) -	79 (1.3)	490 (2.5) -	83 (1.8)
² Spanien	482 (3.3) -	71 (2.0)	483 (3.1) -	72 (1.5)	483 (2.9) -	75 (1.7)
Rumänien	484 (6.3) -	111 (3.8)	478 (6.0) -	105 (4.0)	486 (5.9) -	107 (3.4)
Polen	475 (2.6) -	71 (1.6)	480 (2.6) -	75 (1.6)	493 (2.4) -	79 (1.6)
Türkei	475 (5.4) -	105 (4.4)	469 (4.8) -	102 (3.7)	462 (4.5) -	98 (3.4)
Chile	455 (2.5) -	80 (1.4)	463 (2.5) -	82 (1.2)	469 (2.5) -	86 (1.3)

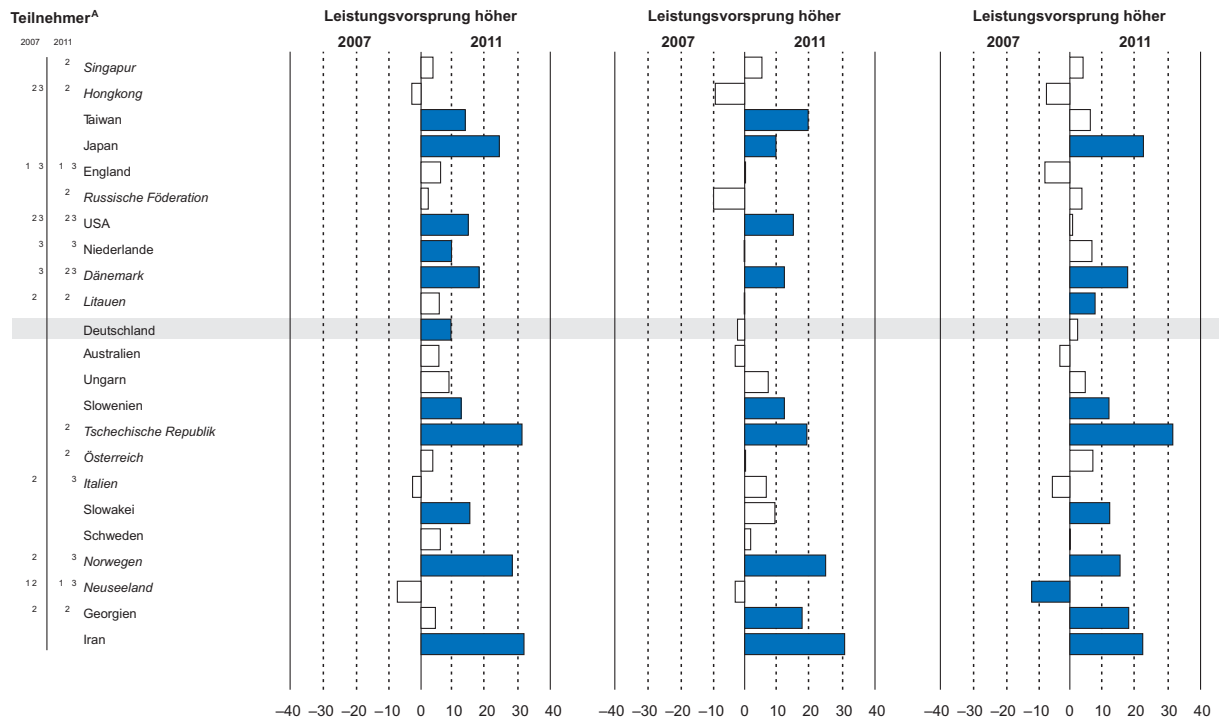


Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss

- In allen drei Bereichen signifikant positiv von den deutschen Mittelwerten abweichende Staaten ($p < .05$).
- ⊖ In allen drei Bereichen signifikant negativ von den deutschen Mittelwerten abweichende Staaten ($p < .05$).
- 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
- 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
- 3 = Die Teilnahmequoten auf Schul- und/oder Schulebene erreichen nicht die internationalen Vorgaben.
- + = Mittelwert statistisch signifikant höher als der deutsche Mittelwert ($p < .05$)
- = Mittelwert statistisch signifikant niedriger als der deutsche Mittelwert ($p < .05$)

Abbildung 3.11: Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 in den kognitiven Anforderungsbereichen *Reproduzieren, Anwenden und Problemlösen*

Teilnehmer ^A	Reproduzieren				Anwenden				Problemlösen									
	2007		2011		Veränderung ^B		2007		2011		Veränderung ^B		2007		2011		Veränderung ^B	
	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)
² Singapur	625 (4.3)	629 (3.5)	4 (5.5)	597 (4.1)	602 (3.4)	5 (5.4)	584 (4.1)	588 (3.7)	4 (5.5)									
²³ ² Hongkong	622 (3.7)	619 (3.2)	-3 (4.9)	606 (3.8)	597 (3.2)	-9 (5.0)	596 (3.8)	589 (3.4)	-7 (5.1)									
Taiwan	586 (1.9)	599 (2.1)	13 (2.8)	574 (1.9)	593 (2.0)	19 (2.8)	571 (2.0)	577 (2.5)	6 (3.2)									
Japan	567 (2.4)	590 (1.7)	24 (2.9)	570 (2.2)	579 (1.6)	9 (2.7)	569 (2.2)	592 (2.0)	22 (3.0)									
^{1 3} ^{1 3} England	546 (3.7)	552 (4.3)	6 (5.6)	542 (3.3)	542 (3.7)	0 (5.0)	539 (3.4)	531 (3.7)	-8 (5.0)									
² Russische Föderation	539 (4.9)	541 (3.4)	2 (5.9)	549 (5.3)	539 (3.9)	-9 (6.6)	544 (5.3)	548 (3.6)	4 (6.4)									
²³ ²³ USA	541 (2.8)	556 (2.1)	14 (3.5)	524 (2.8)	539 (2.1)	15 (3.5)	525 (2.4)	525 (2.2)	1 (3.2)									
³ ³ Niederlande	528 (2.4)	537 (2.0)	9 (3.1)	540 (2.2)	540 (1.6)	0 (2.7)	537 (2.5)	543 (2.6)	7 (3.6)									
³ ²³ Dänemark	514 (2.8)	531 (2.6)	18 (3.9)	527 (2.8)	539 (2.9)	12 (4.1)	525 (2.2)	543 (2.7)	17 (3.5)									
² ² Litauen	520 (2.8)	525 (3.0)	5 (4.1)	540 (2.7)	540 (2.5)	0 (3.6)	529 (2.8)	536 (2.5)	8 (3.7)									
Deutschland	515 (2.1)	524 (2.3)	9 (3.1)	530 (2.4)	528 (2.3)	-2 (3.3)	530 (2.9)	532 (3.0)	2 (4.2)									
Australien	511 (4.4)	516 (3.5)	5 (5.6)	522 (3.8)	519 (3.0)	-3 (4.8)	516 (3.7)	513 (2.6)	-3 (4.6)									
Ungarn	511 (3.6)	519 (3.8)	8 (5.2)	506 (3.8)	513 (3.3)	7 (5.0)	510 (4.2)	514 (3.7)	5 (5.6)									
Slowenien	498 (2.0)	510 (2.8)	12 (3.4)	502 (2.0)	514 (2.3)	12 (3.1)	504 (2.4)	516 (2.9)	12 (3.7)									
² Tschechische Republik	472 (2.5)	502 (2.4)	30 (3.5)	493 (2.9)	512 (2.8)	19 (4.0)	491 (3.7)	523 (2.7)	31 (4.6)									
² Österreich	504 (2.1)	507 (2.5)	4 (3.3)	505 (2.0)	506 (2.6)	0 (3.3)	506 (2.3)	513 (3.3)	7 (4.0)									
² ³ Italien	512 (3.5)	510 (2.7)	-3 (4.4)	499 (3.1)	506 (2.8)	7 (4.2)	511 (3.3)	505 (3.4)	-5 (4.8)									
Slowakei	491 (4.3)	506 (3.8)	15 (5.7)	496 (4.4)	505 (4.0)	9 (5.9)	499 (4.5)	511 (3.9)	12 (6.0)									
Schweden	483 (2.6)	489 (2.2)	6 (3.4)	506 (2.3)	507 (2.2)	2 (3.2)	519 (2.8)	520 (3.0)	0 (4.1)									
² ³ Norwegen	459 (3.0)	487 (3.1)	28 (4.3)	475 (2.9)	499 (3.0)	24 (4.2)	486 (2.9)	501 (3.3)	15 (4.4)									
¹² ^{1 3} Neuseeland	484 (2.7)	476 (3.2)	-7 (4.2)	493 (2.6)	490 (2.4)	-3 (3.5)	502 (2.8)	490 (2.5)	-12 (3.8)									
² ² Georgien	445 (4.2)	449 (3.7)	4 (5.7)	430 (4.7)	447 (3.4)	17 (5.8)	433 (4.6)	450 (3.5)	18 (5.8)									
Iran	404 (3.8)	435 (3.8)	31 (5.4)	397 (3.9)	427 (3.6)	30 (5.3)	401 (4.3)	423 (3.0)	22 (5.3)									



1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.
 B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Vergleich der Kompetenzen in den Anforderungsbereichen 2007 und 2011: In Abbildung 3.11 sind Unterschiede in Bezug auf die mathematischen Kompetenzen zwischen den Studienzyklen TIMSS 2007 und TIMSS 2011 getrennt nach den Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* im internationalen Vergleich dargestellt. Die Gestaltung der Abbildung sowie die Auswahl der Teilnehmerstaaten entsprechen der Abbildung 3.9.

Für Deutschland ist festzuhalten, dass sich für die Bereiche *Anwenden* und *Problemlösen* keine signifikanten Unterschiede zwischen den Leistungsmittelwerten für TIMSS 2007 und TIMSS 2011 ergeben. Dagegen zeigt sich für den Bereich *Reproduzieren* eine signifikant positive Differenz von 9 Punkten.

Die Abbildung zeigt zudem auch hier, dass für die überwiegende Zahl der dargestellten Teilnehmer positive Differenzen zu verzeichnen sind, dass die Muster der Veränderung jedoch nicht einheitlich sind. Während sich für sieben Teilnehmer signifikant positive Differenzen in allen drei Inhaltsbereichen ergeben, zeigen sich in einigen Fällen – wie auch in Deutschland – lediglich signifikante Veränderungen für einzelne Bereiche. Nur für Neuseeland lassen sich in einem Bereich signifikant schlechtere Leistungen für 2011 feststellen.

5.5 Einstellungen und Selbstkonzept

In TIMSS wurden im Anschluss an die Bearbeitung der Leistungstests die beteiligten Viertklässlerinnen und Viertklässler mittels Fragebögen zu ihrer Person, ihren Einstellungen und der Wahrnehmung des Unterrichts befragt. Mit diesen Fragebögen sollten der Kontext und die Hintergründe der Testleistungen erhoben werden.

In diesem Abschnitt werden Ergebnisse aus TIMSS 2011 zu Einstellungen und Selbstkonzept der Schülerinnen und Schüler in Deutschland dargestellt und mit den Ergebnissen aus TIMSS 2007 verglichen. Auf die Darstellung des internationalen Vergleichs der Ergebnisse wurde aus folgendem Grund verzichtet: Fachbezogene Selbstkonzepte entstehen unter anderem durch soziale Vergleichsprozesse (Köller, 2004). Schülerinnen und Schüler vergleichen ihre eigenen Fähigkeiten in einem Fach mit denen von Mitschülerinnen und Mitschülern sowie Peers. Sind die Fähigkeiten dieser Bezugsgruppe beispielsweise sehr hoch ausgeprägt, hat dies zur Folge, dass die einzelne Schülerin beziehungsweise der einzelne Schüler ihre beziehungsweise seine eigenen Fähigkeiten geringer einschätzt, als wenn die Bezugsgruppe weniger leistungsstark wäre.

Diese Auswirkungen der Bezugsgruppe auf das individuelle Selbstkonzept werden als Referenzgruppeneffekte bezeichnet und es wurde gezeigt, dass diese Effekte in ganz unterschiedlichen Kulturen auftreten (Seaton, Marsh & Craven, 2009). Referenzgruppeneffekte treten nicht nur bei fachbezogenen Selbstkonzepten, sondern auch bei affektiv-motivationalen Merkmalen wie Interessen auf. Es ist davon auszugehen, dass Referenzgruppeneffekte auch bei Einstellungen auftreten (Seaton, Marsh & Craven, 2010).

Die Ergebnisse eines internationalen Vergleichs von Einstellungen und Selbstkonzepten wären von solchen Referenzgruppeneffekten betroffen, da sich die Staaten im Niveau mathematischer Kompetenzen unterscheiden (vgl. Abschnitt 5.1). Das Niveau von Einstellungen und Selbstkonzepten wäre in Staaten mit besonders hohem Kompetenzniveau durch Referenzgruppeneffekte ‚nach unten gedrückt‘, in Staaten mit besonders niedrigem Kompetenzniveau hingegen angehoben. So ist etwa in Japan das mathematikbezogene Selbstkonzept

der Schülerinnen und Schüler im internationalen Vergleich sehr niedrig ausgeprägt, obwohl die Schülerinnen und Schüler in Japan hinsichtlich ihrer Leistungen in Mathematik im internationalen Vergleich sehr gut abschneiden. Vor diesem Hintergrund ist eine Interpretation des Niveaus von Einstellungen und Selbstkonzepten im internationalen Vergleich schwierig. In diesem Abschnitt werden daher nur die Ergebnisse für Deutschland in den Blick genommen und Unterschiede zwischen TIMSS 2011 und 2007 dargestellt.

Einstellung zur Mathematik: Die Einstellung zur Mathematik wird in TIMSS mit Hilfe eines Schülerfragebogens erfasst. Darin geben die Schülerinnen und Schüler an, ob sie den folgenden Aussagen über Mathematik *völlig zustimmen*, *eher zustimmen*, *eher nicht zustimmen* oder überhaupt nicht zustimmen.

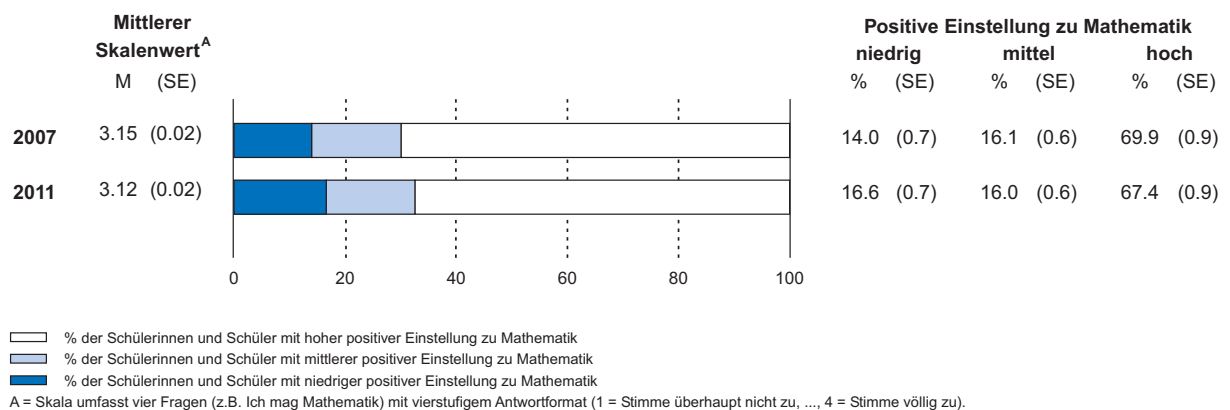
1. Ich lerne gern Mathematik.
2. Mathematik ist langweilig.
3. Ich mag Mathematik.

Die Antworten der Schülerinnen und Schüler werden bei Aussage 1 und 3 mit 1 („stimme überhaupt nicht zu“) bis 4 („stimme völlig zu“) Punkten kodiert, bei Aussage 2 in umgekehrter Reihenfolge. Für jedes Kind wird dann der Mittelwert über die drei Aussagen gebildet. Dieser Skalenwert drückt aus, wie stark positiv die Einstellung des Kindes zu Mathematik ausgeprägt ist. Die interne Konsistenz der Skala liegt bei einem Cronbachs Alpha von $\alpha = .86$.

In einem weiteren Schritt werden Gruppen von Schülerinnen und Schülern mit niedriger, mittlerer und hoher positiver Einstellung zur Mathematik gebildet. Dazu werden Kinder mit einem Mittelwert (M) von $1 \leq M \leq 2$ als ‚niedrig‘, Kinder mit einem Mittelwert von $2 < M < 3$ als ‚mittel‘ und Kinder mit einem Mittelwert von $3 \leq M \leq 4$ als ‚hoch‘ klassifiziert.

Abbildung 3.12 zeigt, wie stark positiv die Einstellungen der Schülerinnen und Schüler zur Mathematik bei TIMSS 2007 und 2011 in Deutschland ausgeprägt sind (mittlerer Skalenwert). Außerdem ist dargestellt, wie sich die Schülerinnen und Schüler auf die Gruppen der Kinder mit niedrigen, mittleren und hoch positiven Einstellungen verteilen.

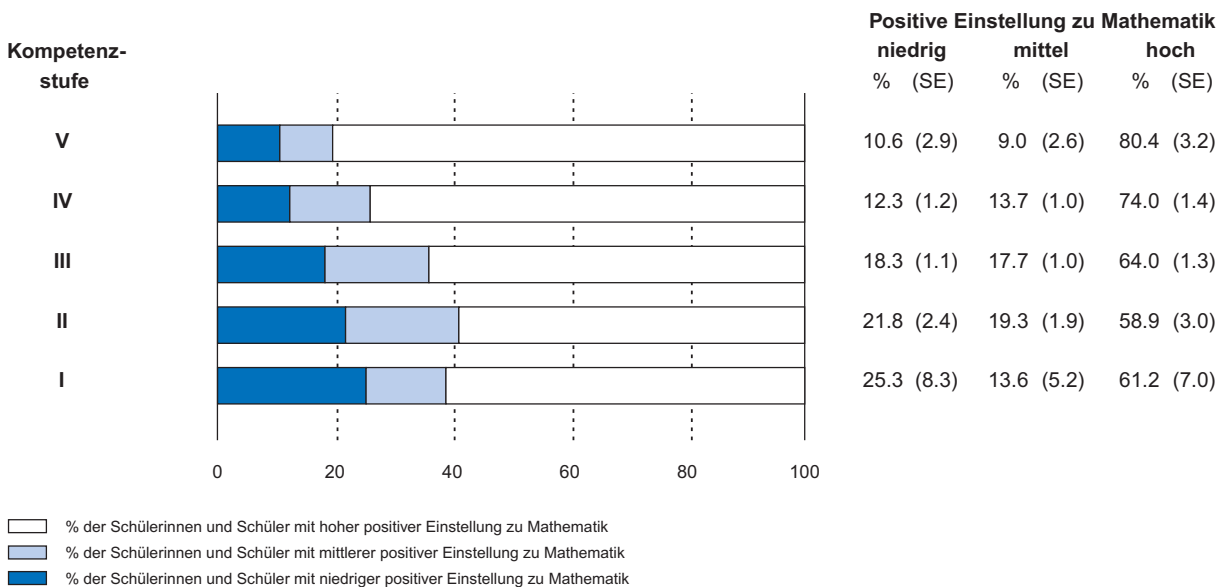
Abbildung 3.12: Mittlere positive Einstellung von Schülerinnen und Schülern zur Mathematik sowie prozentuale Verteilungen nach hoher, mittlerer und niedriger Einstellung – TIMSS 2007 und TIMSS 2011 im Vergleich



Vergleicht man die mittleren Skalenwerte von 2007 und 2011, so ergibt sich für 2011 ein leicht, jedoch nicht signifikant geringerer Wert von 3.12. Die Schülerinnen und Schüler stimmen den oben angegebenen Aussagen also eher zu und artikulieren wie 2007 eine durchaus positive Einstellung zum Fach Mathematik.

Vergleicht man die drei oben angegebenen Gruppen, so fällt auf, dass es gegenüber 2007 leichte Verschiebungen gegeben hat. Zwar bleibt der Anteil der *mittleren* Gruppe mit 16.0 Prozent stabil, der Anteil der Kinder mit *niedriger* positiver Einstellung steigt signifikant um 2.6 Prozentpunkte auf 16.6 Prozent und der Anteil der Schülerinnen und Schüler, die eine *hohe* positive Einstellung angeben, sinkt im Gegenzug entsprechend. Sie ist mit 67.4 Prozent nach wie vor hoch; zwei Drittel der Schülerinnen und Schüler der vierten Jahrgangsstufe in Deutschland geben an, dass sie eine positive Einstellung zur Mathematik haben.

Abbildung 3.13: Prozentuale Verteilungen und Testleistungen von Schülerinnen und Schülern in Mathematik auf den Kompetenzstufen nach positiver Einstellung



Wie erwartet zeigt sich ein positiver Zusammenhang zwischen positiver Einstellung zu Mathematik und der mathematischen Kompetenz der Kinder. Die Korrelation zwischen den mittleren Leistungswerten in Mathematik und der z-standardisierten Skala „Einstellung zur Mathematik“ beträgt in Deutschland $r = .14$ ($p < 0.001$).

In der Abbildung 3.13 wird dieser Zusammenhang illustriert, indem für jede der fünf Kompetenzstufen der jeweilige Anteil der Schülerinnen und Schüler der drei Gruppen angegeben wird (niedrige, mittlere und hohe positive Einstellung). Interessant ist dabei, dass es auf den Kompetenzstufen IV und V jeweils mehr als 10 Prozent Schülerinnen und Schüler gibt, die eine *geringe* positive Einstellung zur Mathematik aufweisen. Zum zweiten ist festzuhalten, dass rund 60 Prozent der Schülerinnen und Schüler auf den unteren beiden Kompetenzstufen eine bemerkenswert *hohe* positive Einstellung zur Mathematik artikulieren.

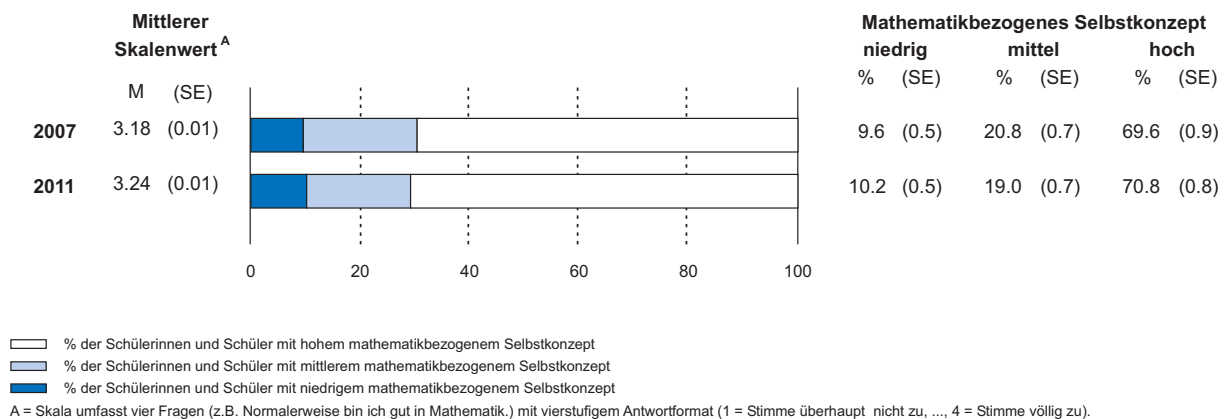
Mathematikbezogenes Selbstkonzept: Auch das mathematikbezogene Selbstkonzept wird in TIMSS mit Hilfe des Schülerfragebogens erfasst. Darin geben die Schülerinnen und Schüler an, ob sie den folgenden Aussagen *völlig zustimmen*, *eher zustimmen*, *eher nicht zustimmen* oder *überhaupt nicht zustimmen*.

1. Normalerweise bin ich gut in Mathematik.
2. Mathematik fällt mir schwerer als vielen meiner Mitschüler.
3. Ich bin einfach nicht gut in Mathematik.
4. Ich lerne schnell in Mathematik.

Die Antworten der Schülerinnen und Schüler werden bei Aussage 1 und 4 mit 1 („stimme überhaupt nicht zu“) bis 4 („stimme völlig zu“) Punkten kodiert, bei den Aussagen 2 und 3 in umgekehrter Reihenfolge. Für jedes Kind wird dann der Mittelwert über die vier Aussagen gebildet. Auch hier werden die Schülerinnen und Schüler in Gruppen mit niedrig, mittel und hoch positivem Selbstkonzept eingeteilt. Kinder mit einem Mittelwert (M) von $1 \leq M \leq 2$ werden als ‚niedrig‘, Kinder mit einem Mittelwert von $2 < M < 3$ als ‚mittel‘ und Kinder mit einem Mittelwert von $3 \leq M \leq 4$ als ‚hoch‘ klassifiziert. Die interne Konsistenz der Skala liegt bei einem Cronbachs Alpha von $\alpha = .85$.

Die Abbildung 3.14 zeigt, wie stark positiv das mathematikbezogene Selbstkonzept der Schülerinnen und Schüler bei TIMSS 2007 und 2011 in Deutschland ausgeprägt ist (mittlerer Skalenwert). Außerdem ist dargestellt, wie sich die Schülerinnen und Schüler auf die Gruppen der Kinder mit niedrigem, mittlerem und hohem positiven Selbstkonzept verteilen.

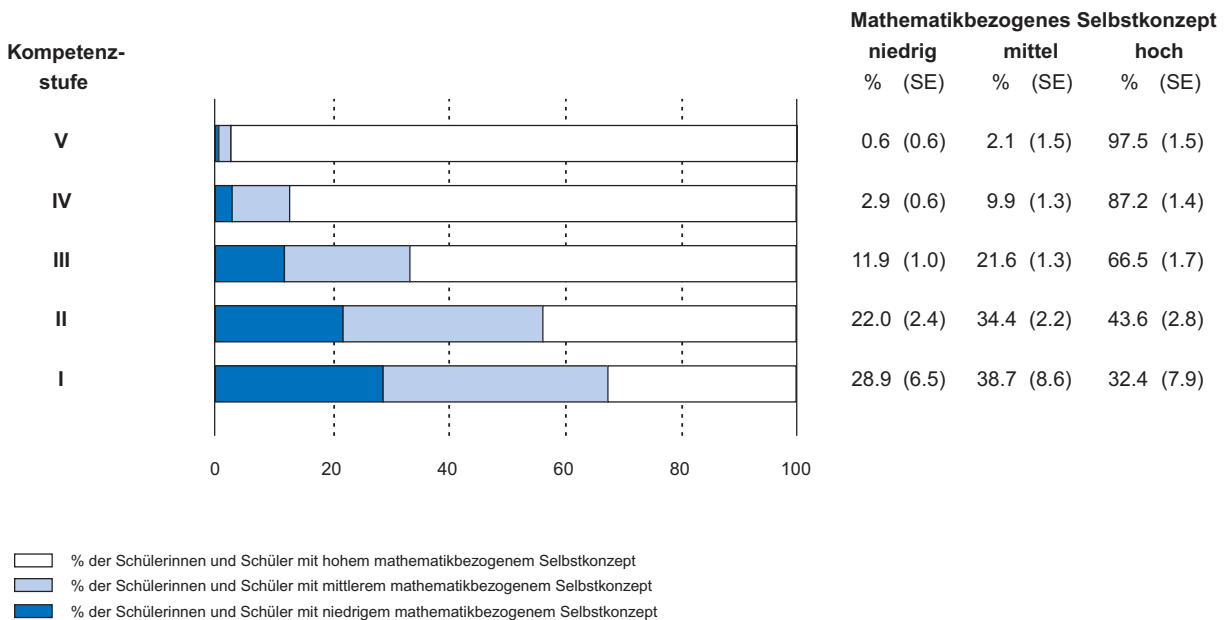
Abbildung 3.14: Mittleres Selbstkonzept von Schülerinnen und Schülern in Mathematik sowie prozentuale Verteilungen nach hohem, mittlerem und niedrigem Selbstkonzept – TIMSS 2007 und TIMSS 2011 im Vergleich



Vergleicht man die mittleren Skalenwerte von 2007 und 2011, so ergibt sich für 2011 ein signifikant höherer Wert von 3.24. Die Schülerinnen und Schüler stimmen den oben angegebenen Aussagen also eher zu und artikulieren wie 2007 ein recht positives mathematikbezogenes Selbstkonzept.

Vergleicht man die drei oben angegebenen Gruppen, so fällt auf, dass es gegenüber 2007 leichte Verschiebungen gegeben hat, die allesamt aber nicht signifikant sind. Der Anteil der Schülerinnen und Schüler mit *niedrigem* mathematischen Selbstkonzept bleibt mit 10.2 Prozent nahezu stabil. Dieser Anteil der Kinder mit *mittlerem* positiven Selbstkonzept fällt leicht auf 19.0 Prozent und der Anteil der Schülerinnen und Schüler, die ein *hohes* mathematikbezogenes Selbstkonzept angeben, steigt im Gegenzug auf 70.8 Prozent. Gut 7 von 10 Schülerinnen und Schülern der vierten Jahrgangsstufe in Deutschland geben an, dass sie über ein *hohes* positives mathematikbezogenes Selbstkonzept verfügen.

Abbildung 3.15: Prozentuale Verteilungen und Testleistungen von Schülerinnen und Schülern in Mathematik auf den Kompetenzstufen nach Selbstkonzept



Wie erwartet zeigt sich ein positiver Zusammenhang zwischen positivem mathematikbezogenen Selbstkonzept und der mathematischen Kompetenz der Kinder. Die Korrelation zwischen den mittleren Leistungswerten in Mathematik und der z-standardisierten Skala „Mathematikbezogenes Selbstkonzept“ beträgt in Deutschland $r = .42$ ($p < 0.001$).

In der Abbildung 3.15 wird dieser Zusammenhang illustriert, indem für jede der fünf Kompetenzstufen der jeweilige Anteil der Schülerinnen und Schüler der drei Gruppen niedrig, mittel und hoch angegeben wird. Der Anteil von Schülerinnen und Schülern mit *hohem* mathematischen Selbstkonzept fällt von 97.5 Prozent auf Kompetenzstufe V auf 32.4 Prozent auf Kompetenzstufe I. Der Anteil von Schülerinnen und Schülern mit *niedrigem* beziehungsweise *mittlerem* Selbstkonzept nimmt erwartungsgemäß von Kompetenzstufe V zu Kompetenzstufe I beständig zu.

6 Zusammenfassung

Die Zielsetzung von TIMSS 2011 besteht in einem internationalen Vergleich der mathematischen und naturwissenschaftlichen Leistungen von Schülerinnen und Schülern der vierten Jahrgangsstufe. In diesem Kapitel wurden die Hauptergebnisse für die Mathematik dargestellt, die sich wie folgt zusammenfassen lassen:

Leistungsmittelwerte: Im internationalen Vergleich erreichen die Schülerinnen und Schüler in Deutschland in Mathematik einen Leistungsmittelwert von 528 Punkten und befinden sich – wie 2007 – noch im oberen Drittel der Rangreihe der insgesamt 50 Teilnehmerstaaten. Damit liegt Deutschland signifikant über dem internationalen Mittelwert (491) sowie den Mittelwerten aller teilnehmenden EU-Staaten (519) sowie aller teilnehmenden OECD-Staaten (521). Deutschland befindet sich zusammen mit Litauen (534), Portugal (532) und Irland (527) in einer Staatengruppe, deren Leistungen sich nicht signifikant unterscheiden.

Ein erheblicher Abstand nach oben besteht zu den Teilnehmern an der Leistungsspitze aus Asien: Singapur (606), Republik Korea (Südkorea) (605), Hongkong (602), Taiwan (591) und Japan (585), zum Teil wohl erklärlich durch den hohen Stellenwert von Bildung, eine andere Lehr-/Lernkultur sowie den privaten Zusatzunterricht in diesen Staaten. Signifikant schlechter schneidet Deutschland auch im Vergleich mit einigen westeuropäischen Staaten, zum Beispiel Finnland (545), England (542), die Niederlande (540) oder Dänemark (537) ab. Trotz insgesamt zufriedenstellender Resultate deutet sich für Deutschland also erhebliches Entwicklungspotential an.

Da Deutschland im Jahr 2011 das zweite Mal an TIMSS-Grundschule teilgenommen hat, können erstmalig Unterschiede in der mathematischen Kompetenz von Schülerinnen und Schülern am Ende der Grundschulzeit angegeben werden. Dabei zeigt sich, dass das mittlere Niveau der mathematischen Kompetenz in Deutschland 2011 mit 528 Punkten im Großen und Ganzen dem Niveau aus 2007 entspricht, als 525 Punkte erreicht wurden. Bei TIMSS 2007 bildeten drei weitere Teilnehmerstaaten, nämlich Dänemark, Litauen und die USA die deutsche Vergleichsgruppe. Im Gegensatz zu Deutschland haben die USA und Dänemark signifikante Zuwächse zu verzeichnen. Dies zeigt, dass eine signifikante Leistungssteigerung durchaus erreichbar ist. Litauen erzielt in etwa denselben Zuwachs wie Deutschland. Insgesamt zeigt sich, dass von 23 Staaten, die an beiden Untersuchungen teilnahmen, neun Staaten signifikant besser abschneiden und sich kein Staat signifikant verschlechtert. 14 Staaten verbleiben in etwa auf dem Niveau von 2007. Zu dieser Gruppe gehört Deutschland.

Bedeutung der curricular nicht validen Aufgaben: In einer gesonderten Berechnung wurden die Leistungsmittelwerte ermittelt, nachdem die für die Schülerinnen und Schüler in Deutschland curricular nicht validen Aufgaben aus den Berechnungen eliminiert wurden. Hier ergibt sich für Deutschland zwar eine nominelle Verbesserung um vier Rangplätze. Allerdings lässt sich insgesamt zur Frage der curricularen Validität festhalten, dass das Abschneiden Deutschlands auf der TIMSS-Leistungsskala nicht auf das Testdesign zurückzuführen ist. Eine Modifikation des Testdesigns würde vermutlich nur zu einer geringfügigen Veränderung der Rangfolge führen, zumal unklar bleibt, ob solche Rangplatzverschiebungen sich auch für andere Teilnehmer ergeben hätten, würde man die jeweils für diesen Staat curricular nicht validen Aufgaben eliminieren.

Leistungsstreuung: Wie bereits in der Untersuchung von 2007 zeigt sich, dass die Leistungsstreuung in Deutschland relativ klein ist, was sowohl an der Größe der Standardabweichung (62) als auch an der Leistungsspanne zwischen den jeweils 5 Prozent schlechtesten und 5 Prozent besten Schülerinnen und Schülern ersichtlich ist. Es gibt nur wenige Staaten mit vergleichbarer ‚Leistungshomogenität‘.

Schaut man sich die Leistungswerte der 5 Prozent Kinder mit der schwächsten Mathematikleistung an, so gibt es lediglich fünf Staaten, in denen höhere Werte erzielt werden. Bei den 5 Prozent leistungsstärksten Schülerinnen und Schülern sind es zwölf Teilnehmerstaaten, die einen höheren oder gleichen Wert erreichten. Die 5 Prozent schwächsten Schülerinnen und Schüler in Deutschland weisen also vergleichsweise hohe Leistungswerte auf, die 5 Prozent stärksten Schüler vergleichsweise niedrige.

Kompetenzstufen und extreme Leistungsgruppen: Die inhaltliche Beurteilung der Mathematikleistung von Schülerinnen und Schülern am Ende des vierten Schuljahres, die über die Angabe von Leistungspunkten hinaus geht, erfolgt im Weiteren zunächst im Hinblick auf die erreichten Kompetenzstufen. Im Vergleich der Untersuchungen TIMSS 2007 und 2011 zeigt sich zunächst, dass sich für Deutschland trotz geringfügiger nomineller Verschiebungen keine signifikanten Veränderungen der Verteilung von Viertklässlerinnen und Viertklässlern auf die Kompetenzstufen ergeben.

Für TIMSS 2011 ist festzuhalten, dass 97 Prozent der Schülerinnen und Schüler in Deutschland mindestens Stufe II, 81 Prozent mindestens Stufe III, 37 Prozent mindestens Stufe IV und lediglich 5 Prozent die höchste Stufe erreichen.

Fast ein Fünftel der Schülerinnen und Schüler befindet sich also auf der ersten und zweiten Kompetenzstufe und verfügt damit allenfalls über elementares mathematisches Wissen sowie elementare mathematische Fertigkeiten und Fähigkeiten, um ihnen vertraute Aufgaben zu bearbeiten. Besorgniserregend sind insbesondere die 3 Prozent der Schülerinnen und Schüler, die sich auf der untersten Kompetenzstufe befinden und nur über rudimentäres mathematisches Wissen verfügen. Hier sind die Prozentsätze in den Teilnehmerstaaten an der Spitze der Leistungsskala kleiner; bis auf England (7%) sind bei den anderen Staaten, die bei den Leistungsmittelwerten signifikant besser abschneiden, vergleichbare Prozentsätze zu verzeichnen. Die Vergleichswerte VG_{EU} und VG_{OECD} liegen mit 7 Prozent ebenfalls über dem deutschen Wert.

Die knapp 75 Prozent der Grundschülerinnen und Grundschüler auf den beiden Kompetenzstufen III und IV wenden elementares mathematisches Wissen sowie elementare mathematische Fertigkeiten und Fähigkeiten in einfachen Situationen, aber auch bei der Lösung von mehrschrittigen Aufgaben mit inner- oder außermathematischem Kontextbezug an. Die 5 Prozent Schülerinnen und Schüler auf der höchsten Kompetenzstufe V nutzen ihre mathematischen Fertigkeiten und Fähigkeiten verständlich beim Lösen verhältnismäßig komplexer Probleme und können ihr Vorgehen erläutern.

Im Hinblick auf die Kompetenzstufe V zeigt sich aber auch, dass in den Staaten Litauen (10%), Irland (9%) und Portugal (8%), die mit Deutschland in der gleichen Leistungsgruppe liegen, der Anteil von Schülerinnen und Schülern auf dem höchsten Niveau fast doppelt so hoch ist. Auch die Vergleichswerte VG_{EU} und VG_{OECD} liegen mit 8 Prozent beziehungsweise 9 Prozent höher. Pro-

zentsätze von mehr als 30 Prozent, ja sogar 40 Prozent, wie sie Singapur (43%), Korea (39%) oder Hongkong (38%) erreichen, liegen in weiter Ferne.

Insgesamt ist festzuhalten, dass es nach wie vor an beiden Enden des Leistungsspektrums Handlungsbedarf gibt. Der Anteil von Schülerinnen und Schülern, die sich in Deutschland auf der untersten Kompetenzstufe I befinden, ist auch mit 3 Prozent noch bedauerlich groß, im internationalen Vergleich ist dies jedoch erfreulicherweise kein hoher Wert. Besonders besorgniserregend ist allerdings der Anteil von knapp 20 Prozent der Kinder auf den ersten beiden Kompetenzstufen, der nach PISA-Ergebnissen etwa auch dem Anteil von schwächeren Schülerinnen und Schülern am Ende der Sekundarstufe I entspricht. Sehr bedenklich ist zudem, dass in Deutschland nur ein vergleichsweise kleiner Anteil von 5 Prozent der Schülerinnen und Schüler die höchste Kompetenzstufe V erreicht.

Mathematische Inhaltsbereiche: In jedem der drei Inhaltsbereiche liegt der Leistungsmittelwert in Deutschland über dem internationalen Mittelwert. Es zeigt sich aber, dass die Leistung der Grundschülerinnen und Grundschüler in Deutschland im Bereich *Arithmetik* (520) unterhalb des Gesamtmittelwerts Mathematik für Deutschland (528) liegt. Im Ausgleich dazu liegt die Leistung in den Bereichen *Geometrie/Messen* (536) und *Umgang mit Daten* (546) oberhalb des Gesamtmittelwerts.

Im Vergleich der Ergebnisse von TIMSS 2007 und 2011 lässt sich konstatieren, dass der erreichte Leistungsmittelwert im Bereich *Arithmetik* mit 520 nicht signifikant von dem 2007 erreichten Wert (524) abweicht. In den beiden anderen Inhaltsbereichen *Geometrie/Messen* und *Umgang mit Daten* sind jedoch signifikante Verbesserungen von 9 beziehungsweise 14 Punkten zu verzeichnen.

Kognitive Anforderungsbereiche: In jedem der drei Kompetenzbereiche liegt der Leistungsmittelwert in Deutschland oberhalb des internationalen Mittelwerts. Die Schülerinnen und Schüler in Deutschland erzielen allerdings im Bereich *Reproduzieren* 524 Punkte, im Bereich *Anwenden* 528 Punkte und im Bereich *Problemlösen* 532 Punkte und liegen damit im erstgenannten Bereich unter und im letztgenannten oberhalb des Gesamtmittelwerts Mathematik.

Im Vergleich der Ergebnisse von TIMSS 2007 und 2011 lässt sich festhalten, dass der erreichte Leistungsmittelwert im Bereich *Reproduzieren* mit 524 Punkten über dem 2007 erreichten Wert (515) liegt. Die Unterschiede sind signifikant, im Gegensatz zu den nicht signifikanten Veränderungen im Bereich *Anwenden*, wo ein Minus von 2 Punkten zu berichten ist, und beim *Problemlösen*, wo ein Plus von 2 Punkten zu verzeichnen ist.

Einstellungen: Vergleicht man die mittleren Skalenwerte von 2007 und 2011, so ergibt sich für 2011 ein leicht, jedoch nicht signifikant geringerer Wert von 3.12 auf einer Skala von 1 bis 4. Die Schülerinnen und Schüler artikulieren wie 2007 eine durchaus positive Einstellung zum Fach Mathematik. Der Anteil der Kinder mit ‚niedriger‘ positiver Einstellung steigt signifikant auf 16.6 Prozent und der Anteil der Schülerinnen und Schüler, die eine hohe positive Einstellung angeben, sinkt im Gegenzug entsprechend. Sie ist mit 67.4 Prozent nach wie vor hoch; zwei Drittel der Schülerinnen und Schüler der vierten Jahrgangsstufe in Deutschland geben an, dass sie eine positive Einstellung zur Mathematik haben.

Bemerkenswert ist weiterhin, dass es auf den Kompetenzstufen IV und V jeweils gut 10 Prozent Schülerinnen und Schüler gibt, die eine geringe positive

Einstellung zur Mathematik aufweisen. Schließlich ist festzuhalten, dass rund 60 Prozent der Schülerinnen und Schüler auf den unteren beiden Kompetenzstufen eine positive mathematische Einstellung artikulieren.

Selbstkonzept: Vergleicht man die mittleren Skalenwerte von 2007 und 2011, so ergibt sich für 2011 ein signifikant höherer Wert von 3.24. Die Schülerinnen und Schüler stimmen den oben angegebenen Aussagen also eher zu und artikulieren wie 2007 ein recht positives mathematikbezogenes Selbstkonzept.

Der Anteil der Schülerinnen und Schüler mit niedrigem mathematischen Selbstkonzept bleibt mit 10.2 Prozent nahezu stabil. Der Anteil der Kinder mit mittlerem positiven Selbstkonzept fällt leicht auf 19.0 Prozent und der Anteil der Schülerinnen und Schüler, die ein hohes mathematikbezogenes Selbstkonzept angeben, steigt im Gegenzug auf 70.8 Prozent. Mehr als sieben von zehn Schülerinnen und Schülern der vierten Jahrgangsstufe in Deutschland geben an, dass sie über ein positives mathematikbezogenes Selbstkonzept verfügen.

Der Anteil von Schülerinnen und Schülern mit hohem mathematischen Selbstkonzept fällt von 97.5 Prozent auf Kompetenzstufe V auf 32.4 Prozent auf Kompetenzstufe I. Verfolgt man die Anteile (hoch) von der Kompetenzstufe V zur Kompetenzstufe I, so nimmt dieser Anteil kontinuierlich ab, während die Anteile der beiden anderen Gruppen (mittel und niedrig) beständig zunehmen, bis auf Kompetenzstufe I jeweils rund ein Drittel der Schülerinnen und Schüler den drei Gruppen zuzuordnen ist.

Resümierend bleibt Folgendes festzuhalten: Die Leistungen der Schülerinnen und Schüler in Deutschland am Ende von Klasse 4 sind nach wie vor recht gut. Sie haben sich aber im Vergleich zu TIMSS 2007 nicht wesentlich verbessert. Zudem sind in Bezug auf die Intensivierung der Förderung der leistungsschwachen Schülerinnen und Schüler (Kompetenzstufen I und II) und insbesondere auch der sehr leistungsstarken Schülerinnen und Schüler (Kompetenzstufe V) kaum Veränderungen zu verzeichnen. In der Gesamtschau der verschiedenen Fragestellungen, die untersucht wurden, erkennt man eine Reihe von Staaten, bei denen im Vergleich zwischen TIMSS 2007 und 2011 erfreuliche Entwicklungen zu konstatieren sind. Dass sich Teilnehmerstaaten in einzelnen Bereichen signifikant verschlechtern, ist nur selten zu beobachten. In der Gesamtschau verbleiben einige Staaten, bei denen sich im Vergleich von 2007 und 2011 keine oder nur wenige bedeutsame Unterschiede feststellen lassen. Zu dieser Gruppe gehört Deutschland.

Literatur

- Baumert, J., Rainer, L., 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., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.). (2008). *TIMSS 2007: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Frey, A., Heinze, A., Mildner, D., Hochweber, J. & Asseburg, R. (2010). Mathematische Kompetenz von PISA 2003 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. 153–176). Münster: Waxmann.
- 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.
- 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.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005). *Bildungsstandards im Fach Mathematik für den Primarbereich (Jahrgangsstufe 4). Beschluss vom 15.10.2004*. München: Wolters Kluwer.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2009). *Kompetenzstufenmodell zu den Bildungsstandards im Fach Mathematik für den Primarbereich (Jahrgangsstufe 4)*. Zugriff am 27.08.2012 unter http://www.iqb.hu-berlin.de/bista?reg=r_4.
- Köller, O. (2004). *Konsequenzen von Leistungsgruppierungen*. Münster: Waxmann.
- Köller, O., Knigge, M. & Tesch, B. (Hrsg.). (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.
- Martin, M. O. & Mullis, I. V. S. (2011). *TIMSS and PIRLS methods and procedures*. Zugriff am 18.10.2012 unter <http://timssandpirls.bc.edu/methods/index.html>.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2008). *Richtlinien und Lehrpläne für die Grundschule in Nordrhein-Westfalen*. Frechen: Ritterbach.
- Mullis, I. V. S., Martin, M. O., Foy, P. & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y. & Preuschoff, C. (2009). *TIMSS 2011 Assessment Framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- 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* (S. 120–141). Weinheim: Beltz.
- 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.
- Seaton, M., Marsh, H. W. & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the Big-Fish-Little-Pond effect (BFLPE) across 41 culturally and economically diverse countries. *Journal of Educational Psychology*, 101 (2), 403-419.
- Seaton, M., Marsh, H. W. & Craven, R. G. (2010). Big-Fish-Little-Pond Effect: Generalizability and moderation – Two sides of the same coin. *American Educational Research Journal*, 47 (2), 390-433.
- 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.
- Walther, G., Geiser, H., Langeheine, R. & Lobemeier, K. (2003). Mathematische Kompetenzen am Ende der vierten Jahrgangsstufe. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 189–226). Münster: Waxmann.
- Walther, G., Selter, C. & Neubrand, J. (2007). Die Bildungsstandards Mathematik. In G. Walther, M. van den Heuvel-Panhuizen, D. Granzer & O. Köller (Hrsg.), *Bildungsstandards für die Grundschule. Mathematik konkret* (S. 16–41). Berlin: Cornelsen Scriptor.
- Walther, G., Selter, C., Bensen, M. & Bos, W. (2008). Mathematische Kompetenz im internationalen Vergleich. Testkonzeption und Ergebnisse. In W. Bos, M. Bensen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische*

und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich (S. 49–85). Münster: Waxmann.

Winter, H. (1975). Allgemeine Lernziele im Mathematikunterricht? *Zentralblatt für Didaktik der Mathematik*, 7 (4), 106-116.

Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 61, 37-46.

Kapitel IV

Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse

Thilo Kleickmann, Theresia Brehl, Steffani Saß, Manfred Prenzel und Olaf Köller

1 Einleitung

Es besteht international ein breiter Konsens darüber, dass eine naturwissenschaftliche Grundbildung für *alle* Bürgerinnen und Bürger in modernen Gesellschaften unabdingbar ist und daher bereits in der Schule angestrebt werden sollte (American Association for the Advancement of Science, 1993; Bybee, 2002; Prenzel, Schöps et al., 2007; Roberts, 2007). Naturwissenschaftliche Erkenntnisse und Denkweisen finden Anwendung in nahezu allen Bereichen des alltäglichen Lebens, von Fragen der Gesundheit bis hin zu Fragen der globalen Erhaltung der Umwelt. Naturwissenschaften helfen dem Einzelnen, sich mit alltäglichen Fragen und Phänomenen verständlich auseinanderzusetzen. Aber auch um an der öffentlichen Diskussion über Themen, die Bezüge zu Naturwissenschaften aufweisen, verantwortungsvoll teilnehmen zu können, sollten Schülerinnen und Schüler ein grundlegendes Verständnis naturwissenschaftlicher Konzepte und Verfahrensweisen entwickeln. Ein naturwissenschaftliches Grundverständnis bildet außerdem das Fundament für späteres Lernen in der beruflichen Ausbildung. Für einen hochqualifizierten Nachwuchs in naturwissenschaftlich-technischen und verwandten Berufen stellt ein hohes naturwissenschaftliches Kompetenzniveau in der Breite eine zentrale Voraussetzung dar (Prenzel, Schöps et al., 2007). Neben naturwissenschaftlichen Kompetenzen sind auch motivationale Orientierungen wie positive Einstellungen und positive Fähigkeitsselbsteinschätzungen der Schülerinnen und Schüler ein weiterer bedeutsamer Zielbereich naturwissenschaftlichen Unterrichts. Sie werden als wichtige Voraussetzung für die Bereitschaft zur lebensspannenübergreifenden Auseinandersetzung mit Naturwissenschaften angesehen (Koballa & Glynn, 2007).

In Deutschland findet die erste schulische Auseinandersetzung mit den Naturwissenschaften im Rahmen des Faches Sachunterricht in der Grundschule statt.¹ Aufgabe des Faches Sachunterricht ist es, „Schülerinnen und Schüler darin zu unterstützen, sich die natürliche, soziale und technisch gestaltete Umwelt bildungswirksam zu erschließen und dabei auch Grundlagen für den Fachunterricht an weiterführenden Schulen zu legen“ (GDSU – Gesellschaft für Didaktik des Sachunterrichts, 2002). Bei dem Fach Sachunterricht handelt es sich somit um ein integratives Fach, in dem neben naturwissenschaftlichen auch sozialwissenschaftliche und technische Bezüge hergestellt werden. Eine Gliederung in Fächer wie Biologie, Chemie, Geographie, Geschichte, Physik und Sozialwissenschaften, wie sie in der Sekundarstufe etabliert ist, ist in der Grundschule nicht vorgesehen.

Vor dem Hintergrund der eingangs skizzierten Bedeutung naturwissenschaftlicher Bildung ist die Frage, über welche Kompetenzen Schülerinnen und Schüler in diesem Bereich verfügen, Gegenstand zahlreicher Untersuchungen gewesen. Deutschland beteiligt sich seit der *Third International Mathematics and Science Study* (TIMSS) aus dem Jahr 1995 und dem *Programme for International Student Assessment* (PISA) aus dem Jahr 2000 regelmäßig an internationalen Schulleistungsstudien im Bereich der Naturwissenschaften. Die TIMS-Studie aus dem Jahr 1995, an der Schülerinnen und Schüler der Jahrgangsstufen 7 und 8 teilnahmen, und die PISA-Studien aus den Jahren 2000, 2003, 2006 und 2009, bei denen 15-jährige Schülerinnen und Schüler untersucht wurden, lieferten Erkenntnisse über naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern der Sekundarstufe I im internationalen Vergleich (Klieme et al., 2010). Hier zeichnete sich im Zeitraum von 2000 bis 2009 ein positiver Trend im Bereich Naturwissenschaften ab: Lagen die Leistungen der Schülerinnen und Schüler im Jahr 2000 noch unter dem Durchschnitt der OECD-Staaten, lagen sie 2006 und 2009 signifikant darüber (Rönnebeck, Schöps, Prenzel, Mildner & Hochweber, 2010).

Mit einer nationalen Erweiterungsstudie zur *Internationalen Grundschul-Lese-Untersuchung* (IGLU-E) aus dem Jahr 2001 und der TIMS-Studie aus dem Jahr 2007 liegen ebenfalls Erkenntnisse zu den naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern am Ende der 4. Jahrgangsstufe vor. Während IGLU-E 2001 im Bereich der naturwissenschaftlichen Kompetenzen noch einigen methodischen Einschränkungen unterlag (u.a. nahmen vier Bundesländer nicht teil), lagen mit der TIMS-Studie 2007 erstmals belastbare, repräsentative Erkenntnisse vor (vgl. Wittwer, Saß & Prenzel, 2008).

Hier zeigte sich, dass die Kompetenz von Grundschülerinnen und -schülern in den Naturwissenschaften in Deutschland mit 528 Punkten im oberen Leistungsdrittel aller Teilnehmerstaaten und signifikant über dem Durchschnitt der teilnehmenden Staaten lag. Allerdings wiesen die Schülerinnen und Schüler in Deutschland einen erheblichen Abstand in der mittleren Leistung zu den Staaten mit den höchsten Leistungsmittelwerten in den Naturwissenschaften auf. Etwa ein Viertel aller Grundschülerinnen und -schüler in Deutschland wurde der ersten und zweiten Kompetenzstufe zugeordnet. Das bedeutet, dass diese Schülerinnen und Schüler allenfalls über elementares Wissen über naturwissenschaftliche Sachverhalte verfügen und Schwierigkeiten haben, dieses Wissen auch anzuwenden (Wittwer et al., 2008).

1 In 5 der 16 Länder in Deutschland wird das Fach anders bezeichnet (z.B. *Heimat- und Sachunterricht* oder *Mensch, Natur und Kultur*). Der Einfachheit halber wird in diesem Kapitel von dem Fach Sachunterricht gesprochen.

Im Vergleich zu den anderen teilnehmenden OECD- und EU-Staaten war Deutschland derjenige Staat mit den größten Unterschieden zwischen Jungen und Mädchen in der naturwissenschaftlichen Kompetenz ($M_M = 520$ gegenüber $M_J = 535$). Die Ergebnisse zeigten zudem deutliche soziale Disparitäten in den Kompetenzen. Der Zusammenhang zwischen dem Besitz an Kulturgütern und der erreichten naturwissenschaftlichen Kompetenz war in keinem der Teilnehmerstaaten so stark wie in Deutschland. Auch waren Unterschiede zwischen Grundschulkindern mit und ohne Migrationshintergrund nur in Österreich noch stärker ausgeprägt als in Deutschland (Bonsen, Frey & Bos, 2008; Bonsen, Kummer & Bos, 2008; Wittwer et al., 2008).

Mit TIMSS werden seit 2003 (in Deutschland seit 2007) Leistungen von Kindern am Ende der 4. Jahrgangsstufe in einem Turnus von vier Jahren untersucht. Dadurch können wie bei PISA perspektivisch auch Trends in den naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler untersucht werden. Vor diesem Hintergrund wurde auch die Bedeutung des Akronyms TIMSS geändert: TIMSS steht jetzt für *Trends in Mathematics and Science Study*.

In diesem Kapitel werden die Ergebnisse der deutschen Beteiligung an TIMSS 2011 für den Bereich der Naturwissenschaften dargestellt. Hierzu wird in Abschnitt 2 zunächst näher auf naturwissenschaftliche Grundbildung im Rahmen des Faches Sachunterricht in Deutschland eingegangen. In den Abschnitten 3 und 4 werden die TIMSS-Rahmenkonzeption und der Test zur Erfassung naturwissenschaftlicher Kompetenzen beschrieben, bevor in Abschnitt 5 die Ergebnisse entlang folgender Fragen dargestellt werden:

- Welches naturwissenschaftliche Kompetenzniveau zeigen Schülerinnen und Schüler in Deutschland am Ende der 4. Jahrgangsstufe im internationalen Vergleich? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und TIMSS 2011? (Abschnitt 5.1)
- Wie lassen sich die TIMSS-Leistungskennwerte auf Kompetenzstufen einordnen? Wie groß sind die Gruppen der auffällig leistungsschwachen und leistungsstarken Kinder? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und TIMSS 2011? (Abschnitt 5.2)
- Welche Ergebnisse erzielen die deutschen Viertklässler in den drei naturwissenschaftlichen Inhaltsbereichen (*Biologie, Physik/Chemie* und *Geographie*)? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und TIMSS 2011? (Abschnitt 5.3)
- Welche Ergebnisse erzielen die deutschen Viertklässler in den drei kognitiven Anforderungsbereichen *Reproduzieren, Anwenden* und *Problemlösen*? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und TIMSS 2011? (Abschnitt 5.4)
- Wie sind Einstellungen zu Naturwissenschaften und naturwissenschaftsbezogene Selbstkonzepte der Schülerinnen und Schüler in Deutschland ausgeprägt? Wie unterscheiden sich die Ausprägungen im Jahre 2011 von den Ergebnissen in 2007? Wie stark hängen Einstellungen zu Naturwissenschaften und das naturwissenschaftsbezogene Selbstkonzept der Schülerinnen und Schüler mit der naturwissenschaftlichen Kompetenz zusammen? (Abschnitt 5.5)

Auf Ergebnisse aus TIMSS 2011 zu Unterschieden in der naturwissenschaftlichen Kompetenz in Abhängigkeit von Geschlecht, sozialer Herkunft und Migrationshintergrund wird in anderen Kapiteln dieses Bandes (Kapitel 6, 7 und 8) eingegangen.

2 Naturwissenschaftliche Grundbildung in Deutschland

Angestoßen unter anderem durch das unerwartet schlechte Abschneiden von Schülerinnen und Schülern bei TIMSS 1995 und PISA 2000 wurde seit Ende der 1990er Jahre die Notwendigkeit einer stärkeren frühen naturwissenschaftlichen Förderung in Deutschland diskutiert. Bereits in den 1970er Jahren gab es eine Phase der Wissenschaftsorientierung in Deutschland, die aber nicht zu einer dauerhaften Implementierung naturwissenschaftlicher Inhalte in der Grundschule führte (Möller, Kleickmann & Sodian, 2011). Es folgte eine Phase, in der naturwissenschaftliche Themen, insbesondere physikalische und chemische Themen, im Sachunterricht kaum eine Rolle spielten, wie durch Lehrplan- und Schulbuchanalysen belegt werden konnte (Blaseio, 2004; Einsiedler, 1998; Strunck, Lück & Demuth, 1998). Die starke Vernachlässigung naturwissenschaftlicher Themen im Sachunterricht der Grundschule in den 1990er Jahren legte die Vermutung nahe, dass die in IGLU-E 2001 erfassten naturwissenschaftlichen Kompetenzen der Grundschul Kinder zu einem beträchtlichen Anteil außerhalb der Schule erworben wurden (Prenzel, Geiser, Langeheine & Lobemeier, 2003).

Zur Wiederbelebung der Diskussion um eine frühe naturwissenschaftliche Förderung gegen Ende der 1990er Jahre trugen neben den oben genannten Schulleistungsstudien auch ein offenkundiger Nachwuchsmangel in naturwissenschaftlich-technischen Berufen und neuere Erkenntnisse aus der Entwicklungspsychologie bei, die die Förderbarkeit anspruchsvoller Fähigkeiten im Grundschulalter zeigten (Einsiedler, Martschinke & Kammermeyer, 2008; Möller et al., 2011). Die Diskussion nahm neben dem Primar- auch den Elementarbereich in den Blick.

Für den Elementarbereich schrieb die *Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland* (KMK) 2004 im *Gemeinsamen Rahmen der Länder für die frühe Bildung in Kindertageseinrichtungen* den Bereich *Mathematik, Naturwissenschaft, (Informations-)Technik* als Bildungsbereich für Kindertagesstätten fest (KMK, 2004). Naturwissenschaften sind daher mittlerweile in den Bildungsplänen der Länder für den Elementarbereich verankert. Die KMK empfiehlt jedoch, die pädagogische Arbeit solle sich in erster Linie an den Interessen, Bedürfnissen und Lebenssituationen der einzelnen Kinder ausrichten (KMK, 2004). Konkrete naturwissenschaftliche Inhalte werden in den Bildungsplänen für den Elementarbereich dementsprechend in der Regel nicht festgelegt (eine Ausnahme bildet z.B. der Bildungsplan des Landes Bayern; Bayerisches Staatsministerium für Arbeit und Sozialordnung, Familie und Frauen/Staatsinstitut für Frühpädagogik, 2006).

Im Primarbereich wurden in den letzten zehn Jahren nahezu alle Lehrpläne zum Sachunterricht erneuert. Naturwissenschaftliche Themen wurden dabei wieder stärker berücksichtigt. Eine neuere Schulbuchanalyse zeigt dementsprechend einen Anstieg naturwissenschaftlicher Inhalte im Sachunterricht seit der Jahrtausendwende (Blaseio, 2009a). Länderübergreifende Bildungsstandards, wie sie für die Fächer Mathematik und Deutsch in der Grundschule entwickelt wurden, gibt es für das Fach Sachunterricht und damit für das naturwissenschaftliche Lernen in der Grundschule nicht. Einen orientierenden Einfluss auf die Lehrplanrevisionen hatte jedoch der *Perspektivrahmen Sachunterricht* der Gesellschaft für Didaktik des Sachunterrichts (GDSU, 2002). Um die Anschlussfähigkeit an die Fächer der Sekundarstufe zu sichern, unterscheidet der Perspektivrahmen fünf fachlich orientierte Perspektiven, die im Sachunterricht erschlossen werden

sollen: eine naturwissenschaftliche, eine sozialwissenschaftliche, eine geographische, eine technische und eine historische Perspektive.

Für die naturwissenschaftliche Perspektive benennt der Perspektivrahmen Kompetenzen, die im Verlaufe der Grundschulzeit entwickelt werden sollen (vgl. GDSU, 2002):

- Naturphänomene sachorientiert wahrnehmen, beobachten, benennen und beschreiben.
- Naturphänomene auf physikalische, chemische und biologische Gesetzmäßigkeiten zurückführen.
- Zwischen Erscheinungen der belebten und der unbelebten Natur unterscheiden können.
- Fragehaltungen aufbauen, Probleme identifizieren und Verfahren der Problemlösung anwenden (z.B. Experimente planen und auswerten).
- Regelmäßigkeiten der unbelebten Natur als Bedingungen für die Existenz der belebten Natur verstehen.
- Gründe für einen verantwortlichen Umgang mit der Natur erfassen.

Es wird deutlich, dass hier inhaltsbezogene Kompetenzen (z.B. Naturphänomene auf physikalische Prinzipien zurückführen) und verfahrensbezogene Kompetenzen (z.B. beobachten oder experimentieren) berücksichtigt wurden. Für jede der Perspektiven beschreibt der Perspektivrahmen einen Kern an Inhalten und Verfahren sowie Beispiele zur Vernetzung der Perspektiven im Unterricht. Eine Zuordnung der Themengebiete, die die TIMSS-Rahmenkonzeption zur Erfassung naturwissenschaftlicher Kompetenzen vorsieht, zu den Perspektiven des Perspektivrahmens wird im Abschnitt 3 vorgenommen.

Eine große Herausforderung im Zusammenhang mit früher naturwissenschaftlicher Bildung in Deutschland besteht in der Abstimmung von naturwissenschaftlichen Inhalten im Elementar-, Primar- und auch Sekundarbereich (Blaseio, 2009b; Möller, 2009). Dieser Abstimmungsbedarf ist durch die Institutionalisierung des naturwissenschaftlichen Lernens im Elementarbereich noch verschärft worden.

Ein Problem für die Implementierung einer naturwissenschaftlichen Grundbildung im Primarbereich stellt in Deutschland wie in anderen Ländern traditionell die Ausbildung von Grundschullehrkräften und auch das oft niedrige Kompetenzzempfinden der Lehrkräfte im Bereich Naturwissenschaften dar (Appleton, 2007). Solche Kompetenzeinschätzungen führen dazu, dass Naturwissenschaften (soweit möglich) schon in der Ausbildung, aber auch später im Beruf als Grundschullehrkraft eher gemieden werden (Appleton, 2007). Zudem werden Grundschullehrkräfte im Vergleich zu Lehrkräften der Sekundarstufe bei geringerem Gesamtstudienumfang als „fachliche Generalisten“ ausgebildet. Entsprechend gering ist häufig der Umfang, der für die Ausbildung für das naturwissenschaftliche Lernen zur Verfügung steht. Lediglich Nordrhein-Westfalen hat mittlerweile den Studienumfang von Grundschullehrkräften dem der Lehrkräfte der Sekundarstufe angeglichen (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2009).

Neben den Bemühungen um Reformen der naturwissenschaftlichen Grundbildung im Rahmen formaler Bildungsangebote in Schulen und Kindertagesstätten gibt es in Deutschland in den letzten circa zehn Jahren auch zahlreiche Aktivitäten, die auf die Förderung informellen und non-formalen naturwissenschaftlichen Lernens außerhalb dieser Institutionen abzielen. Dazu gehören beispielsweise die Einrichtung von Laboratorien für biologisches, chemisches,

physikalisches und technisches Lernen an Universitäten, die Einrichtung von „Science Centern“ und „Kinderakademien“ sowie eine verstärkte Präsenz von Wissenschaftssendungen für Kinder im Fernsehen (Einsiedler et al., 2008).

3 TIMSS-Rahmenkonzeption zur Erfassung naturwissenschaftlicher Kompetenzen

Die Erfassung naturwissenschaftlicher Kompetenzen in TIMSS basiert auf einer Konzeption, die in Zusammenarbeit mit den teilnehmenden Staaten entwickelt wurde. Diese Rahmenkonzeption soll drei Dinge gewährleisten: Erstens soll sichergestellt werden, dass sich große Schnittmengen der Inhaltsbereiche, die in der Rahmenkonzeption vorgesehen sind, mit den Lehrplänen der einzelnen Staaten ergeben. Zweitens sollen naturwissenschaftliche Kompetenzen möglichst breit abgedeckt werden. Drittens soll die Rahmenkonzeption die Analyse von Teilkompetenzen ermöglichen, um auf diese Weise ein differenziertes Bild der naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler zeichnen zu können (Mullis, Martin, Ruddock, O’Sullivan & Preuschoff, 2009). Die Ansprüche der breiten Abdeckung naturwissenschaftlicher Kompetenzen und der Übereinstimmung mit den Lehrplänen in den Teilnehmerstaaten mussten bei der Entwicklung der TIMSS-Rahmenkonzeption notwendigerweise ausbalanciert werden. Dem Anspruch nach differenzierter Darstellung von Teilkompetenzen wurde Rechnung getragen, indem drei Inhaltsbereiche und drei kognitive Anforderungsbereiche unterschieden werden.

Die Inhaltsbereiche umfassen *Biologie*, *Physik/Chemie* und *Geographie*. Die kognitiven Anforderungsbereiche beziehen sich auf grundlegende Denkprozesse, die erforderlich sind, um in einem naturwissenschaftlichen Inhaltsbereich kompetent zu agieren. Sie berücksichtigen das *Reproduzieren* von Wissen, das *Anwenden* von Wissen sowie das Schlussfolgern und *Problemlösen*. In der Rahmenkonzeption von TIMSS werden die drei Inhaltsbereiche durch Themenbereiche weiter ausdifferenziert und die kognitiven Anforderungsbereiche werden durch einzelne kognitive Aktivitäten weiter konkretisiert.

3.1 Naturwissenschaftliche Inhaltsbereiche

Tabelle 4.1 zeigt, welche Themengebiete in der TIMSS-Rahmenkonzeption für das 4. Schuljahr den drei Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geographie* zugeordnet sind.

Im Bereich der *Biologie* werden Eigenschaften der belebten Natur, Fortpflanzungs- und Wachstumsprozesse sowie das Zusammenleben von Organismen in verschiedenen Lebensräumen thematisiert. Außerdem sind Voraussetzungen für eine gesunde Lebensführung und umweltbewusste Verhaltensweisen wichtige Gebiete.

Im Bereich der *Physik/Chemie* liegt ein Schwerpunkt auf der Klassifizierung von Gegenständen und Stoffen nach physikalischen Eigenschaften. Daneben werden Energiequellen, Temperatur und Wärmeleitung thematisiert. Licht, elektrischer Strom, Magnetismus und Kräfte sind weitere wichtige Gebiete. Des Weiteren enthält dieser Bereich auch Inhalte aus der Chemie, zum Beispiel Veränderungen bei Stoffen oder das Herstellen und Trennen von Gemischen.

Im Bereich *Geographie* geht es um die Struktur der Erde, ihre Ressourcen und um zyklische Prozesse. Außerdem wird die Rolle der Erde im Sonnensystem thematisiert.

Bezieht man die in TIMSS vorgesehenen drei Inhaltsbereiche auf die im Perspektivrahmen Sachunterricht dargestellten fünf fachlichen Perspektiven, so wird deutlich, dass die in der TIMSS-Rahmenkonzeption vorgesehenen Anforderungen über die naturwissenschaftliche Anforderungen des Perspektivrahmens hinausgehen. Die TIMSS-Rahmenkonzeption umfasst auch Teile der geographischen und der historischen Perspektive (z.B. Erdgeschichte), in der es im Kern um Zeit und Wandel geht (GDSU, 2002).

Tabelle 4.1: Themengebiete der einzelnen naturwissenschaftlichen Inhaltsbereiche

Biologie	Physik/Chemie	Geographie
<ul style="list-style-type: none"> – Körperstrukturen und deren Funktionen bei Menschen und anderen Lebewesen (Pflanzen und Tiere) – Kennzeichen des Lebendigen; Lebenszyklen und Fortpflanzung bei Pflanzen und Tieren (z.B. Vererbung allgemeiner Eigenschaften, Lebenszyklen bekannter Organismen) – Physische Eigenschaften, Verhalten und Überleben von Organismen in verschiedenen Lebensräumen – Biologische Zusammenhänge in Lebensräumen (z.B. einfache Nahrungsketten, Räuber-Beute-Beziehungen) – Körper und Gesundheit (z.B. Übertragung ansteckender Krankheiten und die Möglichkeit, sich vor Ansteckung zu schützen; Anzeichen für Krankheit; Ernährungsweise, körperliche Betätigung) 	<ul style="list-style-type: none"> – Aggregatzustände (Festkörper, Flüssigkeiten, Gase) und Unterschiede in ihren physikalischen Eigenschaften (Form, Volumen) einschließlich Zustandsänderungen bei Erhitzen und Kühlen – Klassifizierung von Gegenständen/Stoffen nach ihren physikalischen Eigenschaften (z.B. Gewicht/Masse, Volumen, magnetische Anziehung) – Gemische herstellen und trennen – Bekannte Veränderungen bei Stoffen (z.B. Verwesung, Verbrennen, Rosten, Kochen) – Gebräuchliche Energiequellen/ Formen der Energie und deren praktische Anwendung (z.B. Sonne, Elektrizität, Wasser, Wind) – Licht (z.B. Lichtquellen, Verhalten von Licht) – Elektrische Schaltkreise und Eigenschaften von Magneten – Kräfte, die Objekte in Bewegung setzen (z.B. Erdanziehung, Anziehungs- und Abstoßungskräfte) 	<ul style="list-style-type: none"> – Wasser auf der Erde (Vorkommen, Arten, stehende/fließende Gewässer) und Luft (Zusammensetzung, Nachweis des Vorhandenseins, Gebrauch) – Charakteristische Landschaftsmerkmale (z.B. Gebirge, Ebenen, Flüsse, Wüsten) und das Beziehungs- und Wirkungsgefüge von Mensch und Landschaft (z.B. Landwirtschaft, Bewässerung, Landgewinnung) – Wetterbedingungen in der Betrachtung von Tag zu Tag oder über die Jahreszeiten – Versteinerungen von Tieren und Pflanzen (Alter, Fundort, Entstehung) – Unser Sonnensystem (Planeten, Sonne, Mond) – Tag, Nacht und Schatten infolge der Rotation der Erde und ihrer Beziehung zur Sonne

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

3.2 Kognitive Anforderungsbereiche

Tabelle 4.2 gibt Aufschluss darüber, welche Denkprozesse den drei kognitiven Anforderungsbereichen in der TIMSS-Rahmenkonzeption zur Erfassung naturwissenschaftlicher Kompetenzen zugeordnet sind.

Tabelle 4.2: Kognitive Aktivitäten der einzelnen Anforderungsbereiche

Reproduzieren	Anwenden	Problemlösen
– Erinnern/Wiedererkennen	– Vergleichen und Klassifizieren	– Analysieren von Problemen
– Definieren	– Anwenden von Modellen	– Integrieren und Synthetisieren
– Beschreiben	– Zusammenhänge herstellen	– Hypothesen und Vorhersagen formulieren
– Angemessene Beispiele finden	– Informationen interpretieren	– Experimente planen
– Kenntnisse über naturwissenschaftliche Instrumente zeigen	– Lösungen finden	– Schlussfolgerungen ziehen
	– Erklären	– Generalisieren
		– Evaluieren
		– Begründen

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011



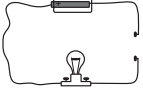

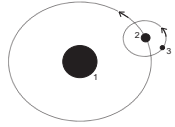
Beim *Reproduzieren* geht es im Kern darum, Fakten zu naturwissenschaftlichen Sachverhalten abzurufen, Sachverhalte zu beschreiben und angemessene Beispiele zu nennen. Beim *Anwenden* liegt der Schwerpunkt hingegen darauf, erlernte naturwissenschaftliche Konzepte auf bekannte Situationen zu übertragen und Zusammenhänge herzustellen, um Lösungen zu finden und Erklärungen zu generieren. Beim *Problemlösen* kommt es darauf an, auch unbekannte naturwissenschaftliche Probleme zu analysieren. Dieser kognitive Anforderungsbereich umfasst bereits anspruchsvolle naturwissenschaftliche Aktivitäten wie das Planen von Experimenten, das Ziehen von Schlussfolgerungen oder das Evaluieren und Begründen.

Vergleicht man die kognitiven Anforderungen, die in der TIMSS-Rahmenkonzeption beschrieben werden, mit denen, die der Perspektivrahmen Sachunterricht vorsieht (GDSU, 2002), so zeigen sich große Überschneidungen der beiden Konzeptionen. Die im Perspektivrahmen Sachunterricht beschriebenen verfahrensbezogenen Kompetenzen umfassen alle drei kognitiven Anforderungsbereiche der TIMSS-Rahmenkonzeption.

3.3 Aufgabenbeispiele zu Inhaltsbereichen und kognitiven Anforderungsbereichen

Um zu illustrieren, wie die drei Inhaltsbereiche und die drei kognitiven Anforderungsbereiche der TIMSS-Rahmenkonzeption in Testaufgaben umgesetzt wurden, zeigt Abbildung 4.1 Beispielaufgaben. Es wird deutlich, dass die Kategorien der kognitiven Anforderungsbereiche nicht immer ganz trennscharf sind. Dies liegt neben theoretischen Überlappungen der Bereiche unter anderem auch daran, dass beispielsweise je nach verfügbarem Vorwissen der Schülerinnen und Schüler unterschiedliche kognitive Prozesse bei der Bearbeitung der Aufgaben vorstellbar sind.

Abbildung 4.1: Zuordnung von Beispielaufgaben zu den Inhaltsbereichen und kognitiven Anforderungsbereichen

	Reproduzieren	Anwenden	Problemlösen										
Biologie	<p>Ein Raubtier ist ein Tier, das sich von anderen Tieren ernährt. Welches dieser Tiere ist ein Raubtier?</p> <p><input type="radio"/> A Hirsch <input checked="" type="radio"/> B Wolf <input type="radio"/> C Kuh <input type="radio"/> D Ziege</p> <p>Bei dieser Aufgabe wird die Information vorgegeben, dass Raubtiere andere Tiere fressen. Aus einer Liste von Tieren muss der Wolf als Raubtier identifiziert werden. Dies erfordert lediglich den Abruf von Wissen über das Verhalten des Wolfes.</p>	<p>Was haben Vögel, Fledermäuse und Schmetterlinge gemeinsam?</p> <p><input type="radio"/> A Federn <input type="radio"/> B Haare <input type="radio"/> C inneres Skelett <input checked="" type="radio"/> D Flügel</p> <p>Bei dieser Aufgabe muss herausgefunden werden, was Vögel, Fledermäuse und Schmetterlinge gemein haben. Hier müssen die drei Tiere verglichen und auf diese Weise die gemeinsame Körperstruktur (Flügel) identifiziert werden.</p>	<p>Bild 1  Bild 2 </p> <p>Bild 1 und 2 zeigen dieselben Augen bei unterschiedlichen Außenbelichtungen. Welche Außenbelichtung ist bei Bild 1 und Bild 2 anders?</p> <p><input type="radio"/> A Das Licht in Bild 1 ist heller. <input checked="" type="radio"/> B Das Licht in Bild 2 ist heller. <input type="radio"/> C Die Temperatur in Bild 1 ist höher. <input type="radio"/> D Die Temperatur in Bild 2 ist höher.</p> <p>Bei dieser Aufgabe werden zwei Augenpaare abgebildet, bei denen die Pupillen unterschiedlich weit geöffnet sind. Aus den Abbildungen muss gefolgert werden, welche „Außenbedingung“ bei den beiden Augenpaaren variierte. Um eine korrekte Schlussfolgerung (Das Licht in Bild 2 ist heller) zu ziehen, muss Wissen über die Funktion des Auges auf die Abbildungen bezogen werden.</p>										
Physik/Chemie	<p>Nenne zwei Dinge, für die man Elektrizität im täglichen Leben braucht.</p> <p>Verwendung 1: <i>Eine Lampe leuchten lassen.</i></p> <p>Verwendung 2: <i>Fernseher</i></p> <p>Bei dieser Aufgabe sollen Dinge des Alltags benannt werden, für die Elektrizität benötigt wird. Um diese Aufgabe zu lösen, müssen zwei angemessene Beispiele gefunden werden (z.B. eine Lampe leuchten lassen).</p>	<p>Florian verbindet eine Batterie, eine Glühlampe und Draht, wie auf dem folgenden Bild gezeigt.</p>  <p>Wird die Birne leuchten? (Kreuze ein Kästchen an.)</p> <p><input type="checkbox"/> Ja <input checked="" type="checkbox"/> Nein</p> <p>Erkläre deine Antwort. <i>Nein, es ist nicht alles verbunden.</i></p> <p>Bei dieser Aufgabe ist ein an einer Stelle unterbrochener „Stromkreis“ abgebildet, und es soll beantwortet werden, ob eine Glühlampe bei diesem Aufbau leuchtet oder nicht. Die Antwort muss erklärt werden. Dabei muss die Vorstellung angewendet werden, dass ein geschlossener Stromkreis vorliegen muss, damit die Glühlampe leuchten kann.</p>	<p>Die Tabelle unten zeigt die Eigenschaften von zwei Materialien.</p> <table border="1" data-bbox="1066 855 1423 981"> <thead> <tr> <th>Eigenschaften von Material 1</th> <th>Eigenschaften von Material 2</th> </tr> </thead> <tbody> <tr> <td>Leitet Wärme schnell</td> <td>Leitet Wärme langsam</td> </tr> <tr> <td>Fest</td> <td>Fest</td> </tr> <tr> <td>Löst sich nicht in Wasser auf</td> <td>Löst sich in Wasser auf</td> </tr> <tr> <td>Wird von Magneten angezogen</td> <td>Wird nicht von Magneten angezogen</td> </tr> </tbody> </table> <p>Welche Aussage über die Materialien 1 und 2 ist am ehesten richtig?</p> <p><input type="radio"/> A Material 1 ist Glas und Material 2 ist Ton. <input type="radio"/> B Material 1 ist Kupfer und Material 2 ist Holz. <input checked="" type="radio"/> C Material 1 ist Eisen und Material 2 ist Zucker. <input type="radio"/> D Material 1 ist Kork und Material 2 ist Gold.</p> <p>Bei dieser Aufgabe werden zwei „unbekannte“ Materialien anhand von vier Eigenschaften beschrieben. Es muss herausgefunden werden, um welche beiden Materialien es sich handelt. Dazu müssen die zur Verfügung stehenden Informationen über die Eigenschaften der Materialien mit dem Wissen über die Eigenschaften der in den Antwortmöglichkeiten vorgegebenen Materialien abgeglichen und es muss eine angemessene Schlussfolgerung gezogen werden.</p>	Eigenschaften von Material 1	Eigenschaften von Material 2	Leitet Wärme schnell	Leitet Wärme langsam	Fest	Fest	Löst sich nicht in Wasser auf	Löst sich in Wasser auf	Wird von Magneten angezogen	Wird nicht von Magneten angezogen
Eigenschaften von Material 1	Eigenschaften von Material 2												
Leitet Wärme schnell	Leitet Wärme langsam												
Fest	Fest												
Löst sich nicht in Wasser auf	Löst sich in Wasser auf												
Wird von Magneten angezogen	Wird nicht von Magneten angezogen												
Geographie	<p>In vielen Teilen der Erde gibt es einen Mangel an Süßwasser. Beschreibe zwei Dinge, die Menschen machen können, um die Verschwendung von Wasser zu verhindern.</p> <p>1. <i>Die Hände mit weniger Wasser waschen.</i></p> <p>2. <i>Darauf achten, dass kein Wasser aus dem Wasserhahn tropft.</i></p> <p>Bei dieser Aufgabe müssen Maßnahmen beschrieben werden, die die Verschwendung von Wasser verhindern können. Um die Aufgabe zu lösen, müssen zwei angemessene Beispiele gefunden werden (z.B. darauf achten, dass kein Wasser aus dem Wasserhahn tropft).</p>	<p>Das Bild zeigt einen Fluss, der durch eine Ebene fließt.</p>  <p>Landwirtschaft wird in der Ebene und entlang des Flusses betrieben. Es hat Vor- und Nachteile, entlang eines Flusses Landwirtschaft zu betreiben.</p> <p>A. Beschreibe einen Vorteil. <i>Man bekommt Wasser.</i></p> <p>Bei dieser Aufgabe soll ein Vorteil genannt werden, Landwirtschaft entlang eines Flusses zu betreiben. Um diese Aufgabe zu lösen, muss bspw. ein Zusammenhang zwischen der Verfügbarkeit von Wasser und der Versorgung von Pflanzen und Tieren hergestellt werden.</p>	<p>Das Bild unten zeigt die Erde, den Mond und die Sonne. Jeder Körper ist mit einer Zahl beschriftet. Die Pfeile zeigen die Richtung, in der sich jeder Körper bewegt.</p>  <p>Schreibe unten neben jedem Körper die richtige Zahl (1, 2 oder 3).</p> <p>Die Erde ist der Körper mit der Nummer: <u>2</u></p> <p>Der Mond ist der Körper mit der Nummer: <u>3</u></p> <p>Die Sonne ist der Körper mit der Nummer: <u>1</u></p> <p>Bei dieser Aufgabe werden drei Himmelskörper schematisch dargestellt und es muss identifiziert werden, bei welchem es sich um die Sonne, die Erde und den Mond handelt. Um diese Aufgabe zu lösen, müssen verfügbare Informationen (z.B. Größe der Himmelskörper, Beziehung der Körper zueinander) mit dem Vorwissen über die Himmelskörper abgeglichen und es muss eine angemessene Schlussfolgerung gezogen werden (Sonne 1, Erde 2, Mond 3).</p>										

4 Der TIMSS-Test zur Erfassung naturwissenschaftlicher Kompetenzen

In diesem Abschnitt wird zunächst auf den Aufbau des Tests eingegangen, der zur Erfassung naturwissenschaftlicher Kompetenzen in TIMSS 2011 eingesetzt wurde (Abschnitt 4.1). Anschließend werden Ergebnisse aus Analysen zur curricularen Validität der Testaufgaben berichtet (Abschnitt 4.2) und es werden die Kompetenzstufen beschrieben, die zur Illustration der Kompetenzen der Schülerinnen und Schüler genutzt werden (Abschnitt 4.3).

4.1 Testaufbau

Die TIMSS-Testaufgaben wurden in einem kooperativen Prozess entwickelt, an dem neben dem *TIMSS & PIRLS International Study Center* am *Boston College* und den nationalen Studienleitungen auch Expertinnen und Experten aus den Bereichen der Naturwissenschaftsdidaktik und der Mathematikdidaktik beteiligt waren (siehe Kapitel 2). Die nationalen Studienleitungen und die beteiligten Vertreterinnen und Vertreter der Fachdidaktik erstellten zunächst Vorschläge für Aufgaben und für Kodiermanuale zu Aufgaben mit offenem Antwortformat. Die Aufgaben und Kodiermanuale wurden anschließend von Experten aus verschiedenen Bereichen begutachtet und überarbeitet. Dabei wurde unter anderem sichergestellt, dass die Aufgaben akkurat in die im Unterricht des jeweiligen Landes übliche Sprache übersetzt wurden. Im Rahmen großer nationaler Feldtests wurden die Aufgaben erprobt und es wurden die finalen Testaufgaben für die Haupterhebung ausgewählt. Durch die umfangreichen Überprüfungsverfahren sollte sichergestellt werden, dass die Aufgaben in psychometrischer Hinsicht und für die Schülerinnen und Schüler des jeweiligen Staates geeignet sind.

Das Ziel der TIMSS-Rahmenkonzeption, naturwissenschaftliche Kompetenzen in den Inhaltsbereichen *Biologie*, *Geographie* und *Physik/Chemie* sowie den kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* breit abzudecken, machte die Entwicklung einer großen Zahl an Aufgaben erforderlich. Der finale Aufgabenpool im Bereich Naturwissenschaften bei TIMSS 2011 umfasst 175 Aufgaben für das vierte Schuljahr. Etwas mehr als die Hälfte der Testaufgaben, nämlich 96 Testaufgaben, werden im *Multiple-Choice*-Format vorgegeben. Die übrigen 79 Testaufgaben erfordern eine kurze schriftliche Antwort von den Schülerinnen und Schülern. 103 der insgesamt 175 Aufgaben (59%) wurden aus den Studien TIMSS 2007 und TIMSS 2003 übernommen, um die Daten aus 2011 mit denen der früheren Studien in Beziehung setzen zu können. Die verbleibenden 72 Aufgaben (41%) wurden für TIMSS 2011 neu entwickelt.

Entsprechend der Struktur der Rahmenkonzeption von TIMSS lassen sich die einzelnen Testaufgaben jeweils einem Inhaltsbereich und einem kognitiven Anforderungsbereich zuordnen (s.o.). Dadurch können zusätzlich zur Erfassung der Kompetenz auf der Gesamtskala Naturwissenschaften auch Kompetenzen im Hinblick auf die einzelnen Inhalts- und Anforderungsbereiche untersucht werden.

Tabelle 4.3 zeigt, wie sich die Aufgaben auf die Inhaltsbereiche *Biologie*, *Physik/Chemie* und *Geographie* verteilen. Mit 45 Prozent ist der Inhaltsbereich *Biologie* am stärksten repräsentiert. Es folgt mit 36 Prozent der Bereich *Physik/Chemie*. Der Inhaltsbereich *Geographie* ist mit 20 Prozent der Testaufgaben am schwächsten vertreten.

Tabelle 4.3: Verteilung der Testaufgaben auf die naturwissenschaftlichen Inhaltsbereiche

Naturwissenschaftliche Inhaltsbereiche	Testaufgaben TIMSS 2011	
	Anzahl	Prozent
Biologie	78	45
Physik/Chemie	63	36
Geographie	34	20
Gesamt	175	100

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Dass Aufgaben zum Inhaltsbereich *Geographie* im Vergleich zu den Bereichen *Biologie* und *Physik/Chemie* den kleinsten Anteil ausmachen, hängt mit der lehrplanorientierten Rahmenkonzeption von TIMSS zusammen (Mullis et al., 2009). Der Aufgabenentwicklung in TIMSS liegt ein internationales Kerncurriculum zugrunde, das eine Schnittmenge der Curricula der Teilnehmerstaaten bildet. In diesem Kerncurriculum sind geographische Inhalte nicht so stark repräsentiert wie biologische oder physikalisch-chemische. Im Sachunterricht in Deutschland sind hingegen geographische Inhalte vergleichsweise stärker repräsentiert. Dies zeigt sich in Schulbuchanalysen (z.B. Blaseio, 2009a) wie auch in normativen Vorgaben, beispielsweise im Perspektivrahmen Sachunterricht (GDSU, 2002). Physikalische und chemische Inhalte sind im Verhältnis zu geographischen und biologischen Inhalten im Sachunterricht – auch nach der Konsolidierung seit der Jahrtausendwende – geringer vertreten (Blaseio, 2009a; GDSU, 2002).

Tabelle 4.4: Verteilung der Testaufgaben auf die kognitiven Anforderungsbereiche

Kognitive Anforderungsbereiche	Testaufgaben TIMSS 2011	
	Anzahl	Prozent
Reproduzieren	72	41
Anwenden	71	41
Problemlösen	32	18
Gesamt	175	100

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Tabelle 4.4 zeigt die Verteilung der Aufgaben auf die drei kognitiven Anforderungsbereiche. Die beiden Bereiche *Reproduzieren* und *Anwenden* sind mit jeweils 41 Prozent der Testaufgaben gleich stark vertreten. Nur 18 Prozent der Testaufgaben erfordern *Problemlösen*.

Die Verteilung der Aufgaben auf die kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen* macht deutlich, dass in TIMSS nicht nur einfaches Faktenwissen erfasst wird. Vielmehr erfordern die Aufgaben unter anderem, Wissen auf neue naturwissenschaftliche Phänomene zu übertragen, um Erklärungen generieren zu können (Anforderungsbereich *Anwenden*). Auch müssen Probleme analysiert, Experimente geplant und zulässige Schlussfolgerungen gezogen werden (Anforderungsbereich *Problemlösen*). Diese anspruchsvollen kognitiven Tätigkeiten sind auch im Perspektivrahmen Sachunterricht bis zum Ende

des vierten Schuljahres vorgesehen und finden sich in aktuellen Konzeptionen naturwissenschaftlicher Grundbildung im Rahmen des Sachunterrichts wieder (Einsiedler et al., 2008).

4.2 Curriculare Validität

Der TIMSS-Rahmenkonzeption zur Erfassung naturwissenschaftlicher Kompetenzen liegt der Anspruch zugrunde, Kompetenzen zu erfassen, die in der Schule aufgrund von curricularen Vorgaben vermittelt wurden. Inwieweit die Erfassung naturwissenschaftlicher Kompetenzen in TIMSS 2011 diesem Anspruch gerecht wird, wurde für Deutschland überprüft. Dazu wurden die in der TIMSS-Rahmenkonzeption vorgesehenen Inhaltsbereiche mit den curricularen Vorgaben zum naturwissenschaftlichen Lernen im Rahmen des Sachunterrichts abgeglichen. So konnte für jede der 175 in TIMSS 2011 eingesetzten Aufgaben zur Erfassung naturwissenschaftlicher Kompetenzen entschieden werden, ob sie einen Inhaltsbereich thematisiert, der curricular vorgesehen ist oder nicht.

Eine Schwierigkeit für die Ermittlung der curricularen Validität der naturwissenschaftlichen Testaufgaben stellt die Heterogenität der curricularen Vorgaben für den Sachunterricht in Deutschland dar. Da Bildungsfragen in Deutschland in die Zuständigkeit der Länder fallen, liegen für jedes Land eigene Lehrpläne zum Sachunterricht vor. Anders als im Fach Mathematik gibt es für das Fach Sachunterricht keine länderübergreifenden Bildungsstandards. Aus diesen Gründen wurde exemplarisch der Lehrplan aus Nordrhein-Westfalen für die Analyse der curricularen Validität der Aufgaben herangezogen. Er spiegelt verschiedene Entwicklungen des Sachunterrichts in Deutschland gut wider. Diese sind zunächst die eingangs beschriebenen Tendenzen bezüglich der Berücksichtigung naturwissenschaftlicher Inhalte im Sachunterricht. Darüber hinaus werden auch naturwissenschaftliche Denk- und Arbeitsweisen als bedeutsame Zielbereiche berücksichtigt und inhalts- und verfahrensbezogene Anforderungen als Kompetenzen im Sinne von „can do statements“ formuliert (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008).

Mit der Prüfung der curricularen Validität wurde eine Expertengruppe beauftragt. Als zentrales Dokument lag den Experten neben den Testaufgaben der nordrhein-westfälische Lehrplan für das Fach Sachunterricht aus dem Jahre 2008 vor (MSW, 2008). In Fällen, in denen keine klare Entscheidung möglich war, wurde zusätzlich der Perspektivrahmen Sachunterricht (GDSU, 2002) hinzugezogen.

Tabelle 4.5 zeigt die Anteile der nicht curricular validen Testaufgaben, aufgeschlüsselt nach den drei Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geographie*. Im Bereich *Physik/Chemie* ist der Anteil nicht curricular valider Aufgaben mit 8 Prozent gering. In den Bereichen *Biologie* (27%) und *Geographie* (42%) sind die Anteile deutlich größer. Diese Ergebnisse der Analyse zur curricularen Validität bedeuten, dass eine vergleichsweise hohe Anzahl an Testaufgaben in den Bereichen *Biologie* und *Geographie* Themen behandelt, die den Schülerinnen und Schülern zumindest aus dem Sachunterricht weitgehend unbekannt sein dürften. Insgesamt wird aber auch anhand der Zahlen deutlich, dass über 75 Prozent der insgesamt eingesetzten Aufgaben curricular verankert sind.

Abbildung 4.2 zeigt zwei Beispiele der Aufgaben, die als nicht curricular valide kategorisiert wurden.

Tabelle 4.5: Anteile curricular nicht valider Testaufgaben nach naturwissenschaftlichen Inhaltsbereichen

Naturwissenschaftliche Inhaltsbereiche	Testaufgaben TIMSS 2011		curricular nicht valide Testaufgaben	
	Anzahl	Anzahl	Anzahl	Prozent ^A
Biologie	78	21		27
Physik/Chemie	63	5		8
Geographie	34	14		42
Gesamt	175	40		23

A = Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

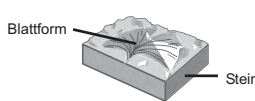
Tabelle 4.6: Übergeordnete Themen von curricular nicht validen Testaufgaben

Inhaltsbereiche	Übergeordnete Themen curricular nicht valider Testaufgaben
Biologie	Vererbung Funktion von Früchten Stoffwechsel der Pflanze (Photosynthese) Stoffwechsel des Menschen (Atmung) Brutpflege bei spezifischen Arten Ontogenetische Entwicklung bei spezifischen Arten (z.B. Motte)
Physik/Chemie	Erdanziehung
Geographie	Fossilien, Erdgeschichte Morphologie der Erde Sonnensystem, Erdrotation

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Abbildung 4.2: Beispiele curricular nicht valider Testaufgaben

Aufgabenbeispiel zur Geographie [Geographie: Fossilien/Erdgeschichte]	Aufgabenbeispiel zur Geographie [Geographie: Die Erde im Sonnensystem]
<p>Die Blattform im Stein stammt von einer Pflanze, die vor langer Zeit gelebt hat.</p> <div style="text-align: center;">  </div> <p>Vor wie langer Zeit hat die Pflanze ungefähr gelebt?</p> <p><input type="radio"/> A vor einem Jahr</p> <p><input type="radio"/> B vor einhundert Jahren</p> <p><input type="radio"/> C vor eintausend Jahren</p> <p><input checked="" type="radio"/> D vor einer Million Jahren (.28/.42)</p>	<p>Wie oft dreht sich die Erde um ihre Achse?</p> <p><input type="radio"/> A einmal alle 12 Stunden</p> <p><input checked="" type="radio"/> B einmal alle 24 Stunden</p> <p><input type="radio"/> C einmal jeden Monat</p> <p><input type="radio"/> D einmal jedes Jahr</p> <p style="text-align: right;">(.53/.54)</p>

Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Tabelle 4.6 zeigt übergeordnete Themenbereiche, denen die Aufgaben, die als nicht curricular valide eingeschätzt wurden, zugeordnet werden können. Im Bereich *Geographie* sieht der Lehrplan Nordrhein-Westfalen in erster Linie topographische Themen vor, beispielsweise über die nähere Region, Deutschland und Europa. Themen wie Erdgeschichte, Morphologie der Erde oder das Sonnensystem sind daher im Lehrplan nicht ausgewiesen.

In Tabelle 4.7 ist dargestellt, wie sich die als nicht curricular valide eingeschätzten Aufgaben auf die drei kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen* verteilen. Hier wird deutlich, dass insbesondere der Bereich *Reproduzieren* betroffen ist. Ein Drittel der Testaufgaben aus diesem Bereich wurde als nicht curricular valide kategorisiert. In den meisten Fällen war dies darauf zurückzuführen, dass die Schülerinnen und Schüler zur Lösung der Aufgaben über Faktenwissen verfügen müssen, das im Lehrplan Nordrhein-Westfalen nicht vorgesehen ist. Die in Abbildung 4.2 gezeigte Aufgabe zur Erdrotation ist ein Beispiel dafür.

Tabelle 4.7: Anteile curricular nicht valider Testaufgaben nach kognitiven Anforderungsbereichen

Kognitive Anforderungsbereiche	Testaufgaben TIMSS 2011	curricular nicht valide Testaufgaben	
	Anzahl	Anzahl	Prozent ^A
Reproduzieren	71	21	30
Anwenden	71	13	18
Problemlösen	35	3	9
Gesamt	177	37	21

A = Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Insgesamt lässt sich festhalten, dass der Anteil der in TIMSS 2011 eingesetzten naturwissenschaftlichen Aufgaben, die als curricular valide eingeschätzt wurden, mit 77 Prozent noch zufriedenstellend hoch ist. Der Anspruch der TIMSS-Rahmenkonzeption, Kompetenzen zu erfassen, die in der Schule aufgrund von curricularen Vorgaben vermittelt werden, wurde also in Deutschland weitgehend erfüllt. In den Inhaltsbereichen *Biologie* (27%) und insbesondere *Geographie* (42%) wurden jedoch recht große Anteile der Aufgaben als nicht curricular valide kategorisiert. Aus diesem Grund wurde gezielt geprüft, ob Einschränkungen der curricularen Validität der Aufgaben einen Einfluss auf die Ergebnisse der Schülerinnen und Schüler im Bereich Naturwissenschaften in Deutschland haben. Darauf wird im Ergebnisteil (Abschnitt 5.1) näher eingegangen.

4.3 Kompetenzstufen in den Naturwissenschaften

Die TIMSS-Rahmenkonzeption deckt ein breites Spektrum von naturwissenschaftlichen Teilkompetenzen ab, da verschiedene Inhaltsbereiche und kognitive Anforderungen angesprochen werden (siehe Abschnitt 3). Um die erreichten Leistungswerte der Schülerinnen und Schüler besser interpretieren zu können, werden in TIMSS fünf Kompetenzstufen unterschieden. Mit Hilfe dieser Kompetenzstufen können Gruppen von Schülerinnen und Schülern identifiziert werden, die sich in ihren inhaltlichen und kognitiven Kompetenzen in den Naturwissenschaften voneinander unterscheiden.

Die Bildung solcher Kompetenzstufen ist möglich, da bei TIMSS Skalierungsmethoden verwendet werden (siehe Kapitel 2), die eine inhaltliche Interpretation der von den Schülerinnen und Schülern erreichten Leistungswerte erlauben. Über eine Analyse der inhaltlichen und kognitiven Anforderungen, die zur erfolgreichen Bearbeitung der Testaufgaben erforderlich sind, können verschiedene Stufen der naturwissenschaftlichen Kompetenz unterschieden werden. Die Kompetenzstufen basieren auf vier Bezugspunkten, sogenannte Benchmarks, die in TIMSS festgelegt werden (Mullis et al., 2009; siehe auch Kapitel 2). Diese Bezugspunkte teilen die Kompetenzskala für die Naturwissenschaften in insgesamt fünf Intervalle, die den Kompetenzstufen entsprechen. Die Leistungswerte der Schülerinnen und Schüler auf der Naturwissenschaftsskala lassen sich somit eindeutig einer der fünf Kompetenzstufen zuordnen. Welcher Kompetenzstufe Schülerinnen und Schüler mit einem bestimmten Wert auf der Kompetenzskala zugeordnet werden, kann Tabelle 4.8 entnommen werden. Die Kompetenzstufen sind hierarchisch geordnet, das heißt Schülerinnen und Schüler, die sich beispielsweise auf Kompetenzstufe III befinden, lösen Testaufgaben, die typisch für diese Stufe sind, mit einer größeren Wahrscheinlichkeit als Testaufgaben der Stufe IV. Tabelle 4.8 fasst für jede der fünf Kompetenzstufen zusammen, über welche naturwissenschaftliche Kompetenz die Schülerinnen und Schüler verfügen. Eine ausführliche Beschreibung befindet sich im Anhang C dieses Bandes.

Schülerinnen und Schüler mit einem Leistungswert unter 400 Punkten befinden sich auf der untersten Kompetenzstufe. Sie verfügen über nur rudimentäres Anfangswissen in den Naturwissenschaften. Mit ansteigendem Punktwert auf der Leistungsskala werden die in den Testaufgaben gestellten inhaltlich-kognitiven Anforderungen komplexer. Während die Testaufgaben der unteren Kompetenzstufen vor allem das Reproduzieren von Faktenwissen und eine direkte Anwendung naturwissenschaftlichen Wissens sowie ein Verständnis einfacher naturwissenschaftlicher Konzepte erfordern (Kompetenzstufen II und III), verlangen die Testaufgaben der oberen Kompetenzstufen das Erklären von Alltagsphänomenen sowie das Anwenden naturwissenschaftlicher Verfahrensweisen und das Verstehen naturwissenschaftlicher Zusammenhänge in weitgehend unbekannt Situationen (Kompetenzstufen IV und V).

Tabelle 4.8: Beschreibung der fünf Kompetenzstufen für die Gesamtskala Naturwissenschaften

Kompetenzstufe V (ab 625): Beginnendes naturwissenschaftliches Denken
<i>Die Schülerinnen und Schüler weisen ein grundlegendes Verständnis des Prozesses naturwissenschaftlichen Arbeitens auf und können ihr Wissen über naturwissenschaftliche Prozesse und Zusammenhänge anwenden. Sie teilen ihr Wissen über Eigenschaften und Prozesse von Organismen, Fortpflanzung und Entwicklung, Ökosysteme und das Interagieren von Organismen mit Lebensräumen sowie über Faktoren gesundheitsbezogener Lebensführung aktiv mit. Sie zeigen ein grundlegendes Verständnis der Eigenschaften des Lichts, der Zusammenhänge verschiedener physikalischer/chemischer Eigenschaften bekannter Materialien und Stoffe sowie der Kräfte im Zusammenhang mit Magnetismus und Erdanziehung. Sie zeigen ferner praktisches Wissen im Bereich der Elektrizität und weisen ein grundlegendes Verständnis des Sonnensystems sowie von Landschaftsmerkmalen und Veränderungen der Erde auf. Die Schülerinnen und Schüler besitzen elementare Fähigkeiten, Ergebnisse im Kontext einfacher Experimente zu interpretieren, Schlüsse zu ziehen, Argumente zu bewerten und diese zu vertreten.</i>
Kompetenzstufe IV (550-624): Erklären von Alltagsphänomenen
<i>Die Schülerinnen und Schüler können mit ihrem Wissen und Verständnis alltägliche Phänomene erklären. Sie besitzen ein grundlegendes Verständnis des Aufbaus von Pflanzen, des Körperbaus von Tieren sowie von Fortpflanzung und Lebenszyklen. Sie besitzen Wissen über natürliche Prozesse und über die Umwelt und verfügen über grundlegende Kenntnisse über Eigenschaften von Materialien und physikalische/chemische Phänomene. Die Schülerinnen und Schüler besitzen ferner grundlegendes Wissen über das Sonnensystem und über die Struktur, die Veränderungen und Ressourcen der Erde. Sie zeigen elementare Kenntnisse und Fähigkeiten bezüglich naturwissenschaftlicher Denk- und Arbeitsweisen. Sie vergleichen, kontrastieren und ziehen einfache Schlussfolgerungen. Sie geben Erklärungen, bei denen sie ihr Wissen über naturwissenschaftliche Konzepte mit alltäglichen Erfahrungen und auch abstrakten Kontexten verbinden.</i>
Kompetenzstufe III (475-549): Anwenden grundlegenden Alltagswissens
<i>Die Schülerinnen und Schüler können Basiswissen und grundlegendes Verständnis auf naturwissenschaftsbezogene Situationen anwenden. Sie kennen grundlegende Begriffe der belebten Natur und ihrer Interaktion mit der Umwelt und weisen ein basales Verständnis der Biologie und der Gesundheit des Menschen auf. Darüber hinaus verfügen sie über ein grundlegendes Wissen über Eigenschaften von Materialien sowie über Licht, elektrischen Strom, Energie und Kräfte. Außerdem kennen sie basale Fakten über das Sonnensystem und die Ressourcen der Erde. Sie verfügen über die Fähigkeiten, Informationen aus Bilddiagrammen zu interpretieren und Faktenwissen auf naturwissenschaftsbezogene Situationen anzuwenden.</i>
Kompetenzstufe II (400-474): Reproduzieren elementaren Faktenwissens
<i>Die Schülerinnen und Schüler besitzen elementares Wissen über Biologie und Physik/Chemie. Sie besitzen Wissen über einfache Fakten zur menschlichen Gesundheit, zu Ökosystemen sowie zum Verhalten und Körperbau von Tieren. Die Schülerinnen und Schüler zeigen basales Wissen über Energie und kennen einige Eigenschaften von Materialien. Sie interpretieren einfache Diagramme, vervollständigen einfache Tabellen und geben kurze schriftliche Erläuterungen auf Fragen nach naturwissenschaftlichen Fakten.</i>
Kompetenzstufe I (unter 400): Rudimentäres Anfangswissen
<i>Die Schülerinnen und Schüler verfügen über rudimentäres schulisches Anfangswissen. Selbst einfache Aufgaben werden nur gelegentlich oder ansatzweise gelöst.</i>

Abbildung 4.3 und Abbildung 4.4 zeigen anhand von Beispielen, wie unterschiedliche Aufgaben den Kompetenzstufen zugeordnet sind. Hier dienen Aufgaben aus den Inhaltsbereichen *Biologie* und *Physik/Chemie* der Illustration.

Beispiel für Kompetenzstufe V. Bei Aufgabe 1 sollen die Schülerinnen und Schüler zunächst zentrale Bestandteile einer Pflanze benennen: Blüte, Stängel, Blatt und Wurzel. Dann sollen sie die Funktion des jeweiligen Pflanzenteils beschreiben. Korrekte Antworten wären hier beispielsweise „(die Blüte) lockt Insekten zur Pflanze, „(der Stängel) leitet Wasser“, „(das Blatt) nimmt Sonnenlicht auf“ und „(die Wurzel) nimmt Wasser auf“. Um die volle Punktzahl

bei dieser Aufgabe zu erhalten, genügt es, die Bestandteile korrekt zu benennen und drei Funktionen korrekt zu beschreiben. Die Aufgabe erfordert somit bereits ein erstes konzeptuelles Verständnis des Aufbaus und der physiologischen Prozesse bei Pflanzen.

Beispiele für Kompetenzstufe IV. Bei Aufgabe 2 werden drei Körper präsentiert, die nach ihrem Volumen angeordnet sind: ein Styroporklotz, ein Ziegel und ein Apfel. Die Schülerinnen und Schüler sollen zunächst entscheiden, ob ein Kind, das behauptet, Objekte mit größerem Volumen wögen mehr, Recht hat. Die Entscheidung soll dann noch begründet werden. Um diese Aufgabe lösen zu können, müssen die Schülerinnen und Schüler die im Grundschulalter oft noch vorzufindende Vorstellung überwunden haben, dass „größere“ Dinge notwendigerweise auch „schwerer“ sind. Außerdem wird in der Aufgabe mit dem Begriff Volumen operiert, was die Schwierigkeit der Aufgabe zusätzlich erhöhen dürfte. Die Erkenntnis, dass Gegenstände mit einem größeren Volumen nicht notwendigerweise auch schwerer sind, ist eine wichtige Voraussetzung für das Verständnis des Konzepts der Dichte.

Bei Aufgabe 3 sollen die Schülerinnen und Schüler in einer Abbildung, die einen Teich darstellt, drei Lebewesen und drei Nicht-Lebewesen identifizieren und benennen. Korrekte Lösungen für Lebewesen in der Abbildung sind beispielsweise Frosch, Fisch und Baum. Korrekte Lösungen für Nicht-Lebewesen sind beispielsweise Stein, Sonne und Wolke. Um diese Aufgabe lösen zu können, müssen die Schülerinnen und Schüler ein erstes konzeptuelles Verständnis des „Lebendigen“ entwickelt haben und dieses auf eine weitgehend bekannte Situation anwenden.

Beispiele für Kompetenzstufe III. In Aufgabe 4 werden den Schülerinnen und Schülern Abbildungen von vier Tieren gezeigt. Anschließend wird jeweils ein Tier anhand von zwei Merkmalen charakterisiert und die Schülerinnen und Schüler sollen entscheiden, um welches der vier Tiere es sich jeweils handelt. Um diese Aufgabe lösen zu können, müssen die Schülerinnen und Schüler Wissen über den Körperbau von Tieren anwenden.

Bei Aufgabe 5 wird den Schülerinnen und Schülern eine Liste aus fünf Materialien vorgegeben und es soll entschieden werden, welche dieser Materialien „brennen“. Die korrekten Lösungen sind Holz und Benzin. Die Lösung dieser Aufgabe erfordert eine Reproduktion von Wissen oder Erfahrungen zur Brennbarkeit von Materialien. Die Aufgabe kann mit Alltagswissen beantwortet werden und erfordert kein tiefergehendes physikalisches Verständnis.

Beispiele für Kompetenzstufe II. In Aufgabe 6 zeigt eine Abbildung einen an einer Stelle unterbrochenen „Stromkreis“ mit einer Glühbirne und einer Batterie. Den Schülerinnen und Schülern wird eine Liste von vier Gegenständen vorgegeben. Sie sollen entscheiden, welcher Gegenstand geeignet ist, die Glühlampe zum Leuchten zu bringen, wenn er an der Stelle eingesetzt wird, an der der „Stromkreis“ unterbrochen ist. Diese Aufgabe kann gelöst werden, indem elementares Wissen über die Leitfähigkeit bekannter Materialien auf den dargestellten einfachen „Stromkreis“ angewendet wird.

In Aufgabe 7 werden drei Tiere genannt, die Grundschulkindern weitgehend bekannt sind: Vögel, Fledermäuse und Schmetterlinge. Die Schülerinnen und Schüler sollen identifizieren, was die Tiere gemeinsam haben. Dazu wird eine Liste aus vier Körperstrukturen (Federn, Haare, inneres Skelett und Flügel) vor-

Abbildung 4.3: Kompetenzstufen und Beispielaufgaben (Gesamtskala Naturwissenschaften) I

Kompetenzstufen

V

625

(.21/.10)

IV

550

III

475

(.42/.51)

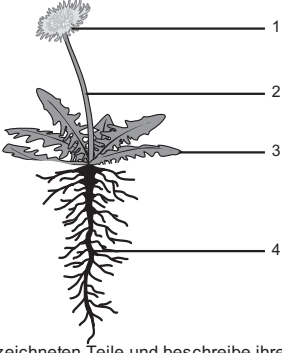
II

400

I

(.47/.40)

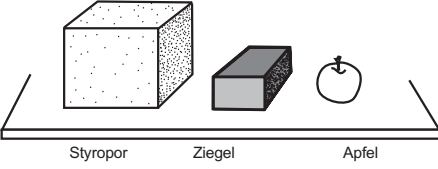
1) Die Abbildung zeigt eine Blütenpflanze. Vier Teile dieser Pflanze sind nummeriert.



Benenne die bezeichneten Teile und beschreibe ihre Aufgabe in der folgenden Tabelle

Nummer des Teils	Name des Teils	Aufgabe des Teils
1	Blüte	lockt Insekten zur Pflanze
2	Stängel	leitet Wasser
3	Blatt	nimmt Sonnenlicht auf
4	Wurzel	nimmt Wasser auf

2) Jakobs Lehrerin legt drei Dinge auf einen Tisch. Das siehst du unten. Sie ordnet diese nach ihrem Volumen an.

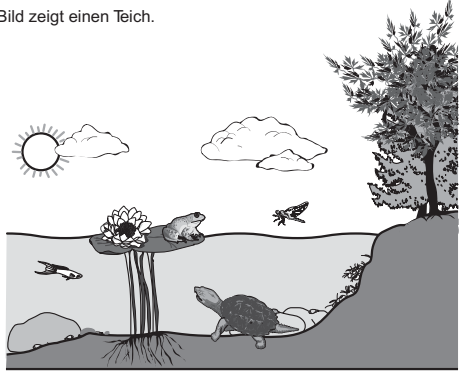


Jakob glaubt, dass Objekte mit größerem Volumen mehr wiegen. Bist du seiner Meinung? Bitte kreuze ein Kästchen an.

Ja
 Nein

Erkläre deine Antwort.
Das Styropor ist viel größer als die anderen beiden Gegenstände und wiegt wahrscheinlich weniger.

3) Das Bild zeigt einen Teich.

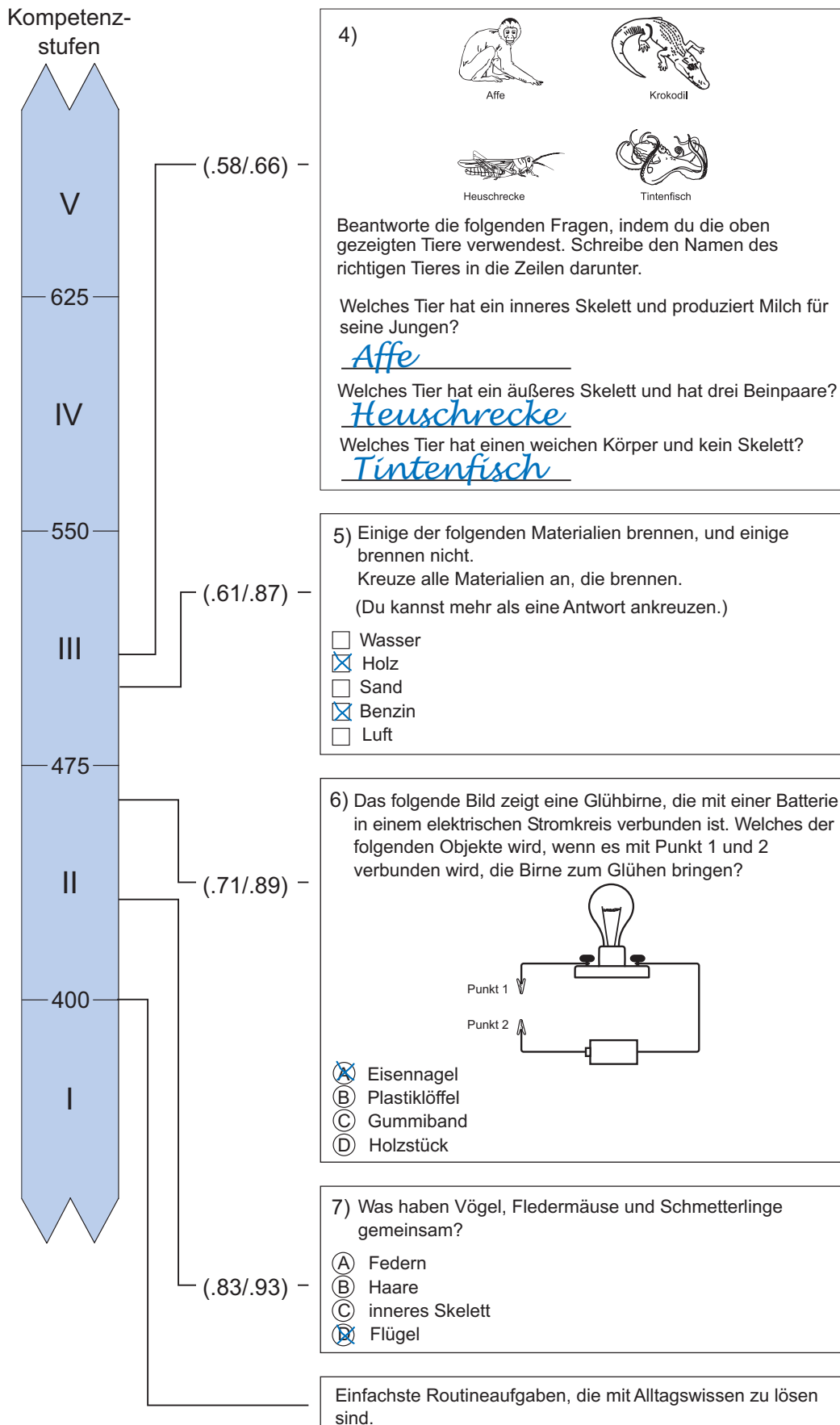


Schreibe drei Lebewesen auf, die in diesem Bild zu sehen sind, und drei Dinge in diesem Bild, die keine Lebewesen sind.

Lebewesen	Keine Lebewesen
1. Frosch	1. Stein
2. Fisch	2. Sonne
3. Baum	3. Wolke

Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an.

Abbildung 4.4: Kompetenzstufen und Beispielaufgaben (Gesamtskala Naturwissenschaften) II



Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an.

gegeben. Um die Aufgabe zu lösen, müssen die Schülerinnen und Schüler die drei Tiere (ohne eine konkrete Darstellung nutzen zu können) vergleichen. Elementares Wissen über Körperstrukturen bei Tieren ist dabei ausreichend.

5 Ergebnisse

Die Ergebnisse des internationalen Vergleichs zur naturwissenschaftlichen Kompetenz von Schülerinnen und Schülern am Ende der 4. Jahrgangsstufe werden in diesem Abschnitt entlang der in Abschnitt 1 aufgeführten Fragestellungen dargestellt. Das heißt, es wird zunächst auf die Ergebnisse des internationalen Vergleichs anhand der Gesamtskala zur naturwissenschaftlichen Kompetenz eingegangen (Abschnitt 5.1). Dann wird berichtet, wie sich die Schülerinnen und Schüler in den Teilnehmerstaaten auf die fünf Kompetenzstufen verteilen (Abschnitt 5.2). Hier ist eine wichtige Frage, wie groß die Gruppen der auffällig leistungsschwachen und leistungsstarken Kinder sind. Anschließend werden die Ergebnisse zu den Teilkompetenzen dargestellt, und zwar aufgeschlüsselt nach Inhaltsbereichen (Abschnitt 5.3) und nach kognitiven Anforderungsbereichen (Abschnitt 5.4). In den Abschnitten 5.1 bis 5.4 wird neben den Ergebnissen aus 2011 berichtet, inwieweit sich die Ergebnisse von TIMSS 2007 und TIMSS 2011 unterscheiden. Der Abschnitt wird abgeschlossen mit Ergebnissen zum Selbstkonzept und zu Einstellungen der Schülerinnen und Schüler im Bereich Naturwissenschaften. Auf Ergebnisse zu Unterschieden in der naturwissenschaftlichen Kompetenz in Abhängigkeit von sozialer Herkunft, Migrationshintergrund und Geschlecht wird in den Kapiteln 6, 7 und 8 dieses Bandes eingegangen.

Deutschland nahm 2011 das zweite Mal an der TIMSS-Grundschuluntersuchung seit Einführung der Studie im Jahr 1995 teil. Im Jahr 2011 waren 50 Staaten und Regionen beteiligt, darunter 21 EU- und 26 OECD-Staaten. Zusätzlich beteiligten sich drei Länder, die mit der Jahrgangsstufe 6 teilnahmen, sowie sieben sogenannte Benchmark-Teilnehmer, bei denen es sich um bestimmte Regionen aus den USA, Kanada und den Vereinigten Arabischen Emiraten handelt (vgl. Kapitel 2).

Damit die Ergebnisse zur naturwissenschaftlichen Kompetenz der Schülerinnen und Schüler in Deutschland eingeordnet werden können, werden in diesem Abschnitt neben dem Vergleich mit einzelnen anderen Staaten drei weitere Vergleichsmöglichkeiten angeboten, für die die Werte unterschiedlicher Staaten zu sogenannten Vergleichsgruppen (VG) zusammengefasst wurden: Die Ergebnisse in Deutschland können (1) mit dem Mittelwert aller Staaten (ohne Benchmark-Teilnehmer und Teilnehmer mit Jahrgangsstufe 6), (2) mit dem Mittelwert der beteiligten OECD-Staaten (VG_{OECD}) und (3) mit dem Mittelwert der beteiligten EU-Staaten (VG_{EU}) verglichen werden. Im Unterschied zu den Abbildungen, in denen Ergebnisse nur aus TIMSS 2011 dargestellt sind, wird bei allen Abbildungen, in denen Unterschiede zwischen TIMSS 2007 und TIMSS 2011 berichtet werden, auf die Angabe des internationalen Mittelwerts sowie des Mittelwerts der VG_{EU} und der VG_{OECD} verzichtet, da es in den beiden Untersuchungen unterschiedliche Teilnehmer und auch unterschiedliche Zusammensetzungen dieser Gruppen gab.

5.1 Kompetenzen im internationalen Vergleich

In diesem Abschnitt werden die Ergebnisse für die Gesamtskala Naturwissenschaften im internationalen Vergleich berichtet. Es wird zunächst auf Unterschiede im Niveau naturwissenschaftlicher Kompetenz zwischen den Staaten, auf die Streuung der Kompetenz innerhalb von Staaten sowie auf die Bedeutung curricular nicht valider Aufgaben für die Ergebnisse zum Niveau naturwissenschaftlicher Kompetenz im internationalen Vergleich eingegangen. Anschließend werden Ergebnisse zu Unterschieden im Niveau der naturwissenschaftlichen Kompetenz zwischen TIMSS 2007 und 2011 sowie Ergebnisse zu Unterschieden in der Streuung der Leistungsmittelwerte dargestellt.

Unterschiede im Niveau der naturwissenschaftlichen Kompetenz im internationalen Vergleich: Abbildung 4.5 gibt für die Kompetenz in den Naturwissenschaften die Mittelwerte (M) und Standardabweichungen (SD) mit ihren jeweiligen Standardfehlern (SE) für jeden Staat an. Die Mittelwerte entsprechen dem Mittelwert der Leistungswerte der Schülerinnen und Schüler eines Staates. Die Standardabweichung ist ein Maß der Streuung der Werte innerhalb eines Staates. Zusätzlich zu den Mittelwerten einzelner Staaten ist der Mittelwert für alle Teilnehmerstaaten (Internationaler Mittelwert), für die teilnehmenden OECD-Staaten (VG_{OECD}) und für die teilnehmenden EU-Staaten (VG_{EU}) angegeben. Die Staaten sind entsprechend ihrem naturwissenschaftlichen Leistungsmittelwert in einer Rangreihe angeordnet.

Das Format der Bezeichnungen der Teilnehmerstaaten sowie die beige gestellten Fußnoten verweisen auf Besonderheiten bezüglich der nationalen Untersuchungspopulationen und -stichproben. Die Bedeutung der einzelnen Fußnoten wird ausführlich in Kapitel 2 erläutert. Um eine internationale Vergleichbarkeit der Ergebnisse zu gewährleisten, verlangt die internationale Studienleitung die Einhaltung strenger methodischer Standards. Teilnehmer, die aus unterschiedlichen Gründen diese Vorgaben nicht eingehalten haben, sind durch eine Kursivsetzung ihres Namens gekennzeichnet. Dies trifft auf 26 der 59 Teilnehmerstaaten und Regionen zu. Während für die anderen 33 Teilnehmer – zu denen auch Deutschland zählt – eine internationale Vergleichbarkeit der Ergebnisse gegeben ist, gilt für die kursiv gesetzten Teilnehmer, dass Schlüsse auf die Population aller Viertklässlerinnen und Viertklässler des jeweiligen Teilnehmerstaates beziehungsweise der Region nur mit gewissen Einschränkungen zulässig sind. Für den Vergleich der Ergebnisse Deutschlands mit kursiv gesetzten Teilnehmerstaaten und Regionen ist unter Berücksichtigung der Tabellen im Anhang A für den Einzelfall zu prüfen, inwiefern ein differentes Abschneiden durch Unterschiede in den Teilnahmebedingungen bedingt sein könnte.

Wie Abbildung 4.5 entnommen werden kann, steht die Republik Korea (Südkorea) mit 587 Punkten an der Spitze der Leistungsskala, gefolgt von Singapur mit 583 Punkten und Finnland mit 570 Punkten. Tunesien, Marokko und Jemen sind hingegen diejenigen Staaten, die mit Leistungsmittelwerten unter 350 Punkten am unteren Ende der Leistungsskala zu finden sind. In Deutschland erreichen die Schülerinnen und Schüler in den Naturwissenschaften einen Leistungsmittelwert von 528 Punkten und sind damit auf dem Rangplatz 17 (von 50) zu finden. Sie liegen mit ihrer Kompetenz deutlich über dem internationalen Mittelwert (486 Punkte). Von dem Mittelwert der OECD-Staaten (523) und dem Mittelwert der EU-Staaten (521) unterscheidet sich der Mittelwert für Deutschland allerdings deutlich weniger. Der Abstand zu den Staaten mit dem

höchsten naturwissenschaftlichen Kompetenzniveau ist mit 59 Punkten (Republik Korea) und 55 Punkten (Singapur) beträchtlich.

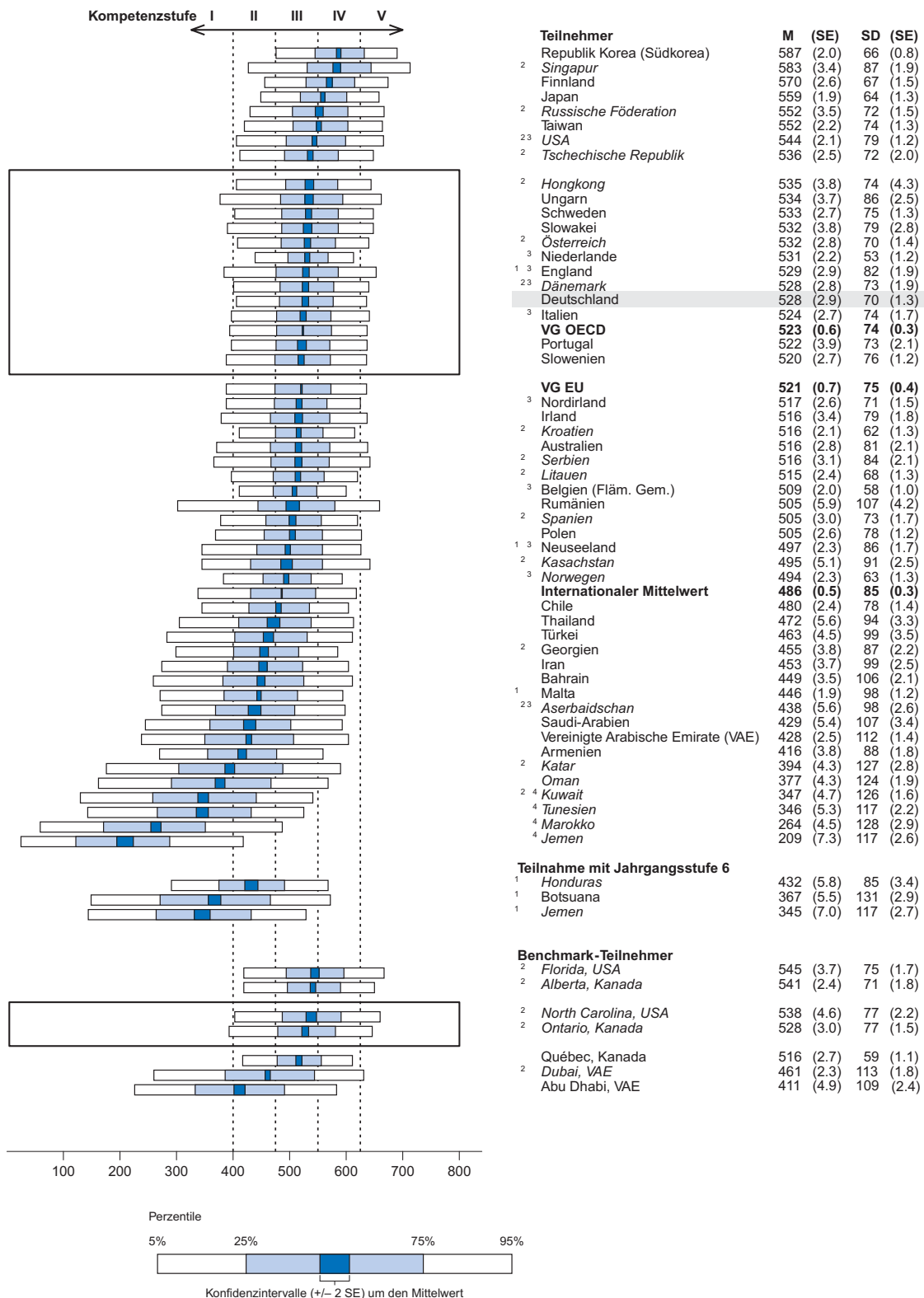
Bei der Interpretation der Ergebnisse ist zu beachten, dass die Anordnung der Staaten nach Rangplätzen nicht notwendigerweise Schlüsse auf Unterschiede in der naturwissenschaftlichen Kompetenz zwischen den Staaten zulässt. Die deskriptiven, numerischen Unterschiede können Unterschiede darstellen, die nicht statistisch bedeutsam sind. Deshalb sind in der schwarzen Umrandung von Abbildung 4.5 diejenigen Staaten zu finden, in denen das Niveau naturwissenschaftlicher Kompetenz der Schülerinnen und Schüler trotz deskriptiver Unterschiede in den Leistungsmittelwerten nicht signifikant höher oder niedriger ist als der für Deutschland ermittelte Leistungsmittelwert. Hierzu gehören unter anderem Hongkong (535), Schweden (533), die Niederlande (531) und England (529) sowie Italien (524), Portugal (522) und Slowenien (520). Signifikant höhere Leistungsmittelwerte als Deutschland weisen allerdings die Staaten oberhalb der schwarzen Umrandung in Abbildung 4.5 auf. Dazu gehören neben der Republik Korea (587) an der Leistungsspitze auch Singapur (583), Finnland (570), Japan (559), die Russische Föderation (552), Taiwan (552), die USA (544) und die Tschechische Republik (536). Betrachtet man den Mittelwert der beteiligten OECD-Staaten und den Mittelwert der beteiligten EU-Staaten, zeigt sich für Deutschland kein signifikant höherer Leistungsmittelwert als für den OECD-Durchschnitt, während ein statistisch bedeutsamer Unterschied zum EU-Durchschnitt vorliegt. Zudem sind in Deutschland die Schülerinnen und Schüler in ihrer naturwissenschaftlichen Kompetenz signifikant besser als einzelne OECD- und EU-Staaten. Hierzu gehören unter anderem Australien (516), die Flämische Gemeinschaft in Belgien (509), Spanien (505), Polen (505), Neuseeland (497), Norwegen (494) und die Türkei (463).

Streuung der naturwissenschaftlichen Kompetenz im internationalen Vergleich: Die im vorigen Abschnitt dargestellten Ergebnisse zeigen, inwieweit sich die Teilnehmerstaaten im Niveau der naturwissenschaftlichen Kompetenz voneinander unterscheiden. Dabei wird der Leistungsmittelwert für jeden Staat betrachtet, der aus dem Mittelwert der naturwissenschaftlichen Kompetenz aller Schülerinnen und Schüler in einem Staat gebildet wird. Von Interesse ist allerdings auch, wie unterschiedlich die naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern innerhalb von Staaten sind.

Daher sind in Abbildung 4.5 neben dem Mittelwert und der Standardabweichung für jeden Staat die sogenannten Perzentilbänder für das 5., 25., 75. und 95. Perzentil angegeben. Der Leistungswert etwa, der dem 75. Perzentil entspricht, sagt aus, dass 75 Prozent der Schülerinnen und Schüler einen niedrigeren oder gleich großen und damit 25 Prozent der Schülerinnen und Schüler einen größeren Leistungswert erzielt haben.

Die Breite des Perzentilbandes vom 5. bis zum 95. Perzentil repräsentiert die Streuung der Leistungswerte, über die die mittleren 90 Prozent der Schülerinnen und Schüler verfügen, die also nicht zu den jeweils 5 Prozent der Schülerinnen und Schüler mit den niedrigsten und höchsten Leistungswerten gehören. Die Breite des Perzentilbandes vom 25. bis zum 75. Perzentil gibt die Streuung der Leistungswerte an, über die die mittleren 50 Prozent der Schülerinnen und Schüler verfügen, die nicht zu den jeweils 25 Prozent der Schülerinnen und Schüler mit den niedrigsten und höchsten Leistungswerten gehören. Die Breite eines Perzentilbandes drückt somit aus, wie stark die Leistungswerte bei den mittleren 90 Prozent beziehungsweise den mittleren 50 Prozent der Schülerinnen

Abbildung 4.5: Testleistung der Schülerinnen und Schüler im internationalen Vergleich – Gesamtskala Naturwissenschaften



□ Nicht statistisch signifikant vom deutschen Mittelwert abweichende Staaten ($p > .05$).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

- 1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
- 2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
- 3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
- 4= Sehr hoher Anteil an Schülerinnen und Schülern mit nicht skalierbaren Leistungswerten.

und Schüler in einem Staat streuen (vgl. Kapitel 2). Damit erlauben die Perzentilbänder eine im Vergleich zur Angabe von Standardabweichungen differenziertere Betrachtung der Streuung der naturwissenschaftlichen Kompetenz. Auf der linken Seite von Abbildung 4.5 sind die Perzentilbänder graphisch dargestellt. Die genauen Werte für das 5., 25., 75. und 95. Perzentil können dem rechten Teil der Tabelle 4.10 entnommen werden.

Die Ergebnisse zeigen, dass es erhebliche Unterschiede in der Streuung der naturwissenschaftlichen Kompetenz zwischen den Staaten gibt. Betrachtet man die mittleren 90 Prozent der Schülerinnen und Schüler, lässt sich der Kompetenzunterschied zwischen den 5 Prozent leistungsschwächsten Schülerinnen und Schülern und den 5 Prozent leistungsstärksten Schülerinnen und Schülern beschreiben. Am geringsten ist die Streuung der naturwissenschaftlichen Kompetenz in den Niederlanden ausgeprägt. Hier beträgt der Unterschied zwischen dem 5. und dem 95. Perzentil 174 Punkte. Das bedeutet, dass die Leistungen der Schülerinnen und Schüler in den Niederlanden vergleichsweise homogen sind und weniger streuen. Die größten Streuungen der Leistungswerte finden sich in Staaten mit sehr niedrigem Niveau naturwissenschaftlicher Kompetenz. In Katar, Kuwait und Marokko betragen die Kompetenzunterschiede zwischen dem 5. und dem 95. Perzentil mehr als 400 Punkte.

In Deutschland ist die Streuung der Kompetenzen vergleichsweise gering ausgeprägt. Die 5 Prozent leistungsschwächsten Schülerinnen und Schüler weisen maximal einen Leistungswert von 406 Punkten auf, die 5 Prozent leistungsstärksten Schülerinnen und Schüler hingegen einen Wert von minimal 636 Punkten. Der Unterschied beträgt somit 230 Punkte. Ähnliche Streuungen in der naturwissenschaftlichen Kompetenz lassen sich auch in den anderen teilnehmenden OECD- und EU-Staaten finden. Eine Ausnahme bilden, wie bereits angesprochen, die Niederlande und auch die Flämische Gemeinschaft in Belgien, in denen die Kompetenzen der Schülerinnen und Schüler besonders homogen ausgeprägt sind.

Für Deutschland lässt sich insgesamt festhalten, dass die Schülerinnen und Schüler am Ende ihrer Grundschulzeit mit einem naturwissenschaftlichen Kompetenzniveau von 528 Punkten auf Rangplatz 17 von 50 Teilnehmerstaaten zu finden sind. Deutschland liegt damit im OECD- und über dem EU-Durchschnitt. Die Kompetenzunterschiede zwischen den leistungsstärksten und den leistungsschwächsten Schülerinnen und Schülern sind in Deutschland vergleichsweise gering.

Zur Bedeutung der curricular nicht validen Aufgaben für die Ergebnisse: Bei den Analysen zur curricularen Validität der in TIMSS 2011 eingesetzten Testaufgaben im Bereich Naturwissenschaften (siehe Abschnitt 4.2) hatte sich gezeigt, dass 23 Prozent der Aufgaben als nicht curricular valide eingeschätzt wurden. Bei diesen Aufgaben ist davon auszugehen, dass sie Inhalte thematisieren, die nicht in den Lehrplänen zum Sachunterricht in Deutschland vorgesehen sind. Um zu prüfen, ob diese Tatsache einen Einfluss auf die Ergebnisse zum Niveau naturwissenschaftlicher Kompetenzen im internationalen Vergleich hat, wurden in einer zusätzlichen Analyse die nicht curricular validen Aufgaben ausgeschlossen (siehe Tabelle 4.9).

Für diese Analyse wurden Neuskalierungen der Daten vorgenommen, da der Veröffentlichungstermin internationaler Itemparameter und Rahmendaten zum internationalen Hintergrundmodell, die für eine genaue Replikation der Ergebnisse aus TIMSS 2011 erforderlich gewesen wären, nach der Drucklegung für die-

Tabelle 4.9: Testwerte und Rangplätze im Bereich Naturwissenschaften – Skala mit für Deutschland curricular validen Testaufgaben

Teilnehmer	Modell I (int. Skalierungsmodell [n = 168 Testaufgaben])					Modell II (Replikation Modell I; Modifikation: 1PL; ohne Hintergrundmodell)					Modell III (Replikation Modell II; Modifikation: nur mit für Deutschland curricular validen Testaufgaben [n = 129])						
	M	(SE)	SD	(SE)	Rang _{INT}	M	(SE)	SD	(SE)	Rang _{1PL}	Δ _{Rang(INT-1PL)}	M	(SE)	SD	(SE)	Rang _{CV}	Δ _{Rang(1PL-CV)}
Republik Korea (Südkorea)	587	(2.0)	66	(0.8)	1	594	(1.5)	77	(1.5)	2	-1	593	(1.8)	76	(1.5)	2	0
² <i>Singapur</i>	583	(3.4)	87	(1.9)	2	595	(2.9)	89	(1.4)	1	1	598	(2.7)	88	(1.3)	1	0
Finland	570	(2.6)	67	(1.5)	3	582	(2.2)	78	(1.5)	3	0	578	(2.0)	79	(1.2)	3	0
Japan	559	(1.9)	64	(1.3)	4	572	(1.6)	76	(0.9)	4	0	575	(1.7)	79	(1.1)	4	0
² <i>Russische Föderation</i>	552	(3.5)	72	(1.5)	5	567	(3.2)	80	(1.5)	5	0	564	(3.2)	80	(1.3)	6	-1
Taiwan	552	(2.2)	74	(1.3)	6	567	(1.7)	80	(1.3)	6	0	568	(1.7)	81	(1.5)	5	1
^{2 3} <i>USA</i>	544	(2.1)	79	(1.2)	7	560	(1.8)	86	(1.1)	7	0	556	(2.0)	84	(0.9)	7	0
² <i>Tschechische Republik</i>	536	(2.5)	72	(2.0)	8	554	(2.3)	80	(1.3)	8	0	552	(2.1)	80	(1.0)	9	-1
² <i>Hongkong</i>	535	(3.8)	74	(4.3)	9	553	(2.9)	80	(2.2)	10	-1	548	(2.8)	79	(2.0)	15	-5
Ungarn	534	(3.7)	86	(2.5)	10	554	(3.2)	88	(1.7)	9	1	551	(2.9)	88	(1.5)	10	-1
Schweden	533	(2.7)	75	(1.3)	11	551	(2.2)	83	(1.2)	12	-1	552	(2.0)	84	(1.5)	8	4
Slowakei	532	(3.8)	79	(2.8)	12	552	(3.1)	84	(1.9)	11	1	550	(2.9)	84	(1.8)	12	-1
² <i>Österreich</i>	532	(2.8)	70	(1.4)	13	549	(2.6)	80	(1.4)	13	0	551	(2.6)	80	(1.2)	11	2
³ <i>Niederlande</i>	531	(2.2)	53	(1.2)	14	546	(1.9)	71	(1.5)	16	-2	546	(2.1)	73	(1.4)	17	-1
^{1 3} <i>England</i>	529	(2.9)	82	(1.9)	15	548	(2.8)	86	(1.7)	14	1	549	(2.9)	86	(1.7)	13	1
^{2 3} <i>Dänemark</i>	528	(2.8)	73	(1.9)	16	546	(2.2)	82	(1.3)	15	1	547	(2.2)	82	(1.7)	16	-1
Deutschland	528	(2.9)	70	(1.3)	17	546	(2.4)	80	(1.2)	17	0	548	(2.3)	81	(1.1)	14	3
³ <i>Italien</i>	524	(2.7)	74	(1.7)	18	544	(2.4)	82	(1.3)	18	0	543	(2.4)	81	(1.3)	18	0
Portugal	522	(3.9)	73	(2.1)	19	543	(3.6)	80	(1.6)	19	0	542	(3.6)	81	(2.0)	20	-1
Slowenien	520	(2.7)	76	(1.2)	20	542	(2.5)	84	(1.3)	20	0	543	(2.2)	84	(1.5)	19	1
³ <i>Nordirland</i>	517	(2.6)	71	(1.5)	21	537	(2.5)	80	(1.4)	21	0	537	(2.6)	81	(1.5)	24	-3
<i>Irland</i>	516	(3.4)	79	(1.8)	22	537	(2.8)	84	(1.6)	24	-2	538	(3.1)	85	(1.5)	22	2
² <i>Kroatien</i>	516	(2.1)	62	(1.3)	23	537	(2.1)	75	(1.2)	22	1	538	(1.8)	76	(1.1)	23	-1
<i>Australien</i>	516	(2.8)	81	(2.1)	24	536	(2.6)	85	(1.6)	25	-1	536	(2.9)	85	(2.1)	25	0
² <i>Serbien</i>	516	(3.1)	84	(2.1)	25	537	(2.8)	87	(1.8)	23	2	541	(3.0)	86	(1.6)	21	2
² <i>Litauen</i>	515	(2.4)	68	(1.3)	26	536	(2.3)	78	(1.2)	26	0	535	(2.4)	79	(1.1)	26	0
³ <i>Belgien (Fläm. Gem.)</i>	509	(2.0)	58	(1.0)	27	529	(1.8)	72	(1.2)	29	-2	529	(2.1)	74	(1.4)	28	1
<i>Rumänien</i>	505	(5.9)	107	(4.2)	28	532	(4.6)	98	(2.5)	27	1	533	(4.6)	97	(2.4)	27	0
² <i>Spanien</i>	505	(3.0)	73	(1.7)	29	529	(2.7)	80	(1.4)	28	1	525	(2.9)	81	(1.6)	30	-2
<i>Polen</i>	505	(2.6)	78	(1.2)	30	527	(2.5)	84	(1.0)	30	0	526	(2.5)	82	(1.2)	29	1
^{1 3} <i>Neuseeland</i>	497	(2.3)	86	(1.7)	31	521	(2.2)	86	(1.4)	32	-1	524	(2.3)	88	(1.2)	31	1
² <i>Kasachstan</i>	495	(5.1)	91	(2.5)	32	522	(4.4)	90	(1.9)	31	1	522	(4.0)	89	(1.7)	32	-1
³ <i>Norwegen</i>	494	(2.3)	63	(1.3)	33	518	(2.1)	75	(1.6)	33	0	516	(2.3)	77	(1.6)	33	0
<i>Chile</i>	480	(2.4)	78	(1.4)	34	510	(2.1)	82	(1.0)	34	0	508	(1.9)	82	(1.4)	34	0
<i>Thailand</i>	472	(5.6)	94	(3.3)	35	506	(4.1)	87	(2.1)	35	0	506	(3.7)	88	(1.9)	35	0
<i>Türkei</i>	463	(4.5)	99	(3.5)	36	501	(3.2)	87	(1.5)	36	0	504	(2.8)	87	(1.3)	36	0
² <i>Georgien</i>	455	(3.8)	87	(2.2)	37	493	(2.7)	81	(1.7)	39	-2	494	(2.7)	82	(1.8)	38	1
<i>Iran</i>	453	(3.7)	99	(2.5)	38	495	(2.8)	88	(1.3)	37	1	496	(2.8)	90	(1.4)	37	0
<i>Bahrain</i>	449	(3.5)	106	(2.1)	39	494	(2.5)	91	(1.5)	38	1	494	(2.6)	92	(1.6)	39	-1
¹ <i>Malta</i>	446	(1.9)	98	(1.2)	40	487	(2.4)	86	(1.4)	41	-1	490	(1.9)	87	(1.3)	40	1
^{2 3} <i>Aserbaidschan</i>	438	(5.6)	98	(2.6)	41	488	(4.0)	85	(1.8)	40	1	489	(4.0)	87	(1.5)	41	-1
<i>Saudi-Arabien</i>	429	(5.4)	107	(3.4)	42	480	(3.9)	88	(2.3)	42	0	478	(3.9)	89	(2.1)	43	-1
<i>Vereinigte Arabische Emirate (VAE)</i>	428	(2.5)	112	(1.4)	43	480	(1.8)	91	(0.9)	43	0	479	(2.0)	91	(0.9)	42	1
<i>Armenien</i>	416	(3.8)	88	(1.8)	44	466	(2.9)	81	(1.4)	44	0	470	(2.9)	83	(1.5)	44	0
² <i>Katar</i>	394	(4.3)	127	(2.8)	45	464	(2.9)	93	(1.7)	45	0	463	(2.6)	93	(2.0)	45	0
<i>Oman</i>	377	(4.3)	124	(1.9)	46	453	(2.7)	89	(1.1)	46	0	452	(2.5)	89	(1.4)	46	0
^{2 4} <i>Kuwait</i>	347	(4.7)	126	(1.6)	47	436	(2.5)	85	(1.3)	47	0	434	(2.5)	85	(1.5)	47	0
⁴ <i>Tunesien</i>	346	(5.3)	117	(2.2)	48	428	(3.1)	84	(1.7)	48	0	434	(3.2)	85	(1.4)	48	0
⁴ <i>Marokko</i>	264	(4.5)	128	(2.9)	49	405	(2.5)	76	(1.5)	49	0	409	(2.5)	77	(1.4)	49	0
⁴ <i>Jemen</i>	209	(7.3)	117	(2.6)	50	385	(2.5)	71	(1.4)	50	0	386	(2.5)	71	(1.4)	50	0

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 4 = Sehr hoher Anteil an Schülerinnen und Schülern mit nicht skalierbaren Leistungswerten.

sen Band lag. Die Neuskalierungen wurden mit einem sogenannten einparametrischen IRT-Modell durchgeführt (vgl. Kapitel 2: *Weitere Skalierungsmodelle*). In einer ersten Analyse wurden wie auch bei der internationalen Skalierung 168 der 175 Testaufgaben eingeschlossen (Modell II). In Modell III wurden dann (unter Verwendung derselben Skalierungsmethode) nur die für Deutschland als curricular valide eingeschätzten Aufgaben zugrunde gelegt. Durch einen Vergleich der Ergebnisse von Modell II und Modell III kann abgeschätzt werden, welchen Einfluss die nicht curricular validen Aufgaben für die Ergebnisse in Deutschland haben.

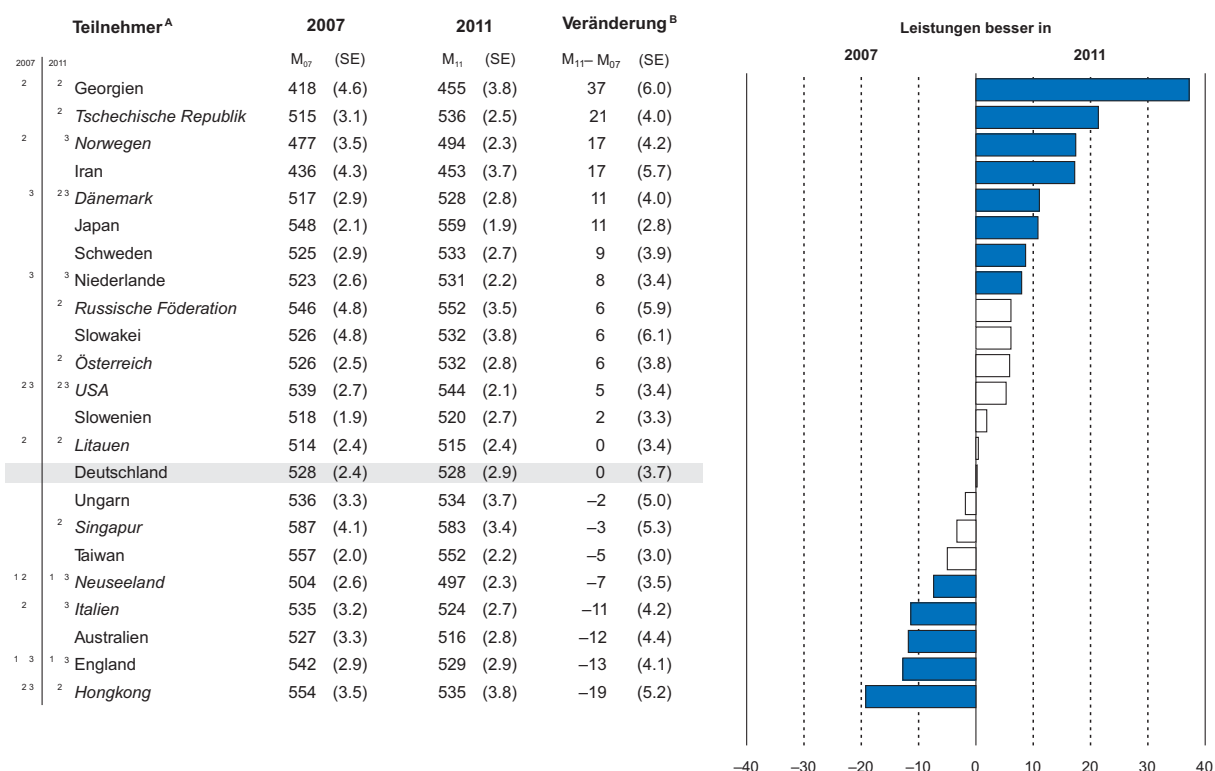
Es zeigt sich, dass sich die in Modell II und die in Modell III ermittelten Ergebnisse für Deutschland nur geringfügig unterscheiden. Das Niveau naturwissenschaftlicher Kompetenz beträgt in Modell II 546 Punkte und in Modell III 548 Punkte. In Modell II liegt Deutschland auf dem Rangplatz 17 (von 50), in Modell III auf Rangplatz 14. Auch in den anderen Staaten zeigen sich nur geringe Rangplatzunterschiede zwischen den beiden Modellen.

Insgesamt betrachtet scheint somit die Tatsache, dass 23 Prozent der Testaufgaben für Deutschland nicht curricular valide sind, nur geringen Einfluss auf die Ergebnisse Deutschlands im internationalen Vergleich zu haben.

Unterschiede in der naturwissenschaftlichen Kompetenz zwischen TIMSS 2007 und 2011: Da Deutschland 2011 nach 2007 das zweite Mal an der TIMSS-Grundschuluntersuchung teilgenommen hat, können erstmalig Unterschiede in der naturwissenschaftlichen Kompetenz von Kindern im Vergleich zu einer vorherigen Erhebung dargestellt werden. In Abbildung 4.6 sind die nationalen Mittelwerte und die zugehörigen Standardfehler der naturwissenschaftlichen Kompetenzen bei TIMSS 2007 und TIMSS 2011 für die Staaten gegenübergestellt, die an beiden Studienzyklen teilgenommen haben (ohne Benchmark-Teilnehmer, Ausnahme siehe Fußnote A in der Abbildung). Außerdem ist für jeden dieser Staaten die Differenz der Mittelwerte aus 2011 und 2007 mit dem dazugehörigen Standardfehler dargestellt. Es zeigt sich, dass sich das Niveau der naturwissenschaftlichen Kompetenz in einigen Staaten zwischen 2007 und 2011 deutlich unterscheidet, in anderen Staaten hingegen weniger oder gar nicht. Die Staaten, die in 2011 ein signifikant höheres Niveau in der naturwissenschaftlichen Kompetenz erzielen als bei TIMSS 2007, sind Georgien, die Tschechische Republik, Norwegen, Iran, Dänemark, Japan, Schweden und die Niederlande. In einer zweiten Gruppe von Staaten zeigen sich keine signifikanten Unterschiede zum Kompetenzniveau in 2007. Zu dieser Gruppe gehören neben Deutschland die Russische Föderation, die Slowakei, Österreich, die USA, Slowenien, Litauen, Ungarn, Singapur und Taiwan. Das mittlere Niveau der naturwissenschaftlichen Kompetenz in Deutschland entspricht 2011 mit 528 Punkten genau dem Niveau in 2007. In einer dritten Gruppe befinden sich Staaten, die 2011 ein niedrigeres naturwissenschaftliches Kompetenzniveau als 2007 aufweisen. Zu dieser Gruppe gehören Neuseeland, Italien, Australien, England und Hongkong.

Unterschiede in der Streuung der Leistungswerte zwischen TIMSS 2007 und 2011: In Tabelle 4.10 sind neben den Angaben zur Streuung der Leistungswerte in 2011, auf die bereits eingegangen wurde, auch die entsprechenden Werte für 2007 dargestellt. So können Unterschiede in der Streuung der Leistungswerte zwischen 2007 und 2011 beschrieben werden. Dargestellt sind das 5., 25., 75. und das 95. Perzentil für Staaten, die sowohl 2007 als auch 2011 an TIMSS mit

Abbildung 4.6: Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 – Gesamtskala Naturwissenschaften



■ Statistisch signifikante Unterschiede ($p < .05$).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3= Die Teilnahmequoten auf Schul- und/oder Schülersebene erreichen nicht die internationalen Vorgaben.

A= Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

B= Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

der vierten Jahrgangsstufe teilgenommen haben. Die Staaten sind hier in alphabetischer Reihenfolge aufgeführt.

Für Deutschland zeigen die Ergebnisse, dass die Streuung der naturwissenschaftlichen Kompetenz der Schülerinnen und Schüler 2011 in etwa so stark ausgeprägt ist wie bei TIMSS 2007. Betrachtet man das 5. Perzentil, also den Leistungswert, den die 5 Prozent leistungsschwächsten Schülerinnen und Schüler maximal erreichen, so liegt dieser Wert 2011 bei 406 Punkten und 2007 bei 393 Punkten. Dieser geringe deskriptive Unterschied ist nicht statistisch signifikant. Auch die Werte für das 95. Perzentil unterscheiden sich zwischen 2011 (636 Punkte) und 2007 (647 Punkte) nur deskriptiv und nicht signifikant. Dementsprechend ist die Streuung der Leistungswerte, über die die mittleren 90 Prozent der Schülerinnen und Schüler verfügen, 2011 nur geringfügig niedriger als 2007: 2011 beträgt der Abstand zwischen dem 5. und dem 95. Perzentil 230 Punkte, 2007 betrug er 254 Punkte. Es ist also eine geringfügige Homogenisierung der Kompetenzen der Schülerinnen und Schüler im Vergleich zu TIMSS 2007 zu beobachten.

Für Deutschland zeigt sich also, dass sich das Niveau naturwissenschaftlicher Kompetenz im Vergleich zu 2007 nicht geändert hat und dass die Streuung der Kompetenzen der Schülerinnen und Schüler 2011 ähnlich stark ausgeprägt ist

Tabelle 4.10: Naturwissenschaftliche Kompetenz nach Perzentilen im Vergleich: TIMSS 2007 und TIMSS 2011 – Gesamtskala Naturwissenschaften

2007	2011	Teilnehmer ^A	TIMSS 2007				TIMSS 2011				
			Perzentile				Perzentile				
			5 (SE)	25 (SE)	75 (SE)	95 (SE)	5 (SE)	25 (SE)	75 (SE)	95 (SE)	
		Australien	384 (9.0)	478 (4.4)	583 (2.6)	651 (4.8)	371 (9.0)	466– (4.3)	571– (3.0)	638 (5.3)	
3	2,3	Dänemark	383 (6.5)	468 (2.6)	570 (5.1)	636 (4.7)	401 (7.0)	483+ (3.7)	578 (3.1)	640 (4.3)	
		Deutschland	393 (7.6)	479 (4.2)	582 (1.6)	647 (6.0)	406 (7.1)	482 (3.6)	577 (2.8)	636 (4.9)	
1	3	England	403 (6.4)	491 (6.0)	596 (3.6)	666 (6.3)	384– (6.3)	476 (5.1)	586 (3.5)	653 (3.0)	
2	2	Georgien	273 (8.6)	361 (5.2)	477 (4.4)	552 (7.9)	299+ (6.5)	401+ (7.2)	516+ (3.7)	585+ (2.9)	
2,3	2	Hongkong	437 (4.6)	511 (3.7)	601 (3.6)	659 (4.2)	406 (16.3)	493– (3.3)	585– (2.7)	644– (5.1)	
		Iran	267 (7.1)	371 (6.5)	506 (5.3)	587 (3.3)	274 (6.7)	390+ (5.6)	523+ (3.7)	604+ (6.0)	
2	3	Italien	395 (5.0)	484 (5.2)	590 (2.9)	664 (4.4)	397 (8.3)	477 (3.4)	573– (3.1)	641– (3.3)	
		Japan	428 (7.0)	505 (3.1)	595 (1.4)	655 (3.2)	449+ (4.1)	519+ (2.6)	601+ (1.9)	658 (2.8)	
2	2	Litauen	401 (5.2)	473 (2.8)	559 (3.2)	615 (4.2)	397 (6.0)	471 (2.9)	561 (2.0)	620 (5.0)	
1,2	1	Neuseeland	344 (4.9)	447 (3.7)	568 (2.9)	643 (2.6)	345 (6.9)	442 (3.2)	558– (2.3)	626– (2.8)	
3	3	Niederlande	421 (6.3)	484 (3.6)	565 (3.9)	617 (5.0)	439 (8.2)	497+ (2.2)	568 (1.8)	613 (2.2)	
2	3	Norwegen	343 (8.8)	429 (4.0)	530 (3.3)	593 (5.2)	383+ (6.1)	453+ (2.4)	538 (3.7)	593 (3.7)	
	2	Österreich	388 (4.4)	477 (3.3)	580 (2.5)	644 (3.5)	408+ (4.8)	485 (4.2)	581 (2.0)	640 (2.5)	
2	2	Russische Föderation	407 (7.6)	495 (6.6)	601 (4.2)	672 (8.8)	430+ (5.2)	505 (3.6)	603 (2.9)	667 (4.8)	
		Schweden	400 (3.3)	478 (4.0)	575 (3.0)	642 (4.0)	403 (6.2)	486 (3.1)	586+ (2.6)	648 (4.2)	
2	2	Singapur	418 (7.0)	531 (6.0)	652 (4.4)	727 (3.9)	427 (6.7)	531 (5.8)	644 (4.1)	713– (4.0)	
		Slowakei	376(13.9)	476 (4.7)	584 (2.8)	652 (6.9)	390 (13.9)	486 (4.3)	586 (2.7)	648 (5.4)	
		Slowenien	383 (6.2)	471 (3.7)	571 (2.8)	634 (2.7)	388 (4.9)	474 (2.8)	572 (4.2)	636 (4.1)	
		Taiwan	423 (6.5)	508 (2.0)	609 (2.4)	679 (3.0)	420 (6.5)	506 (4.3)	603 (2.6)	664– (4.2)	
	2	Tschechische Republik	386 (8.9)	467 (4.4)	567 (3.8)	635 (4.3)	412+ (9.3)	491+ (2.3)	586+ (3.8)	648+ (3.8)	
		Ungarn	383 (9.4)	485 (3.8)	595 (3.4)	661 (4.7)	377 (8.7)	484 (6.4)	594 (4.0)	662 (4.2)	
2,3	2,3	USA	392 (4.9)	484 (3.2)	597 (3.4)	668 (3.2)	406+ (3.9)	494+ (2.4)	599 (2.5)	666 (2.3)	

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerenebene erreichen nicht die internationalen Vorgaben.

A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

+ = Mittelwert für 2011 signifikant höher als 2007 ($p < .05$).

– = Mittelwert für 2011 signifikant niedriger als 2007 ($p < .05$).

wie 2007. Es ist lediglich eine schwache Homogenisierung der Leistungen zu beobachten.

5.2 Kompetenzstufen

In Abschnitt 4.3 wurden fünf Stufen naturwissenschaftlicher Kompetenz unterschieden. Die fünf Stufen lassen sich anhand der Anforderungen, die der Bearbeitung der Testaufgaben auf einer Kompetenzstufe zugrunde liegen, inhaltlich beschreiben. Diese inhaltliche Charakterisierung der fünf Kompetenzstufen findet sich in Tabelle 4.8 (Abschnitt 4.3). Für die einzelnen Kompetenzstufen können die Anteile der Schülerinnen und Schüler bestimmt werden. Auf diese Weise lässt sich beschreiben, wie groß die Gruppe der Schülerinnen und Schüler in einem Staat ist, die eine hohe naturwissenschaftliche Kompetenz aufweisen (Kompetenzstufen IV und V), und wie groß die Gruppe der Schülerinnen und Schüler ist, die nur über eine niedrige naturwissenschaftliche Kompetenz verfügen, die es ihnen nur in Ansätzen erlaubt, dieses Wissen produktiv einzusetzen (Kompetenzstufen I und II).

In diesem Abschnitt werden zunächst die Anteile der Schülerinnen und Schüler auf den fünf Kompetenzstufen im internationalen Vergleich berichtet. Anschließend werden Unterschiede in den Anteilen der Schülerinnen und Schüler auf den Kompetenzstufen zwischen 2007 und 2011 dargestellt.

Verteilung der Schülerinnen und Schüler auf die fünf Stufen naturwissenschaftlicher Kompetenz im internationalen Vergleich: In Abbildung 4.7 sind für jeden der hier ausgewählten Teilnehmerstaaten die Anteile der Schülerinnen und Schüler dargestellt, die den fünf Kompetenzstufen zugeordnet wurden. Die Staaten sind entsprechend des Anteils von Schülerinnen und Schülern auf Kompetenzstufe V in einer Rangreihe angeordnet.

Es wird deutlich, dass Singapur und die Republik Korea hier eine besondere Position einnehmen. Der Anteil der Schülerinnen und Schüler, die in diesen Staaten der Kompetenzstufe V zugeordnet wurden, liegt etwa doppelt so hoch (in Singapur 33 %, in der Republik Korea 29 %) wie in den übrigen Staaten (mit Ausnahme von Finnland). Die Republik Korea sticht außerdem dadurch hervor, dass sich hier im Vergleich zu den anderen Staaten der geringste Anteil von Schülerinnen und Schülern auf den unteren Kompetenzstufen I und II befindet (5 %). Demgegenüber fallen Staaten wie die Türkei, Chile und Malta mit sehr großen Anteilen von Schülerinnen und Schülern auf den unteren Kompetenzstufen auf.

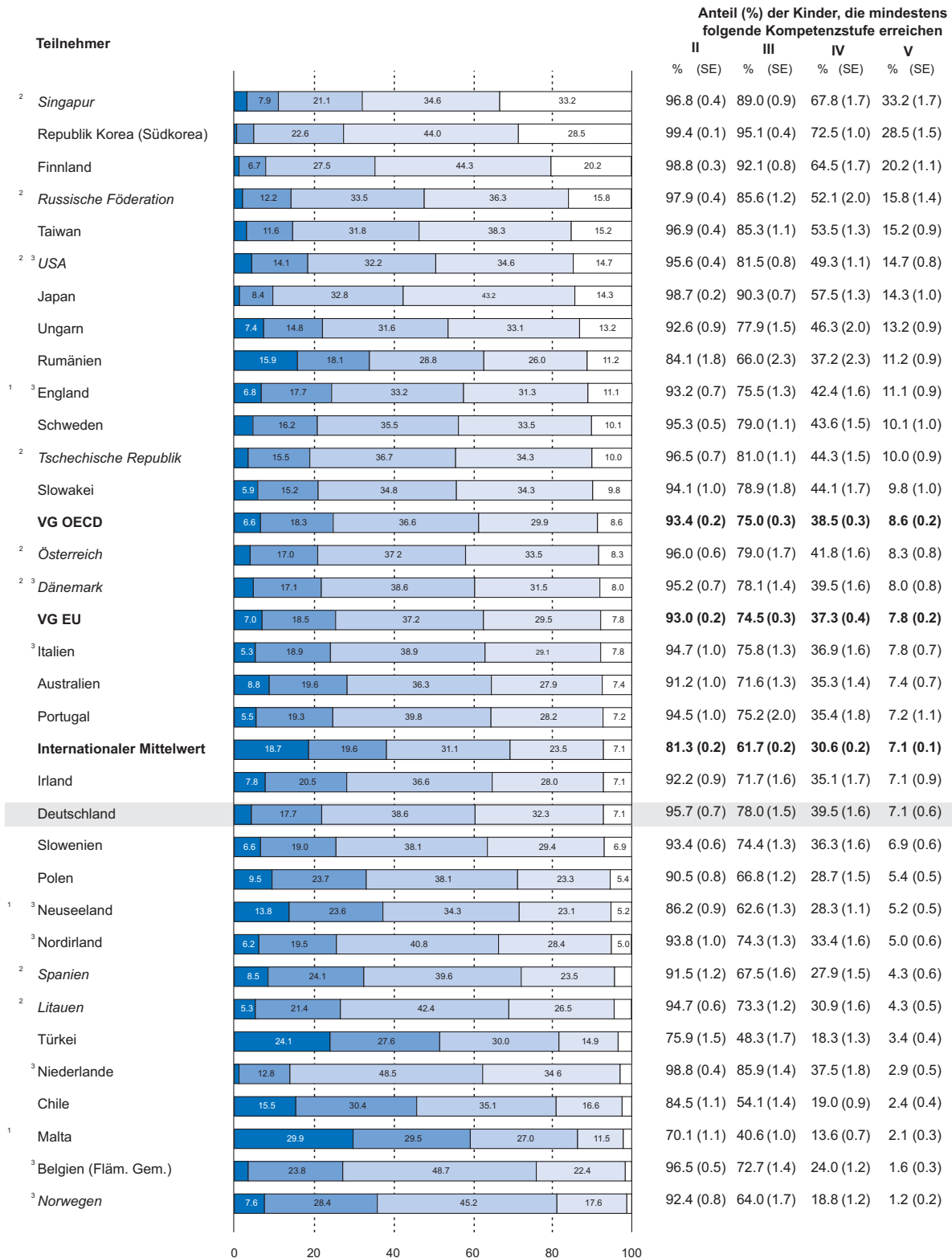
In Deutschland befinden sich 7 Prozent der Schülerinnen und Schüler auf Kompetenzstufe V. Diese Schülerinnen und Schüler verstehen und begründen bereits am Ende der Grundschulzeit naturwissenschaftliche Zusammenhänge, sie interpretieren einfache naturwissenschaftliche Experimente und ziehen Schlussfolgerungen daraus. Diese Gruppe von Schülerinnen und Schülern verfügt somit über ausgezeichnete Voraussetzungen für eine weiterführende Beschäftigung mit den Naturwissenschaften. Weitere 32 Prozent der Schülerinnen und Schüler befinden sich auf Kompetenzstufe IV. Sie verfügen über naturwissenschaftliche Konzepte, mit denen sie Alltagsphänomene und Sachverhalte beschreiben können. Auch sie besitzen ausreichendes Basiswissen, welches zu einer erfolgreichen Teilnahme an weiterführendem naturwissenschaftlichen Unterricht beitragen kann. Demgegenüber sind die Voraussetzungen der Gruppe von Schülerinnen und Schülern in Deutschland, die sich auf den Kompetenzstufen I und II befinden, als schlecht zu bezeichnen. Sie können zwar elementares Faktenwissen abrufen, allerdings fehlt es ihnen an einem grundlegenden naturwissenschaftlichen Verständnis. Besonders kritisch ist der Anteil der Schülerinnen und Schüler (4 %) einzuschätzen, die am Ende ihrer Grundschulzeit nicht in der Lage sind, einfache Aufgaben zu elementarem naturwissenschaftlichen Faktenwissen sicher zu lösen (Kompetenzstufe I).

Die Anteile der Schülerinnen und Schüler auf den einzelnen Kompetenzstufen liegen in Deutschland allerdings in etwa in der Größenordnung der Anteile im OECD- und im EU-Durchschnitt.

Unterschiede in den Anteilen der Schülerinnen und Schüler auf den Kompetenzstufen zwischen TIMSS 2007 und 2011: Tabelle 4.11 zeigt, wie sich die prozentualen Anteile der Schülerinnen und Schüler auf den fünf Kompetenzstufen in den einzelnen Staaten zwischen TIMSS 2007 und 2011 unterscheiden. In der Tabelle sind nur Staaten aufgeführt, die sowohl an TIMSS 2007 als auch an TIMSS 2011 teilgenommen haben. Die Staaten sind in alphabetischer Reihenfolge angeordnet.

Für Deutschland zeigen die Ergebnisse, dass sich der Anteil der Schülerinnen und Schüler auf den unteren Kompetenzstufen I und II nur minimal und nicht statistisch signifikant verändert hat. 2007 betrug der Anteil der Kinder auf Kompetenzstufe I 5,8 Prozent, 2011 4,3 Prozent. 2007 befanden sich 17,9 Prozent auf Kompetenzstufe II, 2011 17,7 Prozent. Im Vergleich zu TIMSS 2007 ist der Anteil der Kinder auf Kompetenzstufe III 2011 geringfügig, aber signifikant

Abbildung 4.7: Prozentuale Verteilung der Schülerinnen und Schüler auf die fünf Kompetenzstufen (Naturwissenschaften) im internationalen Vergleich



□ % der Schülerinnen und Schüler, die genau Kompetenzstufe V erreichen
 □ % der Schülerinnen und Schüler, die genau Kompetenzstufe IV erreichen
 □ % der Schülerinnen und Schüler, die genau Kompetenzstufe III erreichen
 □ % der Schülerinnen und Schüler, die genau Kompetenzstufe II erreichen
 ■ % der Schülerinnen und Schüler, die genau Kompetenzstufe I erreichen

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

Tabelle 4.11: Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007 und in TIMSS 2011 (Angaben in Prozent)

Teilnehmer ^A	TIMSS 2007					TIMSS 2011					
	Kompetenzstufen					Kompetenzstufen					
	I	II	III	IV	V	I	II	III	IV	V	
2007	2011	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	
Australien		6.7 (0.8)	16.9 (1.1)	35.5 (1.3)	30.6 (2.1)	10.2 (0.7)	8.8 (1.0)	19.6 (1.0)	36.3 (1.1)	27.9 (1.2)	7.4 – (0.7)
³ Dänemark		7.1 (0.8)	20.8 (1.0)	37.6 (1.6)	27.7 (1.1)	6.9 (0.8)	4.8 – (0.7)	17.1 – (1.1)	38.6 (1.2)	31.5 + (1.3)	8.0 (0.8)
Deutschland		5.8 (0.6)	17.9 (0.9)	35.4 (0.7)	31.3 (1.1)	9.6 (0.7)	4.3 (0.7)	17.7 (1.3)	38.6 + (1.2)	32.3 (1.3)	7.1 – (0.6)
^{1 3} England		4.7 (0.6)	14.7 (0.8)	32.7 (1.2)	33.5 (1.1)	14.2 (1.2)	6.8 + (0.7)	17.7 + (0.9)	33.2 (1.3)	31.3 (1.3)	11.1 – (0.9)
² Georgien		41.1 (2.1)	32.9 (1.5)	20.6 (1.7)	4.9 (0.7)	0.5 (0.2)	24.8 – (1.6)	31.4 (1.2)	31.3 + (1.3)	11.2 + (1.1)	1.4 (0.4)
^{2 3} Hongkong		2.0 (0.4)	10.1 (1.0)	33.1 (1.4)	40.7 (1.5)	14.1 (1.4)	4.5 (1.2)	13.9 + (1.1)	37.0 (1.6)	35.5 – (1.7)	9.1 – (0.9)
Iran		34.7 (1.9)	29.0 (1.1)	24.6 (1.2)	10.1 (0.9)	1.6 (0.3)	27.9 – (1.5)	27.9 (1.1)	28.0 + (1.1)	13.4 + (0.9)	2.8 + (0.4)
^{2 3} Italien		5.5 (0.7)	16.1 (0.9)	34.5 (1.0)	30.9 (1.2)	13.0 (1.0)	5.3 (1.0)	18.9 (1.2)	38.9 + (1.4)	29.1 (1.4)	7.8 – (0.7)
Japan		2.7 (0.4)	11.3 (0.9)	35.2 (0.9)	38.4 (0.9)	12.4 (1.0)	1.3 – (0.2)	8.4 – (0.7)	32.8 (1.2)	43.2 + (1.2)	14.3 (1.0)
² Litauen		4.9 (0.6)	21.2 (1.2)	43.6 (1.3)	27.0 (1.3)	3.3 (0.4)	5.3 (0.6)	21.4 (1.0)	42.4 (1.2)	26.5 (1.4)	4.3 (0.5)
^{1 3} Neuseeland		13.4 (1.0)	21.7 (0.8)	32.7 (0.9)	24.5 (0.9)	7.8 (0.5)	13.8 (0.9)	23.6 (0.9)	34.3 (1.2)	23.1 (0.9)	5.2 – (0.5)
³ Niederlande		2.5 (0.5)	18.2 (1.3)	45.4 (1.4)	30.2 (1.5)	3.7 (0.8)	1.2 (0.4)	12.8 – (1.3)	48.5 (1.2)	34.6 (1.7)	2.9 (0.5)
² Norwegen		16.0 (1.4)	30.3 (1.2)	37.1 (1.1)	15.1 (1.3)	1.5 (0.4)	7.6 – (0.8)	28.4 (1.3)	45.2 + (1.3)	17.6 (1.2)	1.2 (0.2)
² Österreich		6.5 (0.6)	17.8 (1.1)	36.2 (1.1)	30.7 (1.2)	8.7 (0.7)	4.0 – (0.6)	17.0 (1.5)	37.2 (1.4)	33.5 (1.5)	8.3 (0.8)
² Russische Föderation		4.3 (0.9)	13.9 (1.3)	32.3 (1.3)	33.7 (1.3)	15.8 (1.9)	2.1 – (0.4)	12.2 (0.9)	33.5 (1.4)	36.3 (1.4)	15.8 (1.4)
Schweden		5.0 (0.6)	19.0 (1.1)	38.5 (1.0)	29.4 (1.3)	8.0 (0.6)	4.7 (0.5)	16.2 – (0.9)	35.5 – (1.0)	33.5 + (1.0)	10.1 (1.0)
² Singapur		3.8 (0.5)	7.9 (0.8)	20.1 (1.2)	32.2 (1.4)	35.9 (1.9)	3.2 (0.4)	7.9 (0.8)	21.1 (1.1)	34.6 (1.1)	33.2 (1.7)
Slowakei		7.5 (1.3)	17.0 (1.2)	33.7 (1.1)	31.1 (1.6)	10.6 (0.8)	5.9 (1.0)	15.2 (1.2)	34.8 (1.4)	34.3 (1.1)	9.8 (1.0)
Slowenien		7.0 (0.6)	19.5 (0.8)	38.0 (1.2)	29.0 (1.0)	6.5 (0.6)	6.6 (0.6)	19.0 (1.2)	38.1 (1.0)	29.4 (1.5)	6.9 (0.6)
Taiwan		3.0 (0.4)	11.0 (0.6)	30.9 (1.1)	36.4 (1.3)	18.7 (1.0)	3.1 (0.4)	11.6 (0.9)	31.8 (1.0)	38.3 (1.2)	15.2 – (0.9)
² Tschechische Republik		7.0 (0.8)	21.5 (1.1)	38.5 (1.5)	26.1 (1.7)	7.0 (0.7)	3.5 – (0.7)	15.5 – (1.1)	36.7 (1.4)	34.3 + (1.4)	10.0 + (0.9)
Ungarn		6.6 (0.8)	15.1 (1.3)	31.1 (1.5)	33.7 (1.5)	13.4 (1.0)	7.4 (0.9)	14.8 (1.1)	31.6 (1.2)	33.1 (1.6)	13.2 (0.9)
^{2 3} USA		6.0 (0.6)	16.1 (0.7)	31.2 (0.9)	31.8 (0.9)	15.0 (0.9)	4.4 – (0.4)	14.1 – (0.6)	32.2 (0.7)	34.6 + (0.8)	14.7 (0.8)

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

+ = Mittelwert für 2011 signifikant höher als 2007 ($p < .05$).

– = Mittelwert für 2011 signifikant niedriger als 2007 ($p < .05$).

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

größer (2007: 35.4%, 2011: 38.6%). Während sich der Anteil der Schülerinnen und Schüler auf Kompetenzstufe IV zwischen 2011 und 2007 nicht signifikant unterscheidet, ist der Anteil der Kinder auf Kompetenzstufe V 2011 geringfügig, aber signifikant kleiner (2007: 9.6%, 2011: 7.1%). Es zeigt sich also eine geringfügige Verschiebung der Verteilung in die Mitte: Während der Anteil der Kinder auf Kompetenzstufe V geringfügig kleiner geworden ist, ist der Anteil der Schülerinnen und Schüler auf Kompetenzstufe III geringfügig größer geworden.

Auffällig ist, dass in Georgien der Anteil der Kinder auf der Kompetenzstufe I 2011 um 16.3 Prozentpunkte niedriger liegt als 2007. Auch Norwegen und Iran zeigen im Vergleich zu 2007 einen deutlich geringeren Anteil von Kindern auf Kompetenzstufe I. Allerdings handelt es sich bei diesen Staaten auch um diejenigen, die unter den aufgeführten Staaten 2007 die größten Anteile von Kindern auf Kompetenzstufe I aufwiesen. Die Tschechische Republik fällt dadurch auf, dass hier im Vergleich zu 2007 der Anteil von Schülerinnen und Schülern auf Kompetenzstufe I in 2011 geringer und der Anteil von Kindern auf Kompetenzstufe V größer geworden ist. Ein eher ungünstiges, gegenläufiges Muster zeigt England.

5.3 Inhaltsbereiche

Die TIMSS-Rahmenkonzeption zur Erfassung naturwissenschaftlicher Kompetenzen ermöglicht es, Teilbereiche der naturwissenschaftlichen Kompetenz darzustellen und auf diese Weise ein differenziertes Bild über Stärken und Schwächen der Schülerinnen und Schüler zu zeichnen. Unterschieden werden wie in Abschnitt 3 beschrieben die Inhaltsbereiche *Biologie*, *Physik/Chemie* und *Geographie* sowie die kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen*. In diesem Abschnitt werden zunächst Ergebnisse zu den Kompetenzen in den naturwissenschaftlichen Inhaltsbereichen im internationalen Vergleich sowie Unterschiede in diesen Ergebnissen zwischen 2007 und 2011 berichtet. In Abschnitt 5.4 wird dann auf Ergebnisse zu den drei kognitiven Anforderungsbereichen eingegangen.

Kompetenzen in den naturwissenschaftlichen Inhaltsbereichen im internationalen Vergleich: Abbildung 4.8 zeigt die in den einzelnen Staaten erreichten Leistungsmittelwerte in den Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geographie*. Aufgeführt sind hier nur OECD- und EU-Staaten sowie Staaten, die eine signifikant höhere Kompetenz auf der Gesamtskala als Deutschland aufweisen. Die Staaten sind entsprechend ihres Leistungsmittelwertes auf der Gesamtskala Naturwissenschaften in einer Rangfolge angeordnet.

Für Deutschland zeigen die Ergebnisse eine leichte relative Stärke der Schülerinnen und Schüler im Bereich *Physik/Chemie*. Hier liegt der Leistungsmittelwert bei 535 Punkten, also 7 Punkte über dem Wert der Gesamtskala Naturwissenschaften. Für den Bereich *Geographie* zeigt sich eine leichte relative Schwäche der Schülerinnen und Schüler in Deutschland. Hier liegt der Leistungsmittelwert bei 520 Punkten, also 8 Punkte unter dem Wert des Gesamttests. Das Kompetenzniveau im Bereich *Biologie* liegt mit 525 Punkten nahe bei dem Niveau der Gesamtskala zur naturwissenschaftlichen Kompetenz.

Im Vergleich zum OECD- und EU-Durchschnitt zeigt sich bei den Schülerinnen und Schülern in Deutschland ebenfalls eine relative Stärke im Bereich *Physik/Chemie*. Sowohl der OECD- als auch der EU-Mittelwert liegen in diesem Bereich signifikant unter dem Mittelwert in Deutschland. Die Differenz beträgt 13 Punkte zum OECD-Mittelwert und 17 Punkte zum EU-Mittelwert. In den anderen beiden Bereichen, *Biologie* und *Geographie*, zeigen sich keine signifikanten Abweichungen vom OECD- und vom EU-Mittelwert.

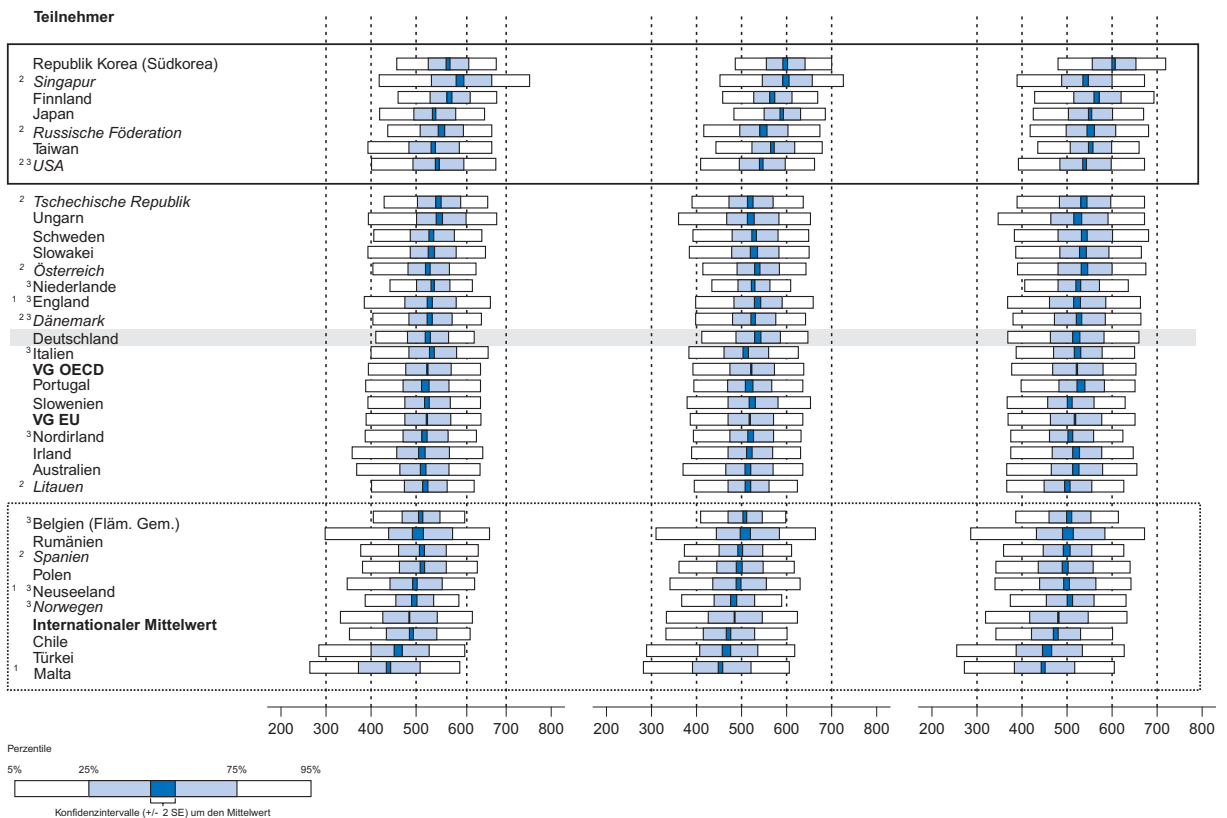
Sieben Staaten, die auch die Leistungsspitze in der Gesamtskala Naturwissenschaften bilden, weisen in allen drei Inhaltsbereichen ein signifikant höheres Kompetenzniveau auf als Deutschland. Im Bereich *Geographie* beträgt die Differenz zwischen Deutschland und dem Staat an der Spitze der Leistungsskala, der Republik Korea, beachtliche 83 Punkte. Im Bereich *Biologie* beträgt der Unterschied zu Singapur 72 Punkte und im Bereich *Physik/Chemie* 63 Punkte.

Neun Staaten erreichen in allen drei Inhaltsbereichen ein signifikant niedrigeres Kompetenzniveau als Deutschland. Zu dieser Gruppe gehören Spanien, Neuseeland, Norwegen und die Türkei. Auch die internationalen Mittelwerte für die Kompetenzen in den drei Inhaltsbereichen liegen signifikant unter den Werten der Schülerinnen und Schüler in Deutschland.

Unterschiede in den Kompetenzen in den naturwissenschaftlichen Inhaltsbereichen zwischen TIMSS 2007 und 2011: In Abbildung 4.9 sind Kompetenzunterschiede in den naturwissenschaftlichen Inhaltsbereichen zwischen TIMSS

Abbildung 4.8: Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den Inhaltsbereichen *Biologie, Physik/Chemie und Geographie*

Teilnehmer	Biologie		Physik/Chemie		Geographie	
	M (SE)	SD (SE)	M (SE)	SD (SE)	M (SE)	SD (SE)
Republik Korea (Südkorea)	571 (2.2) +	67 (1.1)	597 (2.6) +	65 (0.9)	603 (1.8) +	73 (1.4)
² Singapur	597 (4.3) +	102 (2.2)	598 (3.5) +	83 (1.9)	541 (3.0) +	85 (1.7)
Finnland	574 (2.8) +	68 (1.4)	568 (2.8) +	65 (1.6)	566 (2.9) +	81 (1.7)
Japan	540 (1.9) +	70 (0.9)	589 (1.9) +	61 (1.0)	551 (1.8) +	74 (1.1)
² Russische Föderation	556 (3.6) +	71 (1.5)	548 (4.0) +	79 (2.3)	552 (4.1) +	80 (2.1)
Taiwan	538 (2.4) +	83 (1.2)	569 (2.0) +	71 (1.4)	553 (2.5) +	69 (1.2)
^{2,3} USA	547 (2.1) +	84 (1.4)	544 (2.0) +	77 (1.2)	539 (2.1) +	85 (1.7)
² Tschechische Republik	550 (3.0) +	70 (1.5)	519 (3.1) -	75 (2.2)	537 (3.4) +	86 (1.9)
Ungarn	552 (3.5) +	86 (2.6)	520 (3.8) -	89 (2.8)	524 (4.4)	98 (2.7)
Schweden	534 (2.7) +	73 (1.4)	528 (2.5)	78 (1.5)	538 (3.2) +	90 (1.8)
Slowakei	534 (3.5) +	80 (2.8)	527 (4.0)	82 (2.8)	535 (3.8) +	86 (3.3)
² Österreich	526 (2.6)	69 (1.1)	535 (2.9)	70 (1.5)	539 (3.6) +	87 (1.7)
³ Niederlande	537 (1.8) +	56 (1.0)	526 (2.0) -	53 (1.2)	525 (2.7)	70 (2.0)
¹ England	530 (2.8)	86 (1.7)	535 (3.5)	79 (2.3)	522 (3.8)	91 (1.9)
^{2,3} Dänemark	530 (2.8)	73 (1.8)	526 (2.5) -	74 (1.9)	527 (3.0)	86 (1.9)
Deutschland	525 (2.6)	67 (1.6)	535 (3.1)	72 (1.7)	520 (3.7)	89 (1.5)
³ Italien	535 (2.7) +	79 (2.1)	509 (3.0) -	74 (1.7)	523 (3.6)	80 (2.0)
VG OECD	525 (0.6)	76 (0.4)	522 (0.6)	75 (0.4)	522 (0.7)	84 (0.4)
Portugal	520 (4.2)	78 (2.8)	517 (4.2)	74 (3.0)	531 (4.4)	76 (2.2)
Slowenien	524 (2.6)	76 (1.2)	524 (3.4)	84 (1.8)	506 (2.7) -	80 (1.8)
VG EU	524 (0.7)	77 (0.4)	518 (0.7)	76 (0.5)	518 (0.8)	85 (0.5)
³ Nordirland	519 (2.9)	76 (1.5)	520 (3.2)	73 (1.9)	507 (2.7) -	76 (2.4)
Irland	513 (3.6) -	88 (2.5)	517 (3.1) -	74 (2.5)	520 (3.8)	82 (2.3)
Australien	516 (3.1) -	83 (2.0)	514 (3.2) -	81 (2.2)	520 (3.5)	88 (2.0)
² Litauen	520 (2.9)	70 (1.9)	514 (3.1) -	69 (2.1)	501 (3.0) -	79 (1.8)
³ Belgien (Fläm. Gem.)	510 (2.4) -	62 (1.9)	507 (2.1) -	57 (0.9)	505 (2.8) -	69 (1.7)
Rumänien	504 (6.1) -	110 (4.1)	508 (5.7) -	108 (3.6)	502 (6.0) -	116 (4.0)
² Spanien	513 (2.8) -	79 (1.8)	497 (2.7) -	72 (2.1)	499 (3.8) -	81 (2.3)
Polen	514 (2.5) -	78 (1.2)	495 (3.3) -	78 (1.2)	496 (3.3) -	91 (1.9)
³ Neuseeland	497 (2.5) -	86 (1.3)	493 (2.7) -	88 (1.8)	499 (3.2) -	92 (1.5)
³ Norwegen	496 (3.0) -	63 (1.8)	482 (3.4) -	67 (1.6)	506 (3.0) -	78 (2.8)
Internationaler Mittelwert	485 (0.5)	89 (0.3)	484 (0.5)	89 (0.3)	481 (0.6)	96 (0.3)
Chile	490 (2.2) -	81 (1.7)	471 (2.5) -	82 (2.0)	475 (2.7) -	79 (2.4)
Türkei	460 (4.5) -	99 (3.5)	466 (4.7) -	100 (3.6)	456 (5.1) -	113 (4.3)
Malta	439 (2.4) -	101 (1.8)	453 (2.5) -	98 (1.7)	447 (2.2) -	100 (1.8)



Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

☐ In allen drei Bereichen signifikant positiv von den deutschen Mittelwerten abweichende Staaten ($p < .05$).

☐ In allen drei Bereichen signifikant negativ von den deutschen Mittelwerten abweichende Staaten ($p < .05$).

¹ = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

² = Der Ausschöpfungsgang und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

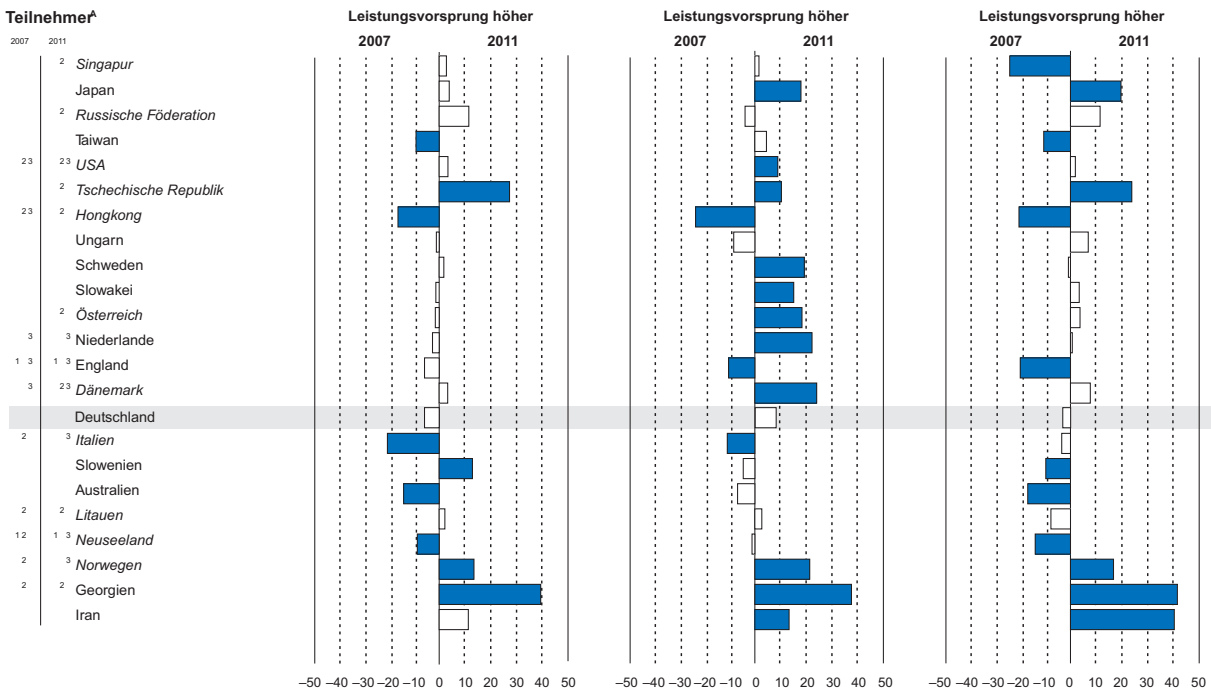
³ = Die Teilnahmequoten auf Schul- und/oder Schülerenebene erreichen nicht die internationalen Vorgaben.

+ = Mittelwert statistisch signifikant höher als der deutsche Mittelwert ($p < .05$).

- = Mittelwert statistisch signifikant niedriger als der deutsche Mittelwert ($p < .05$).

Abbildung 4.9: Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 in den Inhaltsbereichen *Biologie, Physik/Chemie und Geographie*

Teilnehmer ^A	Biologie			Physik/Chemie			Geographie		
	2007	2011	Veränderung ^B	2007	2011	Veränderung ^B	2007	2011	Veränderung ^B
	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)
² ²⁰⁰⁷ ²⁰¹¹ <i>Singapur</i>	595 (4.7)	597 (4.3)	3 (6.4)	597 (4.3)	598 (3.5)	2 (5.6)	565 (4.0)	541 (3.0)	-24 (5.0)
Japan	536 (2.3)	540 (1.9)	4 (2.9)	571 (2.9)	589 (1.9)	18 (3.4)	532 (3.5)	551 (1.8)	20 (4.0)
² <i>Russische Föderation</i>	545 (4.7)	556 (3.6)	12 (5.9)	552 (5.4)	548 (4.0)	-4 (6.7)	541 (5.5)	552 (4.1)	11 (6.8)
Taiwan	547 (2.8)	538 (2.4)	-9 (3.7)	564 (2.5)	569 (2.0)	5 (3.3)	563 (2.9)	553 (2.5)	-10 (3.9)
²³ ²³ <i>USA</i>	544 (2.9)	547 (2.1)	3 (3.5)	535 (3.1)	544 (2.0)	9 (3.7)	537 (3.2)	539 (2.1)	2 (3.9)
² <i>Tschechische Republik</i>	522 (3.5)	550 (3.0)	27 (4.6)	509 (3.4)	519 (3.1)	10 (4.6)	514 (3.5)	537 (3.4)	24 (4.9)
²³ ² <i>Hongkong</i>	540 (4.0)	524 (3.7)	-16 (5.4)	562 (4.0)	539 (4.4)	-23 (5.9)	568 (4.2)	548 (3.3)	-20 (5.4)
Ungarn	553 (3.3)	552 (3.5)	-1 (4.8)	529 (3.7)	520 (3.8)	-8 (5.3)	517 (4.3)	524 (4.4)	7 (6.1)
Schweden	532 (2.8)	534 (2.7)	2 (3.9)	509 (3.2)	528 (2.5)	19 (4.0)	539 (3.7)	538 (3.2)	-1 (4.9)
Slowakei	535 (4.6)	534 (3.5)	-1 (5.8)	512 (5.2)	527 (4.0)	15 (6.6)	532 (6.5)	535 (3.8)	3 (7.5)
² <i>Österreich</i>	528 (2.3)	526 (2.6)	-2 (3.5)	517 (3.1)	535 (2.9)	18 (4.2)	535 (2.5)	539 (3.6)	4 (4.4)
³ <i>Niederlande</i>	539 (2.7)	537 (1.8)	-3 (3.3)	503 (3.1)	526 (2.0)	22 (3.7)	524 (3.3)	525 (2.7)	1 (4.2)
¹ ³ <i>England</i>	536 (3.1)	530 (2.8)	-6 (4.2)	546 (3.3)	535 (3.5)	-10 (4.8)	542 (3.4)	522 (3.8)	-19 (5.1)
³ ²³ <i>Dänemark</i>	527 (3.2)	530 (2.8)	3 (4.3)	502 (2.8)	526 (2.5)	24 (3.7)	519 (3.3)	527 (3.0)	8 (4.5)
<i>Deutschland</i>	531 (2.2)	525 (2.6)	-6 (3.4)	527 (3.2)	535 (3.1)	8 (4.4)	524 (2.8)	520 (3.7)	-4 (4.6)
² ³ <i>Italien</i>	555 (3.6)	535 (2.7)	-20 (4.5)	520 (3.7)	509 (3.0)	-11 (4.8)	527 (4.1)	523 (3.6)	-3 (5.5)
Slowenien	511 (2.0)	524 (2.6)	13 (3.3)	528 (2.3)	524 (3.4)	-5 (4.1)	516 (3.2)	506 (2.7)	-10 (4.2)
Australien	529 (3.6)	516 (3.1)	-14 (4.7)	521 (3.8)	514 (3.2)	-7 (5.0)	536 (4.5)	520 (3.5)	-17 (5.7)
² ² <i>Litauen</i>	518 (2.2)	520 (2.9)	2 (3.7)	511 (2.0)	514 (3.1)	3 (3.7)	508 (2.9)	501 (3.0)	-8 (4.1)
¹ ² ³ <i>Neuseeland</i>	506 (2.8)	497 (2.5)	-8 (3.7)	494 (3.4)	493 (2.7)	-1 (4.3)	513 (3.4)	499 (3.2)	-14 (4.7)
² ³ <i>Norwegen</i>	482 (3.0)	496 (3.0)	13 (4.3)	461 (3.5)	482 (3.4)	21 (4.9)	490 (3.8)	506 (3.0)	17 (4.9)
² ² <i>Georgien</i>	421 (4.0)	461 (3.6)	39 (5.4)	403 (4.9)	440 (4.2)	37 (6.4)	416 (5.4)	458 (4.3)	42 (6.9)
Iran	437 (5.2)	449 (4.1)	11 (6.6)	440 (4.9)	453 (4.0)	13 (6.3)	416 (5.0)	457 (3.5)	40 (6.1)



■ Statistisch signifikante Unterschiede (p < .05).
 Kursiv gesetzt sind Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.
 B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

2007 und 2011 dargestellt. In der Abbildung sind nur Staaten aufgeführt, die sowohl an TIMSS 2007 als auch an TIMSS 2011 teilgenommen haben. Positive Werte in der Spalte Veränderung bedeuten, dass im Jahr 2011 mehr Punkte erzielt wurden als 2007.

Für Deutschland zeigt sich, dass die Kompetenz der Schülerinnen und Schüler im Bereich *Biologie* 2007 um 6 Punkte höher lag als 2011. Im Bereich *Physik/Chemie* zeigt sich hingegen 2011 ein um 8 Punkte höherer Leistungsmittelwert. Der Leistungsmittelwert im Bereich *Geographie* lag 2007 um 4 Punkte höher. Keiner dieser Unterschiede ist jedoch statistisch signifikant.

Norwegen, Georgien und die Tschechische Republik sind die Staaten, die sich in allen drei Inhaltsbereichen zwischen 2007 und 2011 signifikant verbessern konnten. Bei den übrigen Staaten finden sich zum Teil signifikante Verbesserungen und Verschlechterungen in einzelnen Bereichen.

5.4 Kognitive Anforderungsbereiche

Analog zu den Ergebnissen im vorigen Abschnitt werden in diesem Abschnitt Ergebnisse zu den drei kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* berichtet.

Kompetenzen in den kognitiven Anforderungsbereichen im internationalen Vergleich: Abbildung 4.10 zeigt die in den einzelnen Staaten erreichten Leistungsmittelwerte in den kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen*. Aufgeführt sind auch hier wie bei den Ergebnissen zu den drei Inhaltsbereichen nur OECD- und EU-Staaten sowie Staaten, die eine signifikant höhere Kompetenz auf der Gesamtskala als Deutschland aufweisen. Die Staaten sind entsprechend ihres Leistungsmittelwertes auf der Gesamtskala Naturwissenschaften in einer Rangfolge angeordnet.

Für Deutschland zeigen die Ergebnisse keine relativen Stärken oder Schwächen der Schülerinnen und Schüler in einem der drei kognitiven Anforderungsbereiche. Das Kompetenzniveau liegt in allen drei Anforderungsbereichen nahe bei dem Kompetenzniveau auf der Gesamtskala Naturwissenschaften.

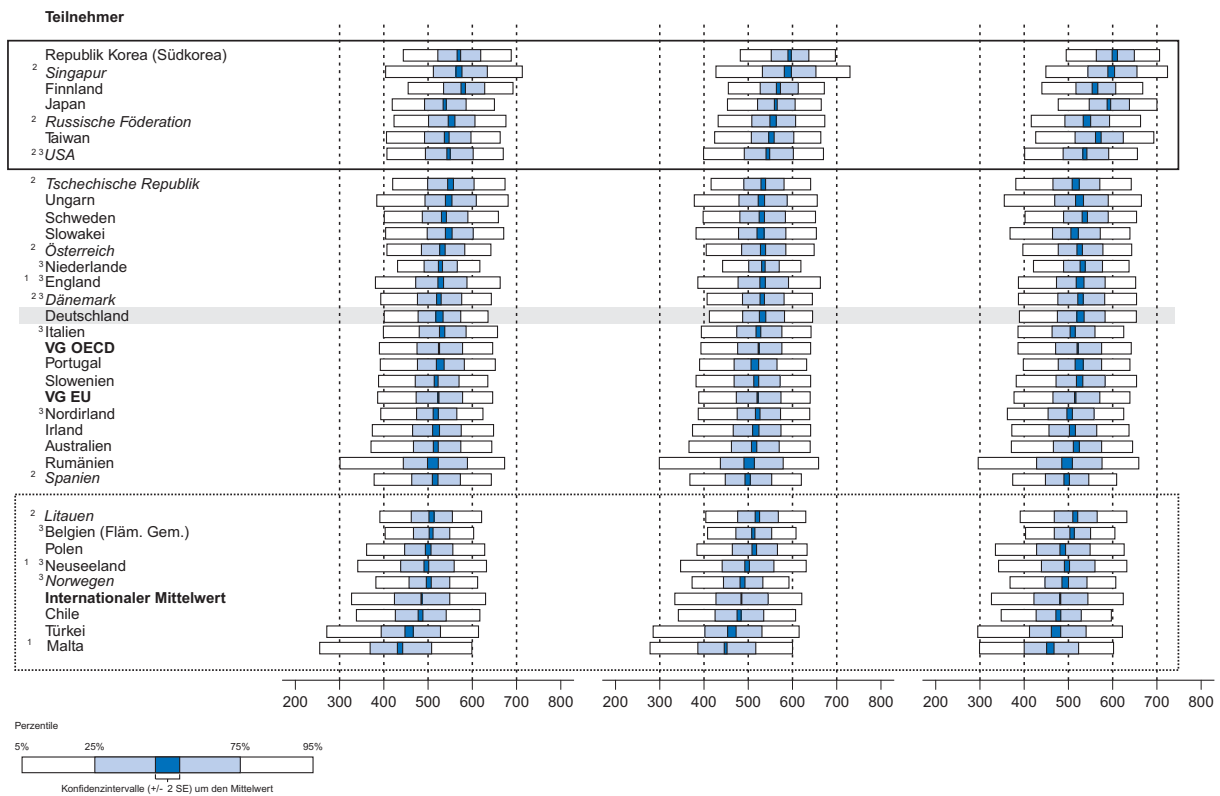
Während sich im Bereich *Reproduzieren* keine signifikanten Unterschiede zum OECD- und EU-Durchschnitt zeigen, liegt der Mittelwert der Schülerinnen und Schüler in Deutschland im Bereich des *Anwendens* signifikant über dem OECD- und dem EU-Durchschnitt. Die Differenz beträgt 9 Punkte zum OECD-Mittelwert und 12 Punkte zum EU-Mittelwert. Im Bereich des *Problemlösens* liegt Deutschland signifikant über dem EU-Mittelwert (Differenz: 11 Punkte).

Sieben Staaten, die auch die Leistungsspitze in der Gesamtskala Naturwissenschaften bilden, weisen in allen drei kognitiven Anforderungsbereichen ein signifikant höheres Kompetenzniveau auf als Deutschland. Im Bereich des *Reproduzierens* beträgt die Differenz zwischen Deutschland und Finnland 55 Punkte. Im Bereich des *Anwendens* beträgt der Unterschied zu der Republik Korea 60 Punkte und im Bereich des *Problemlösens* beachtliche 79 Punkte.

Acht Staaten erreichen in allen drei Inhaltsbereichen ein signifikant niedrigeres Kompetenzniveau als Deutschland. Zu dieser Gruppe gehören Polen, Neuseeland, Norwegen und die Türkei. Auch die internationalen Mittelwerte für die Kompetenzen in den drei kognitiven Anforderungsbereichen liegen signifikant unter den Werten der Schülerinnen und Schüler in Deutschland.

Abbildung 4.10: Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den kognitiven Anforderungsbereichen *Reproduzieren, Anwenden* und *Problemlösen*

Teilnehmer	Reproduzieren		Anwenden		Problemlösen	
	M (SE)	SD (SE)	M (SE)	SD (SE)	M (SE)	SD (SE)
Republik Korea (Südkorea)	570 (2.0) +	74 (1.2)	593 (1.9) +	65 (1.4)	605 (3.0) +	65 (1.2)
² Singapur	570 (3.4) +	93 (1.9)	590 (4.0) +	92 (2.0)	597 (3.8) +	83 (1.8)
Finnland	579 (2.5) +	72 (2.5)	568 (2.3) +	66 (1.9)	560 (3.2) +	70 (2.7)
Japan	538 (1.8) +	71 (1.8)	562 (1.6) +	64 (1.3)	591 (2.0) +	68 (1.7)
² Russische Föderation	553 (3.8) +	77 (2.0)	556 (3.6) +	73 (1.4)	542 (4.2) +	75 (1.7)
Taiwan	542 (2.7) +	78 (1.1)	552 (3.1) +	73 (1.3)	568 (3.2) +	81 (1.7)
^{2,3} USA	546 (1.9) +	80 (1.3)	544 (2.1) +	82 (1.2)	537 (2.3) +	77 (1.3)
² Tschechische Republik	551 (3.3) +	78 (2.1)	534 (2.6)	69 (1.9)	516 (4.0)	80 (2.3)
Ungarn	547 (3.7) +	91 (3.0)	530 (3.6)	85 (2.3)	525 (4.5)	95 (2.9)
Schweden	536 (2.8) +	79 (1.7)	531 (3.0)	77 (1.2)	537 (3.0) +	76 (1.9)
Slowakei	547 (3.8) +	82 (2.8)	528 (4.0)	83 (2.8)	514 (4.2) -	83 (3.0)
² Österreich	532 (3.1)	72 (1.3)	533 (2.9)	74 (1.4)	525 (3.1)	75 (1.8)
³ Niederlande	528 (2.3)	56 (1.7)	534 (2.0)	53 (1.0)	532 (2.9)	66 (1.5)
¹ England	529 (3.2)	87 (2.3)	532 (3.1)	84 (2.0)	526 (4.4)	81 (2.0)
^{2,3} Dänemark	524 (2.6)	76 (2.0)	532 (2.5)	72 (2.2)	527 (3.1)	81 (2.2)
Deutschland	524 (4.0)	71 (1.1)	533 (2.6)	70 (1.4)	526 (3.6)	80 (2.2)
³ Italien	532 (3.0)	79 (1.9)	523 (2.7) -	76 (1.6)	510 (2.9) -	73 (2.4)
VG OECD	525 (0.6)	78 (0.4)	524 (0.6)	75 (0.4)	521 (0.7)	78 (0.4)
Portugal	528 (4.4)	79 (2.5)	515 (4.3) -	74 (2.5)	524 (4.6)	74 (3.1)
Slowenien	518 (2.2)	76 (1.6)	518 (2.8) -	79 (1.5)	525 (3.6)	83 (1.8)
VG EU	523 (0.7)	79 (0.5)	521 (0.7)	77 (0.4)	515 (0.8)	80 (0.5)
³ Nordirland	517 (2.9)	70 (1.9)	521 (2.6) -	76 (1.9)	503 (3.1) -	80 (2.7)
Irland	518 (3.9)	83 (2.0)	517 (3.6) -	81 (2.2)	509 (3.4) -	81 (1.9)
Australien	517 (2.8)	83 (2.0)	513 (3.0) -	83 (2.2)	518 (3.4)	83 (2.1)
Rumänien	511 (6.1)	112 (4.0)	502 (5.9) -	109 (3.9)	497 (6.0) -	111 (3.8)
² Spanien	516 (3.2)	81 (1.9)	499 (3.0) -	77 (1.6)	496 (3.1) -	72 (2.0)
² Litauen	508 (2.9) -	70 (1.7)	521 (2.5) -	69 (1.5)	515 (2.8) -	73 (1.9)
³ Belgien (Fläm. Gem.)	507 (2.2) -	61 (1.4)	511 (1.8) -	61 (1.0)	508 (2.5) -	61 (1.7)
Polen	500 (3.2) -	81 (1.1)	514 (2.6) -	76 (1.5)	487 (3.2) -	89 (1.4)
³ Neuseeland	496 (2.7) -	89 (1.7)	497 (2.6) -	86 (1.5)	497 (2.9) -	88 (1.7)
³ Norwegen	502 (2.8) -	70 (1.3)	487 (2.8) -	66 (1.3)	493 (3.7) -	73 (2.0)
Internationaler Mittelwert	485 (0.6)	92 (0.3)	484 (0.5)	87 (0.3)	481 (0.6)	91 (0.3)
Chile	483 (2.7) -	85 (1.4)	479 (2.3) -	80 (1.3)	477 (2.8) -	75 (2.1)
Türkei	457 (4.7) -	104 (3.7)	463 (4.8) -	99 (3.7)	472 (5.3) -	99 (3.9)
¹ Malta	437 (3.0) -	103 (2.8)	449 (1.6) -	98 (1.5)	459 (4.2) -	92 (1.6)



Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

□ In allen drei Bereichen signifikant positiv von den deutschen Mittelwerten abweichende Staaten (p < .05).

⋯ In allen drei Bereichen signifikant negativ von den deutschen Mittelwerten abweichende Staaten (p < .05).

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

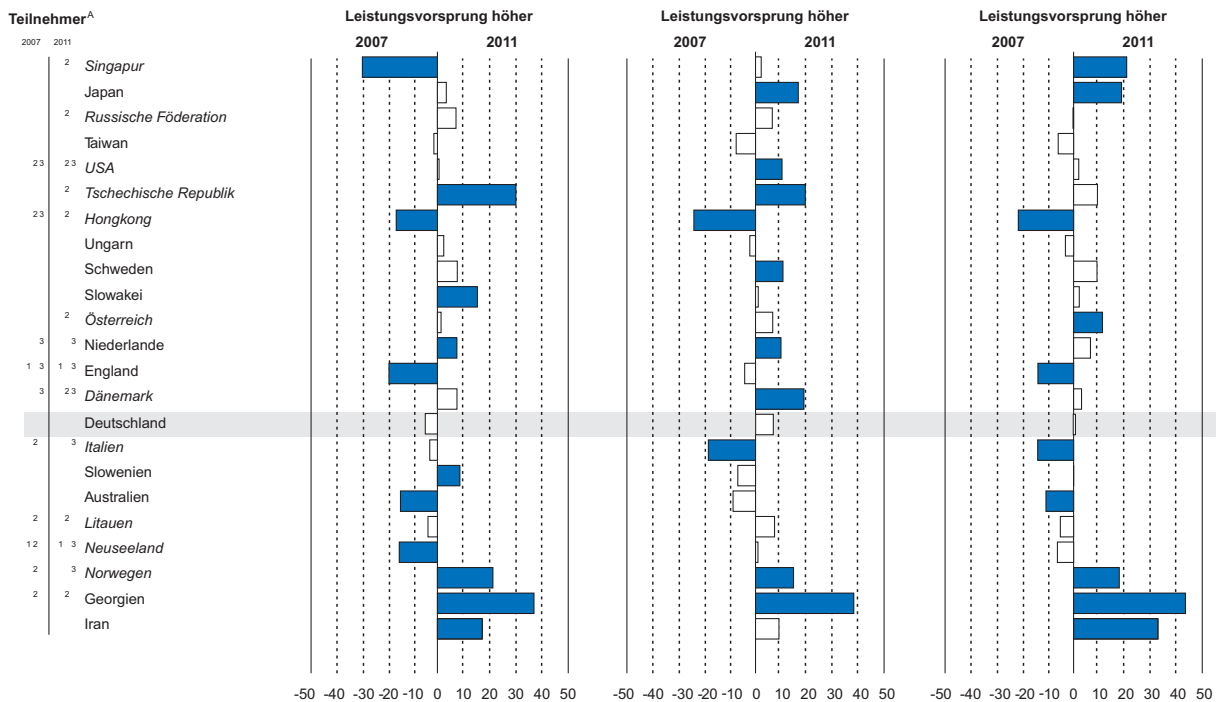
3 = Die Teilnahmequoten auf Schul- und/oder Schülerbene erreichen nicht die internationalen Vorgaben.

+ = Mittelwert statistisch signifikant höher als der deutsche Mittelwert (p < .05).

- = Mittelwert statistisch signifikant niedriger als der deutsche Mittelwert (p < .05).

Abbildung 4.11: Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 in den kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen*

Teilnehmer ^A	Reproduzieren			Anwenden			Problemlösen		
	2007	2011	Veränderung ^B	2007	2011	Veränderung ^B	2007	2011	Veränderung ^B
	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)	M ₀₇ (SE)	M ₁₁ (SE)	M ₁₁ -M ₀₇ (SE)
² Singapur	599 (4.4)	570 (3.4)	-29 (5.6)	587 (4.2)	590 (4.0)	2 (5.8)	576 (4.0)	597 (3.8)	20 (5.5)
Japan	534 (2.7)	538 (1.8)	3 (3.3)	546 (3.2)	562 (1.6)	16 (3.6)	573 (2.3)	591 (2.0)	18 (3.0)
² Russische Föderation	546 (5.6)	553 (3.8)	7 (6.8)	550 (5.3)	556 (3.6)	6 (6.4)	542 (5.2)	542 (4.2)	0 (6.7)
Taiwan	544 (2.8)	542 (2.7)	-1 (3.9)	560 (2.2)	552 (3.1)	-7 (3.8)	574 (3.2)	568 (3.2)	-6 (4.5)
^{2,3} USA	546 (2.7)	546 (1.9)	1 (3.3)	534 (3.1)	544 (2.1)	10 (3.7)	535 (3.0)	537 (2.3)	2 (3.8)
² Tschechische Republik	521 (2.9)	551 (3.3)	30 (4.4)	515 (3.4)	534 (2.6)	19 (4.3)	507 (3.6)	516 (4.0)	9 (5.4)
^{2,3} Hongkong	553 (3.9)	537 (3.6)	-16 (5.3)	552 (3.4)	529 (3.5)	-24 (4.8)	563 (4.9)	541 (4.2)	-21 (6.5)
Ungarn	544 (3.5)	547 (3.7)	2 (5.0)	532 (3.8)	530 (3.6)	-2 (5.2)	528 (4.2)	525 (4.5)	-3 (6.2)
Schweden	528 (3.0)	536 (2.8)	8 (4.1)	520 (3.0)	531 (3.0)	11 (4.2)	528 (4.3)	537 (3.0)	9 (5.2)
Slowakei	531 (4.8)	547 (3.8)	15 (6.1)	527 (4.9)	528 (4.0)	1 (6.3)	512 (5.3)	514 (4.2)	2 (6.7)
² Österreich	531 (2.5)	532 (3.1)	1 (3.9)	527 (2.7)	533 (2.9)	7 (3.9)	514 (2.9)	525 (3.1)	11 (4.2)
³ Niederlande	521 (2.6)	528 (2.3)	7 (3.4)	525 (2.4)	534 (2.0)	10 (3.1)	526 (2.7)	532 (2.9)	6 (4.0)
^{1,3} England	547 (3.4)	529 (3.2)	-19 (4.7)	537 (3.2)	532 (3.1)	-4 (4.5)	540 (2.8)	526 (4.4)	-14 (5.2)
³ Dänemark	517 (3.3)	524 (2.6)	7 (4.2)	513 (3.2)	532 (2.5)	19 (4.1)	524 (4.5)	527 (3.1)	3 (5.4)
Deutschland	529 (2.4)	524 (4.0)	-4 (4.6)	526 (2.5)	533 (2.6)	7 (3.6)	525 (2.6)	526 (3.6)	1 (4.5)
² Italien	535 (4.2)	532 (3.0)	-3 (5.1)	541 (3.4)	523 (2.7)	-18 (4.4)	523 (3.6)	510 (2.9)	-14 (4.6)
Slowenien	510 (2.0)	518 (2.2)	9 (2.9)	525 (2.5)	518 (2.8)	-7 (3.8)	525 (2.0)	525 (3.6)	0 (4.1)
Australien	532 (3.6)	517 (2.8)	-14 (4.5)	522 (3.8)	513 (3.0)	-9 (4.8)	528 (4.1)	518 (3.4)	-11 (5.3)
² Litauen	511 (2.3)	508 (2.9)	-4 (3.7)	513 (3.3)	521 (2.5)	7 (4.2)	521 (2.9)	515 (2.8)	-5 (4.0)
^{1,2} Neuseeland	511 (3.4)	496 (2.7)	-15 (4.3)	496 (2.7)	497 (2.6)	1 (3.7)	503 (4.0)	497 (2.9)	-6 (5.0)
² Norwegen	480 (3.2)	502 (2.8)	21 (4.3)	472 (3.5)	487 (2.8)	15 (4.5)	475 (3.2)	493 (3.7)	17 (4.9)
² Georgien	429 (4.3)	466 (3.9)	37 (5.8)	415 (4.5)	452 (4.4)	38 (6.3)	379 (6.0)	422 (5.0)	43 (7.8)
Iran	431 (5.0)	448 (4.3)	17 (6.6)	443 (4.9)	452 (3.8)	9 (6.2)	427 (4.6)	459 (3.9)	32 (6.0)



■ Statistisch signifikante Unterschiede ($p < .05$).

Kursiv gesetzt sind Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Unterschiede in den Kompetenzen in den kognitiven Anforderungsbereichen zwischen TIMSS 2007 und 2011: In Abbildung 4.11 sind Kompetenzunterschiede in den kognitiven Anforderungsbereichen zwischen TIMSS 2007 und 2011 dargestellt. In der Tabelle sind nur Staaten aufgeführt, die sowohl an TIMSS 2007 als auch an TIMSS 2011 teilgenommen haben. Positive Werte in der Spalte Veränderung bedeuten, dass im Jahr 2011 mehr Punkte erzielt wurden als 2007.

Für Deutschland zeigt sich, dass die Kompetenz der Schülerinnen und Schüler im Bereich *Reproduzieren* 2007 um 4 Punkte höher lag als 2011. Im Bereich *Anwenden* zeigte sich hingegen 2011 ein um 7 Punkte höherer Leistungsmittelwert. Der Leistungsmittelwert im Bereich *Problemlösen* lag 2011 um 1 Punkt höher als 2007. Keiner dieser Unterschiede ist jedoch statistisch signifikant.

Norwegen und Georgien sind die Staaten, die sich in allen drei kognitiven Anforderungsbereichen von 2007 nach 2011 signifikant verbessern konnten. In Hongkong hingegen lagen die Leistungsmittelwerte in allen drei Anforderungsbereichen 2011 niedriger als 2007. Bei den übrigen Staaten finden sich zum Teil signifikante Verbesserungen und Verschlechterungen in einzelnen Bereichen.

5.5 Einstellungen und Selbstkonzept

Neben den in der TIMSS-Rahmenkonzeption beschriebenen Bereichen naturwissenschaftlicher Kompetenzen (siehe Abschnitt 3) stellen auch positive Einstellungen gegenüber Naturwissenschaften und ein positives naturwissenschaftsbezogenes Selbstkonzept wichtige Zielbereiche naturwissenschaftlichen Unterrichts dar (Koballa & Glynn, 2007; Prenzel, Schütte & Walter, 2007). Unter Einstellungen zu Naturwissenschaften werden affektive Bewertungen von Naturwissenschaften verstanden. Sie sind eine wichtige Voraussetzung für die Bereitschaft, sich – auch nach der Schulzeit – mit naturwissenschaftlichen Sachverhalten aktiv auseinanderzusetzen. Dies gilt auch für die Bereitschaft, naturwissenschaftliche Kurse in der Schule zu wählen und später naturwissenschaftliche Berufe zu ergreifen (Koballa & Glynn, 2007).

Unter dem naturwissenschaftsbezogenen Selbstkonzept wird die Selbsteinschätzung der Schülerinnen und Schüler hinsichtlich ihrer Leistungsstärke im Bereich der Naturwissenschaften verstanden. Ein positives naturwissenschaftsbezogenes Selbstkonzept ist wichtig, da es Vertrauen in die eigenen naturwissenschaftlichen Kompetenzen gibt und somit auch eine wichtige Voraussetzung für die Bereitschaft zur Auseinandersetzung mit naturwissenschaftlichen Themen darstellt (z.B. Marsh & Yeung, 1997).

Positive Einstellungen und ein positives Selbstkonzept werden auch deshalb als wichtige Zielbereiche naturwissenschaftlichen Unterrichts angesehen, weil von einer wechselseitigen Beeinflussung von Einstellungen und Selbstkonzept einerseits und der Entwicklung naturwissenschaftlicher Kompetenzen andererseits auszugehen ist (z.B. Guay, Marsh & Boivin, 2003).

In diesem Abschnitt werden Ergebnisse aus TIMSS 2011 zu Einstellungen und zum Selbstkonzept der Schülerinnen und Schüler aus Deutschland dargestellt und mit den Ergebnissen aus TIMSS 2007 verglichen. Auf die Darstellung des internationalen Vergleichs der Ergebnisse wurde vor folgendem Hintergrund verzichtet: Fachbezogene Selbstkonzepte entstehen unter anderem durch soziale Vergleichsprozesse (Köller, 2004). Schülerinnen und Schüler vergleichen ihre eigenen Fähigkeiten in einem Fach mit denen von Mitschülerinnen und

Mitschülern sowie Peers. Sind die Fähigkeiten dieser Bezugsgruppe beispielsweise sehr hoch ausgeprägt, hat dies zur Folge, dass die einzelne Schülerin beziehungsweise der einzelne Schüler ihre beziehungsweise seine eigenen Fähigkeiten geringer einschätzt, als wenn die Bezugsgruppe weniger leistungsstark wäre. Diese Auswirkungen der Bezugsgruppe auf das individuelle Selbstkonzept werden als Referenzgruppeneffekte bezeichnet und es wurde gezeigt, dass diese Effekte in verschiedenen Kulturen unterschiedlich auftreten (Seaton, Marsh & Craven, 2009). Referenzgruppeneffekte treten nicht nur bei fachbezogenen Selbstkonzepten, sondern auch bei affektiv-motivationalen Merkmalen wie Interessen auf. Es ist davon auszugehen, dass Referenzgruppeneffekte auch bei Einstellungen auftreten (Seaton, Marsh & Craven, 2010).

Die Ergebnisse eines internationalen Vergleichs von Einstellungen und Selbstkonzepten wären von solchen Referenzgruppeneffekten betroffen, da sich die Staaten auch im Niveau naturwissenschaftlicher Kompetenzen unterscheiden (siehe Abschnitt 5.1). Das Niveau von Einstellungen und Selbstkonzepten könnte in Staaten mit besonders hohem Kompetenzniveau durch Referenzgruppeneffekte „nach unten gedrückt“ sein, in Staaten mit besonders niedrigem Kompetenzniveau hingegen angehoben. So ist beispielsweise in Japan das naturwissenschaftsbezogene Selbstkonzept der Schülerinnen und Schüler im internationalen Vergleich sehr niedrig ausgeprägt, obwohl die Schülerinnen und Schüler in Japan hinsichtlich der naturwissenschaftlichen Kompetenz im internationalen Vergleich sehr gut abschneiden. Neben Referenzgruppeneffekten kommt hier noch die Tendenz asiatischer Schülerinnen und Schüler hinzu, sich generell eher niedrig einzuschätzen (Kitayama, Markus, Matsumoto & Norasakkunkit, 1997). Vor diesem Hintergrund ist eine Interpretation des Niveaus von Einstellungen und Selbstkonzepten im internationalen Vergleich schwierig. In diesem Abschnitt werden daher nur die Ergebnisse für Deutschland in den Blick genommen und Unterschiede zwischen TIMSS 2011 und 2007 dargestellt.

Im Rahmen von TIMSS wurden die Einstellungen und das Selbstkonzept nicht mit Bezug auf die Naturwissenschaften, sondern im Hinblick auf das Fach Sachunterricht erfasst. Da in Deutschland naturwissenschaftliche Inhalte in der Grundschule nicht in einem eigenen Fach unterrichtet werden, ist davon auszugehen, dass die Schülerinnen und Schüler noch keine klare Vorstellung davon haben, was naturwissenschaftliche Inhalte von anderen Inhalten des Sachunterrichts unterscheidet. Aus diesem Grund werden Einstellungen und Selbstkonzept in TIMSS in Bezug auf das gesamte Fach Sachunterricht erfasst.

Einstellung zum Fach Sachunterricht: Die Einstellung zum Fach Sachunterricht wird in TIMSS mit Hilfe eines Schülerfragebogens erfasst. Darin geben die Schülerinnen und Schüler an, ob sie den folgenden Aussagen über das Lernen im Sachunterricht *völlig zustimmen, eher zustimmen, eher nicht zustimmen* oder *überhaupt nicht zustimmen*:

1. Ich lerne gern im Sachunterricht.
2. Sachunterricht ist langweilig.
3. Ich mag Sachunterricht.

Die Antworten der Schülerinnen und Schüler werden bei Aussage 1 und 3 mit 1 („stimme überhaupt nicht zu“) bis 4 („stimme völlig zu“) Punkten kodiert, bei Aussage 2 in umgekehrter Reihenfolge. Für jedes Kind wird dann der Mittelwert über die drei Aussagen gebildet. Dieser Skalenwert drückt aus, wie stark posi-

tiv die Einstellung des Kindes zum Sachunterricht ausgeprägt ist. Die interne Konsistenz dieser aus drei Items bestehenden Skala liegt bei einem Cronbachs Alpha von $\alpha = .86$.

In einem weiteren Schritt wurden Gruppen von Schülerinnen und Schülern mit niedriger, mittlerer und hoher positiver Einstellung zum Sachunterricht gebildet. Dazu wurden Kinder mit einem Mittelwert (M) von $1 \leq M \leq 2$ als *niedrig*, Kinder mit einem Mittelwert von $2 < M < 3$ als *mittel* und Kinder mit einem Mittelwert von $3 \leq M \leq 4$ als *hoch* klassifiziert. Abbildung 4.12 zeigt, wie stark positiv die Einstellungen der Schülerinnen und Schüler zum Sachunterricht bei TIMSS 2007 und 2011 in Deutschland ausgeprägt waren (mittlerer Skalenwert). Außerdem ist dargestellt, wie sich die Schülerinnen und Schüler auf die Gruppen der Kinder mit *niedrigen*, *mittleren* und *hohen* positiven Einstellungen verteilen.

Es zeigt sich, dass die Einstellungen der Schülerinnen und Schüler zum Sachunterricht am Ende der Grundschulzeit sehr positiv ausgeprägt sind. Der Mittelwert von 3.37 liegt recht nah am theoretischen Maximum der Skala (4). Im Vergleich zu TIMSS 2007 zeigt sich nur ein marginaler deskriptiver Unterschied, der nicht signifikant ist. Dem hohen Mittelwert von 3.37 entsprechend groß ist der Anteil der Schülerinnen und Schüler, die eine *hohe* positive Einstellung angeben: Etwa 80 Prozent der Schülerinnen und Schüler erfüllten dieses Kriterium. 11 Prozent weisen eine *mittlere* positive Einstellung auf und nur etwa 9 Prozent eine *niedrige* positive Einstellung. Die Anteile der Schülerinnen und Schüler in diesen drei Gruppen unterscheiden sich zwischen TIMSS 2007 und 2011 praktisch nicht. Keiner der deskriptiven Unterschiede ist signifikant.

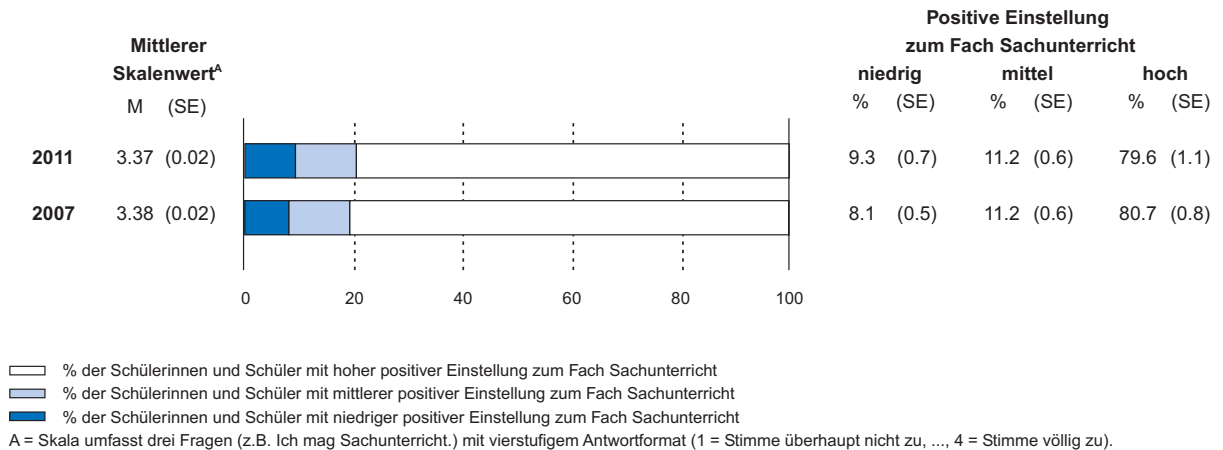
Wie theoretisch zu erwarten war, zeigt sich ein positiver Zusammenhang zwischen positiven Einstellungen zum Fach Sachunterricht und der naturwissenschaftlichen Kompetenz der Kinder. Die Korrelation ist allerdings mit $r = .10$ ($p < .001$) schwach ausgeprägt. In Abbildung 4.13 wird dieser Zusammenhang noch einmal illustriert, indem dargestellt wird, wie groß die Anteile der Schülerinnen und Schüler mit *hohen*, *mittleren* und *niedrigen* positiven Einstellungen auf den fünf Kompetenzstufen sind.

Der Anteil der Schülerinnen und Schüler mit *hoher* positiver Einstellung zum Sachunterricht ist auf den beiden unteren Kompetenzstufen am geringsten (je 74%). Auf den Kompetenzstufen III, IV und V steigt der Anteil dieser Gruppe von Kindern bis auf 86 Prozent (Kompetenzstufe V) an. Der geringste Anteil an Schülerinnen und Schülern mit *niedrigen* positiven Einstellungen zum Sachunterricht findet sich auf den beiden höchsten Kompetenzstufen (6% auf Kompetenzstufe V und 8% auf Kompetenzstufe IV).

Sachunterrichtsbezogenes Selbstkonzept: Auch das sachunterrichtsbezogene Selbstkonzept wird in TIMSS mit Hilfe eines Schülerfragebogens erfasst. Darin geben die Schülerinnen und Schüler an, ob sie den folgenden Aussagen *völlig zustimmen*, *eher zustimmen*, *eher nicht zustimmen* oder *überhaupt nicht zustimmen*.

1. Normalerweise bin ich gut im Sachunterricht.
2. Sachunterricht fällt mir schwerer als vielen meiner Mitschüler.
3. Ich bin einfach nicht gut im Sachunterricht.
4. Ich lerne schnell im Sachunterricht.

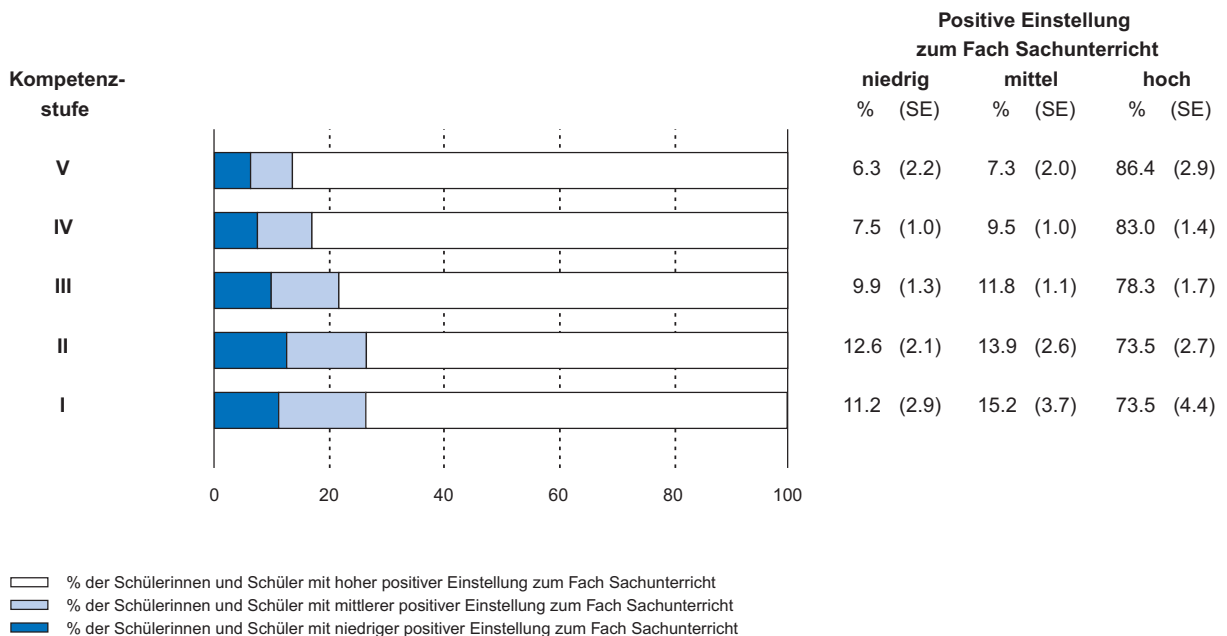
Abbildung 4.12: Mittlere positive Einstellung von Schülerinnen und Schülern zum Fach Sachunterricht sowie prozentuale Verteilung nach hoher, mittlerer und niedriger Einstellung – TIMSS 2007 und TIMSS 2011 im Vergleich



IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Abbildung 4.13: Prozentuale Verteilung von Schülerinnen und Schülern auf den fünf Kompetenzstufen nach Einstellungen zum Fach Sachunterricht



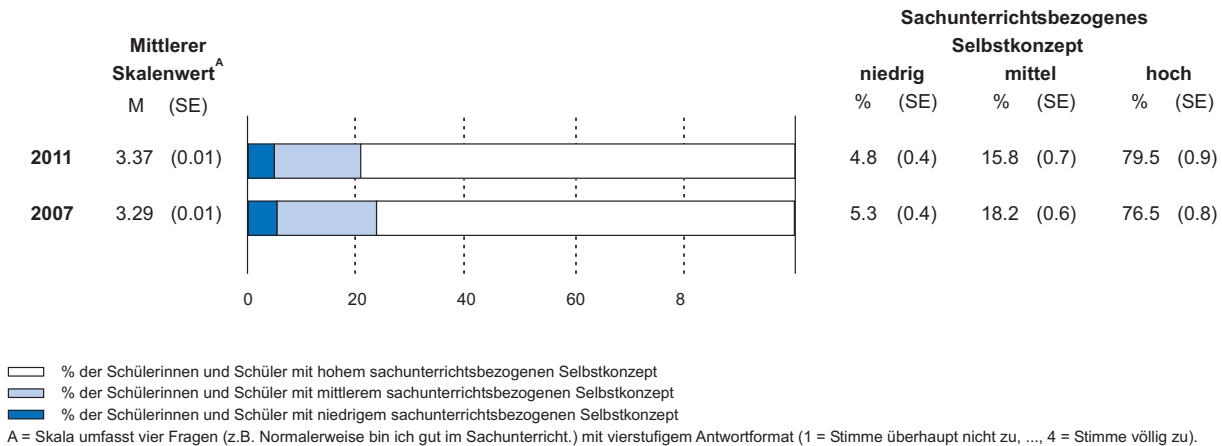
IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Die Antworten der Schülerinnen und Schüler werden bei Aussage 1 und 4 mit 1 („stimme überhaupt nicht zu“) bis 4 („stimme völlig zu“) Punkten kodiert, bei den Aussagen 2 und 3 in umgekehrter Reihenfolge. Für jedes Kind wird dann der Mittelwert über die vier Aussagen gebildet. Die interne Konsistenz dieser aus vier Items bestehenden Skala liegt bei einem Cronbachs Alpha von $\alpha = .79$.

Auch hier wurden die Schülerinnen und Schüler in Gruppen mit *niedrigem*, *mittlerem* und *hohem* Selbstkonzept eingeteilt. Kinder mit einem Mittelwert (M) von $1 \leq M \leq 2$ wurden als *niedrig*, Kinder mit einem Mittelwert von $2 < M < 3$ als *mittel* und Kinder mit einem Mittelwert von $3 \leq M \leq 4$ als *hoch*

Abbildung 4.14: Mittleres sachunterrichtsbezogenes Selbstkonzept der Schülerinnen und Schüler sowie prozentuale Verteilung nach *niedrigem*, *mittlerem* und *hohem* Selbstkonzept – TIMSS 2007 und TIMSS 2011 im Vergleich

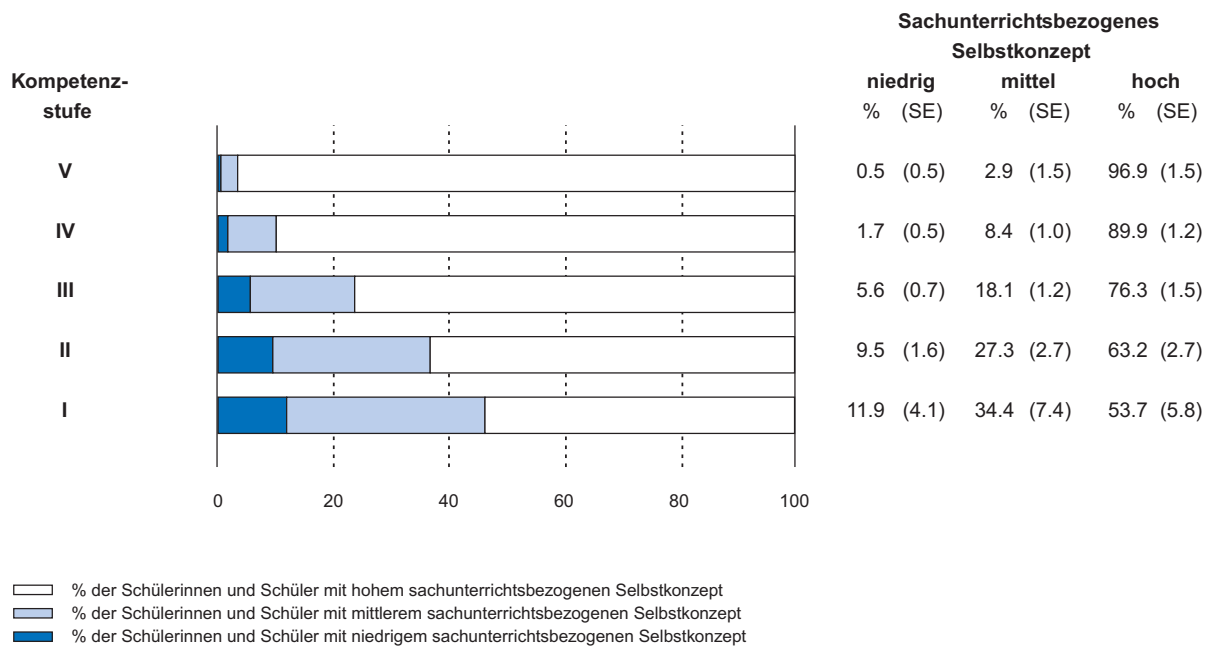


klassifiziert. Abbildung 4.14 zeigt, wie stark positiv das sachunterrichtsbezogene Selbstkonzept der Schülerinnen und Schüler bei TIMSS 2007 und 2011 in Deutschland ausgeprägt war (mittlerer Skalenwert). Außerdem ist dargestellt, wie sich die Schülerinnen und Schüler auf die Gruppen der Kinder mit *niedrigem*, *mittlerem* und *hohem* Selbstkonzept verteilen.

Es zeigt sich wie schon bei den Einstellungen zum Sachunterricht, dass auch das sachunterrichtsbezogene Selbstkonzept der Schülerinnen und Schüler am Ende der Grundschulzeit sehr positiv ausgeprägt ist. Der Mittelwert von 3.37 liegt recht nah am theoretischen Maximum der Skala (4). Im Vergleich zu TIMSS 2007 zeigt sich ein signifikanter Unterschied von 0.08 Punkten. Diese Differenz entspricht einer Effektgröße von Cohens $d = .11$. Hierbei handelt es sich um eine sehr kleine Effektgröße. Dem hohen Mittelwert von 3.37 entsprechend groß ist der Anteil der Schülerinnen und Schüler, die ein *hohes* sachunterrichtsbezogenes Selbstkonzept aufweisen: 80 Prozent der Schülerinnen und Schüler erfüllten dieses Kriterium. 16 Prozent weisen ein *mittleres* Selbstkonzept auf und nur etwa 5 Prozent ein *niedriges* Selbstkonzept. Die Anteile der Schülerinnen und Schüler in den Gruppen mit *hohem* und *mittlerem* Selbstkonzept unterscheiden sich zwischen TIMSS 2007 und 2011 signifikant. Für beide Gruppen liegen die Anteile der Schülerinnen und Schüler 2011 höher als 2007.

Wie erwartet zeigt sich ein positiver Zusammenhang zwischen dem sachunterrichtsbezogenen Selbstkonzept und der naturwissenschaftlichen Kompetenz der Kinder. Die Korrelation beträgt $r = .31$ ($p < .001$). In Abbildung 4.15 wird dieser Zusammenhang illustriert, indem gezeigt wird, wie groß die Anteile der Schülerinnen und Schüler mit *niedrigem*, *mittlerem* und *hohem* Selbstkonzept auf den fünf Kompetenzstufen sind.

Der Anteil der Schülerinnen und Schüler mit *hohem* sachunterrichtsbezogenem Selbstkonzept ist auf der untersten Kompetenzstufe am geringsten (54%) und steigt dann bis auf 97 Prozent in der Gruppe der Schülerinnen und Schüler auf Kompetenzstufe V an. Der geringste Anteil an Schülerinnen und Schülern mit *niedrigem* Selbstkonzept findet sich auf den beiden höchsten Kompetenzstufen (0.5% auf Kompetenzstufe V und 1.7% auf Kompetenzstufe IV).

Abbildung 4.15: Prozentuale Verteilung der Schülerinnen und Schüler auf den fünf Kompetenzstufen nach sachunterrichtsbezogenem Selbstkonzept

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

6 Zusammenfassung

In diesem Kapitel wurde dargestellt, über welche naturwissenschaftlichen Kompetenzen Schülerinnen und Schüler in Deutschland am Ende ihrer Grundschulzeit im internationalen Vergleich verfügen und welche Unterschiede sich im Vergleich zu der Erhebung von 2007 abzeichnen.

Niveau und Streuung der naturwissenschaftlichen Kompetenz: Das Niveau naturwissenschaftlicher Kompetenz von Schülerinnen und Schülern am Ende ihrer Grundschulzeit liegt bei 528 Punkten auf der TIMSS-Skala. Deutschland liegt damit in der oberen Hälfte der Rangreihe. Die naturwissenschaftliche Kompetenz der Schülerinnen und Schüler ist in Deutschland deutlich über dem internationalen Mittelwert von 486 Punkten, während der Abstand zum EU-Durchschnitt (521 Punkte) zwar statistisch signifikant, aber eher gering ist. Der Unterschied zum Durchschnitt der OECD-Staaten ist nicht signifikant. Im Vergleich zu den Staaten mit dem höchsten naturwissenschaftlichen Kompetenzniveau schneidet Deutschland deutlich schlechter ab. Der Abstand zur Republik Korea, dem Teilnehmerstaat an der Spitze der Leistungsskala, und zum in der Rangreihe folgenden Singapur ist mit 59 Punkten beziehungsweise 55 Punkten beträchtlich. Deutschland befindet sich zusammen mit Schweden, den Niederlanden, England, Italien und weiteren Staaten in einer Gruppe von Staaten, deren naturwissenschaftliches Kompetenzniveau ähnlich ausgeprägt ist.

Bei TIMSS 2011 erzielt Deutschland genau denselben Mittelwert in der naturwissenschaftlichen Kompetenz wie in 2007. Es finden sich also keine Unterschiede im Niveau der naturwissenschaftlichen Kompetenz zwischen 2007 und 2011. Bei einer Reihe weiterer Staaten, darunter die Russische Föderation, die USA und Taiwan, finden sich ebenfalls keine Veränderungen im

Kompetenzniveau. Staaten, die hingegen eine stark positive Veränderung des Niveaus der naturwissenschaftlichen Kompetenz zeigen, sind beispielsweise Georgien, die Tschechische Republik und Iran.

In Deutschland ist die Streuung der Kompetenzen vergleichsweise gering ausgeprägt, das heißt die Schülerinnen und Schüler verfügen über eine vergleichsweise homogene naturwissenschaftliche Kompetenz. Der Unterschied zwischen den 5 Prozent leistungsschwächsten und den 5 Prozent leistungsstärksten Schülerinnen und Schülern ist im Vergleich zu anderen Staaten eher gering. Ähnliche Streuungen in der naturwissenschaftlichen Kompetenz lassen sich in vielen der anderen OECD- und EU-Staaten finden. Eine Ausnahme bilden die Niederlande und auch die Flämische Gemeinschaft in Belgien, in denen die Kompetenzen der Schülerinnen und Schüler besonders homogen ausgeprägt sind. In 2011 ist die Streuung der Leistungswerte der Schülerinnen und Schüler in Deutschland im Vergleich zu 2007 ähnlich ausgeprägt.

Kompetenzstufen: Anhand der in Abschnitt 4 beschriebenen fünf Kompetenzstufen können die naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler inhaltlich illustriert werden. In Deutschland befinden sich 7 Prozent der Schülerinnen und Schüler auf der höchsten Kompetenzstufe V. Diese Schülerinnen und Schüler verstehen und begründen bereits am Ende der Grundschulzeit naturwissenschaftliche Zusammenhänge, sie interpretieren einfache naturwissenschaftliche Experimente und ziehen Schlussfolgerungen daraus. Weitere 32 Prozent der Schülerinnen und Schüler befinden sich auf der Kompetenzstufe IV. Diese Schülerinnen und Schüler verfügen über naturwissenschaftliche Konzepte, mit denen sie Alltagsphänomene und Sachverhalte beschreiben können. Sie besitzen ausreichendes Basiswissen, welches zu einer erfolgreichen Teilnahme an weiterführendem naturwissenschaftlichen Unterricht beitragen kann. Demgegenüber sind die Voraussetzungen der Gruppe von Schülerinnen und Schülern in Deutschland, die sich auf den Kompetenzstufen I und II befinden, als ungünstig zu bezeichnen. Sie können zwar elementares Faktenwissen abrufen, allerdings fehlt es ihnen an einem grundlegenden naturwissenschaftlichen Verständnis. Besonders kritisch ist der Anteil der Schülerinnen und Schüler (4%) einzuschätzen, die am Ende ihrer Grundschulzeit nicht in der Lage sind, einfache Aufgaben zu elementarem naturwissenschaftlichen Faktenwissen sicher zu lösen (Kompetenzstufe I). Die Anteile der Schülerinnen und Schüler auf den einzelnen Kompetenzstufen liegen in Deutschland aber in etwa in der Größenordnung der Anteile im OECD- und im EU-Durchschnitt.

Im Vergleich zu TIMSS 2007 zeigt sich eine geringfügige Verschiebung der Verteilung in die Mitte: Während der Anteil der Kinder auf Kompetenzstufe V geringfügig kleiner geworden ist, ist der Anteil der Schülerinnen und Schüler auf Kompetenzstufe III geringfügig größer geworden.

Naturwissenschaftliche Kompetenzbereiche: Die TIMSS-Rahmenkonzeption zur Erfassung naturwissenschaftlicher Kompetenzen erlaubt die differenzierte Darstellung von Ergebnissen in den drei Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geographie* sowie in den drei kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen*. Auf diese Weise können relative Stärken und Schwächen in den einzelnen Bereichen identifiziert werden.

Hinsichtlich der Inhaltsbereiche zeigen sich in Deutschland eine leichte relative Stärke im Bereich *Physik/Chemie* und eine leichte relative Schwäche im Bereich *Geographie*. Hinsichtlich der kognitiven Anforderungsbereiche zei-

gen sich im *Reproduzieren*, im *Anwenden* und auch im *Problemlösen* ähnlich hohe Leistungsmittelwerte der Schülerinnen und Schüler in Deutschland. Hier gibt es offensichtlich keine relativen Stärken oder Schwächen am Ende der Grundschulzeit.

Der Abstand zu den Staaten mit dem höchsten Niveau naturwissenschaftlicher Kompetenzen ist allerdings in allen drei Anforderungsbereichen groß. Im Vergleich zu TIMSS 2007 zeigen sich weder in einem der Inhaltsbereiche noch in einem der kognitiven Anforderungsbereiche signifikante Unterschiede.

Einstellungen und Selbstkonzept: In TIMSS werden nicht nur die durch Inhalts- und kognitive Anforderungsbereiche aufgespannten naturwissenschaftlichen Kompetenzen erfasst, sondern auch motivationale Orientierungen. Diese stellen neben den Kompetenzen einen wichtigen Zielbereich naturwissenschaftlichen Unterrichts dar. Da in Deutschland naturwissenschaftliche Inhalte in der Grundschule im Rahmen des integrativen Faches Sachunterricht unterrichtet werden, haben Grundschul Kinder oft noch keine Vorstellung davon, was naturwissenschaftliche Inhalte von anderen Inhalten des Sachunterrichts unterscheidet. Aus diesem Grund werden bei TIMSS nicht naturwissenschaftsbezogene, sondern auf den Sachunterricht bezogene Einstellungen und Selbstkonzepte der Schülerinnen und Schüler erfasst. Es zeigte sich, dass die Schülerinnen und Schüler am Ende der Grundschulzeit über insgesamt sehr positive Einstellungen zum Sachunterricht verfügen. 80 Prozent der Kinder befinden sich in der Gruppe der Schülerinnen und Schüler mit *hohen* positiven Einstellungen. Beim sachunterrichtsbezogenen Selbstkonzept zeigt sich ein ähnliches Bild: 80 Prozent der Schülerinnen und Schüler verfügen über ein *hohes* sachunterrichtsbezogenes Selbstkonzept.

Sowohl die Einstellungen zum Sachunterricht als auch das sachunterrichtsbezogene Selbstkonzept korrelieren schwach, aber signifikant positiv mit der naturwissenschaftlichen Kompetenz. Kinder, die sich auf den oberen Kompetenzstufen befinden, haben also eher auch positive Einstellungen und ein positives Selbstkonzept. Umgekehrt finden sich auf den unteren Kompetenzstufen eher auch Kinder mit weniger positiven Einstellungen und Selbstkonzepten. Bei dieser Gruppe von Kindern zeichnen sich möglicherweise sowohl leistungsbezogene als auch motivationale Probleme hinsichtlich der weiteren Auseinandersetzung mit Naturwissenschaften ab. Vergleicht man die Werte für Einstellungen und Selbstkonzept von 2011 mit denen von 2007, so zeigen sich bei den Einstellungen keine Unterschiede und beim Selbstkonzept geringfügig höhere Werte in 2011. Insgesamt sind die Einstellungen zum Sachunterricht und die sachunterrichtsbezogenen Selbstkonzepte der Schülerinnen und Schüler in Deutschland sehr positiv ausgeprägt, was erwarten lässt, dass die Bereitschaft, sich aktiv mit naturwissenschaftlichen Fragestellungen auch nach der Grundschulzeit auseinanderzusetzen, hoch ist.

Literatur

- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.
- Appleton, K. (2007). Elementary Science Teaching. In S. K. Abell & N. G. Lederman (Hrsg.), *Handbook of research on science education* (S. 493–535). Mahwah, NJ: Lawrence Erlbaum Associates.

- Bayerisches Staatsministerium für Arbeit und Sozialordnung, Familie und Frauen/Staatsinstitut für Frühpädagogik. (2006). *Der Bayerische Bildungs- und Erziehungsplan für Kinder in Tageseinrichtungen bis zur Einschulung* (2. Aufl.). Weinheim: Beltz.
- Blaseio, B. (2004). *Entwicklungstendenzen der Inhalte des Sachunterrichts. Eine Analyse von Lehrwerken von 1970 bis 2000*. Bad Heilbrunn: Klinkhardt.
- Blaseio, B. (2009a). Natur in den Bildungsplänen des Elementarbereichs. In R. Lauterbach, H. Giest & B. Marquardt-Mau (Hrsg.), *Lernen und kindliche Entwicklung. Elementarbildung und Sachunterricht* (S. 85–92). Bad Heilbrunn: Klinkhardt.
- Blaseio, B. (2009b). Neue Entwicklungstendenzen der Inhalte des Sachunterrichts. *Zeitschrift für Grundschulforschung*, 2 (1), 117–131.
- Bonsen, M., Frey, K. A. & Bos, W. (2008). Soziale Herkunft. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 141–156). Münster: Waxmann.
- Bonsen, M., Kummer, N. & Bos, W. (2008). Schülerinnen und Schüler mit Migrationshintergrund. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 157–175). Münster: Waxmann.
- Bybee, R. (2002). Toward an understanding of scientific literacy. In R. Evans, W. Gräber, T. Koballa & P. Nentwig (Hrsg.), *Scientific Literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (S. 37–68). Opladen: Leske + Budrich.
- Einsiedler, W. (1998). The curricula of elementary science education in Germany. In Japanisch-Deutsches Zentrum (Hrsg.), *Mathematics and elementary science education* (S. 25–40). Berlin: JDZB.
- Einsiedler, W., Martschinke, S. & Kammermeyer, G. (2008). Die Grundschule zwischen Heterogenität und gemeinsamer Bildung. In J. Baumert, K. S. Cortina, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (S. 325–374). Reinbek bei Hamburg: Rowohlt.
- GDSU – Gesellschaft für Didaktik des Sachunterrichts. (2002). *Perspektivrahmen Sachunterricht*. Bad Heilbrunn: Klinkhardt.
- Guay, F., Marsh, H. W. & Boivin, M. (2003). Academic self-concept and academic achievement. Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, 95 (1), 124–136.
- Kitayama, S., Markus, H. R., Matsumoto, H., & Norasakkunkit, V. (1997). Individual and collective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan. *Journal of Personality & Social Psychology*, 72, 1245–1267.
- 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.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2004). *Gemeinsamer Rahmen der Länder für die frühe Bildung in Kindertageseinrichtungen*. Zugriff am 26.10.2012 unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_06_03-Fruehe-Bildung-Kindertageseinrichtungen.pdf.
- Koballa, T. R. & Glynn, S. M. (2007). Attitudinal and motivational constructs in science learning. In S. K. Abell & N. G. Lederman (Hrsg.), *Handbook of research on science education* (S. 75–102). Mahwah, NJ: Lawrence Erlbaum Associates.
- Köller, O. (2004). *Konsequenzen von Leistungsgruppierungen*. Münster: Waxmann.
- Marsh, H. W. & Yeung, A. S. (1997). Coursework Selection: Relations to Academic Self-Concept and Achievement. *American Educational Research Journal*, 34 (4), 691–720.
- MSW – Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2008). *Richtlinien und Lehrpläne für die Grundschule in Nordrhein-Westfalen*. Frechen: Ritterbach.
- MSW – Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2009). *Gesetz über die Ausbildung für Lehrämter an öffentlichen Schulen (Lehrerausbildungsgesetz – LABG) vom 12. Mai 2009*. Zugriff am 17.10.2012 unter <http://www.schulministerium.nrw.de/ZBL/Reform/LABG.pdf>.

- Möller, K. (2009). Was lernen Kinder über Naturwissenschaften im Elementar- und Primarbereich? Einige kritische Bemerkungen. In R. Lauterbach, H. Giest & B. Marquardt-Mau (Hrsg.), *Lernen und kindliche Entwicklung. Elementarbildung und Sachunterricht* (S. 165–172). Bad Heilbrunn: Klinkhardt.
- Möller, K., Kleickmann, T. & Sodian, B. (2011). Naturwissenschaftlich-technischer Lernbereich. In W. Einsiedler, M. Götz, F. Heinzel, J. Kahlert & U. Sandfuchs (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (1. Aufl., S. 509–517). Stuttgart: UTB.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y. & Preuschoff, C. (2009). *TIMSS 2011 Assessment Framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Prenzel, M., Geiser, H., Langeheine, R. & Lobemeier, K. (2003). Das naturwissenschaftliche Verständnis am Ende der Grundschule. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 143–187). Münster: Waxmann.
- Prenzel, M., Schöps, K., Rönnebeck, S., Senkbeil, M., Walter, O., Carstensen, C. H. & Hammann, M. (2007). Naturwissenschaftliche Kompetenz im internationalen Vergleich. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 63–105). Münster: Waxmann.
- Prenzel, M., Schütte, K. & Walter, O. (2007). Interesse an den Naturwissenschaften. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 107–124). Münster: Waxmann.
- Roberts, D. A. (2007). Scientific literacy/science literacy. In S. K. Abell & N. G. Lederman (Hrsg.), *Handbook of research on science education* (S. 729–780). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rönnebeck, S., Schöps, K., Prenzel, M., Mildner, D. & Hochweber, J. (2010). Naturwissenschaftliche Kompetenz von PISA 2006 bis 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. 177–198). Münster: Waxmann.
- Seaton, M., Marsh, H. W. & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the Big-Fish-Little-Pond effect (BFLPE) across 41 culturally and economically diverse countries. *Journal of Educational Psychology*, 101 (2), 403–419.
- Seaton, M., Marsh, H. W. & Craven, R. G. (2010). Big-Fish-Little-Pond Effect: Generalizability and Moderation – Two Sides of the Same Coin. *American Educational Research Journal*, 47 (2), 390–433.
- Strunck, U., Lück, G. & Demuth, R. (1998). Der naturwissenschaftliche Sachunterricht in Lehrplänen, Unterrichtsmaterialien und Schulpraxis. *Zeitschrift für Didaktik der Naturwissenschaften*, 4 (1), 69–80.
- Wittwer, J., Saß, S. & Prenzel, M. (2008). Naturwissenschaftliche Kompetenz im internationalen Vergleich: Testkonzeption und Ergebnisse. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 87–124). Münster: Waxmann.



Kapitel V

Merkmale der Lehr- und Lernbedingungen im Primarbereich

Kerstin Drossel, Heike Wendt, Silvia Schmitz und Birgit Eickelmann

1 Einleitung

Im Mittelpunkt von TIMSS steht der internationale Vergleich der mathematischen und naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe. Dabei werden die Leistungen der getesteten Schülerinnen und Schüler als Ergebnis eines Lernprozesses betrachtet, der von vielfältigen Merkmalen gekennzeichnet und beeinflusst ist.

Um die verschiedenen Ebenen und wechselseitigen Verflechtungen der Bedingungen des schulischen Lernens darzustellen, wurde auf das Rahmenmodell in IGLU zurückgegriffen, da es zum einen die hier berichteten Merkmale in einen Ordnungsrahmen stellt (siehe Kapitel 2, Abschnitt 4.1.2 in Bos, Tarelli, Bremerich-Vos & Schwippert, 2012) und zum anderen die Möglichkeit bietet, Befunde der gemeinsamen IGLU- und TIMSS-Erhebung zu parallelisieren (Tarelli, Lankes, Drossel & Gegenfurtner, 2012). Das Modell veranschaulicht, dass drei große Bereiche von Bedingungen in engem Zusammenhang mit den schulischen Ergebnissen der Schülerinnen und Schülern stehen, auf die mit unterschiedlichem Wirkungsgrad Einfluss genommen werden kann: Im Zentrum aller Maßnahmen zur Qualitätssicherung und Qualitätsentwicklung stehen die innerschulischen Prozesse auf Ebene der Schule und des Unterrichts (Helmke, 2009). Die familiären und außerschulischen Kontexte spielen zwar eine entscheidende Rolle für die Leistungsergebnisse der Schülerinnen und Schüler, sie entziehen sich aber weitgehend dem Einfluss der Schule. Bildungspolitische Rahmenbedingungen lassen sich nur auf lange Sicht beeinflussen, und es erweist sich als schwierig, Zusammenhänge der meist strukturellen Entscheidungen mit dem Lernerfolg zu belegen (Klieme, Jude, Baumert & Prenzel, 2010).

Im Folgenden werden schulische Merkmale und Rahmenbedingungen beleuchtet, in denen das Lernen der Schülerinnen und Schüler der vierten Jahrgangsstufe stattfindet. Dabei werden in diesem Kapitel zwei Schwerpunkte gesetzt, bei de-

nen sich vermutlich Entwicklungen gegenüber der Erhebung im Jahr 2007 zeigen. Zum einen wird der Ausbau der Ganztagschullandschaft und zum anderen die Ausstattung und Nutzung digitaler Medien betrachtet. Im Kapitel „Lehr- und Lernbedingungen an Grundschulen im internationalen Vergleich“ (Tarelli et al., 2012) in „IGLU 2011 – Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich“ (Bos et al., 2012) liegt die Fokussierung hingegen auf den äußeren schulischen Rahmenbedingungen wie Klassengröße und Unterrichtszeit, dem Schulleitungshandeln sowie spezifischen Aspekten des Leseunterrichts.

Die Merkmale von Lehr- und Lernbedingungen im Primarbereich in Deutschland lassen sich zum einen im internationalen Vergleich darstellen, zum anderen bietet die erneute Teilnahme Deutschlands an TIMSS die Möglichkeit, ausgewählte Schul- und Unterrichtsmerkmale im Vergleich von TIMSS 2007 und TIMSS 2011 zu beschreiben und somit Veränderungen und Entwicklungen der vergangenen fünf Jahre einzuschätzen. Für die Interpretation der Ergebnisse ist allerdings zu beachten, dass das querschnittliche Erhebungsdesign in TIMSS keine kausalen Interpretationen zur Erklärung von Kompetenzergebnissen zulässt, das heißt, Aussagen zu Wirkungen der Rahmenbedingungen auf das Lernen beziehungsweise die Kompetenzentwicklung der Schülerinnen und Schüler lassen sich nicht ableiten. Vielmehr entspricht die Darstellung der Ergebnisse einer Momentaufnahme der Lernumgebungen zum Zeitpunkt der Testung, die als Information zum Verständnis der berichteten Leistungsergebnisse herangezogen werden kann. Darüber hinaus schafft der Vergleich mit anderen Teilnehmerstaaten einen Referenzrahmen auf internationaler Ebene. Als Vergleichsgruppe wurden für dieses Kapitel Teilnehmerstaaten, die Mitglieder der *Europäischen Union* (EU) sind, ausgewählt (VG_{EU}).

1.1 Grundschule in Deutschland

Als eigenständige Schulform mit einem ausgeprägten pädagogischen Profil bildet die Grundschule die Primarstufe des deutschen Bildungssystems. Als erster institutionalisierter Bildungseinrichtung kommen ihr im deutschen Bildungssystem zwei tragende Funktionen zu (Schorch, 2007; Valtin, 2006): Zum einen werden eigenständige pädagogische Zielsetzungen verfolgt, welche Schülerinnen und Schülern die Erschließung ihrer Lebenswirklichkeit ermöglichen und sie bei der Bewältigung ihrer eigenen Lebensaufgaben unterstützen sollen. Zum anderen bereitet die Grundschule Schülerinnen und Schüler auf die Angebote und Anforderungen der weiterführenden Schulen vor (vgl. Tarelli et al., 2012).

Anders als die horizontal gegliederte Sekundarstufe zeichnet sich die Grundschule dadurch aus, dass sie grundsätzlich von allen Kindern eines Jahrgangs besucht wird. Eine Ausnahme bilden hierbei diejenigen Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf, die nicht an Regelschulen unterrichtet werden. Sie machen einen Anteil von drei bis vier Prozent der Gesamtschülerschaft im Primarbereich aus (Statistisches Bundesamt, 2011). In Zahlen betrachtet besuchten im Schuljahr 2010/2011 in Deutschland 2 837 737 Schülerinnen und Schüler eine Grundschule, die sich auf 134 323 Klassen in 16 290 Grundschulen aufteilen (Statistisches Bundesamt, 2012).

2 Bildungspolitische Merkmale und äußere schulische Rahmenbedingungen

Bildungspolitische Merkmale und äußere schulische Rahmenbedingungen können von der Einzelschule selbst nicht beeinflusst werden, geben dieser jedoch den Rahmen vor, in welchem sie agieren kann. Dazu zählen unter anderem der Umfang von bereitgestellten finanziellen Mitteln, die Schul- und Klassengrößen, die personellen Ressourcen und die Unterrichtszeit. Diese Rahmenbedingungen variieren je nach bildungspolitischen Vorgaben und der Situation der öffentlichen Gesamthaushalte.

2.1 Bildungsausgaben

Die Bildungsausgaben als zentrale Größe bildungspolitischer Rahmenbedingungen bestimmen mit, wie und in welchem Umfang äußere schulische Rahmenbedingungen, beispielsweise die Klassengröße, zusätzliches pädagogisches Personal und die Unterrichtszeit gestaltet werden können. Die Höhe der Bildungsausgaben ist nicht unabhängig von der wirtschaftlichen Situation eines Staates und den zur Verfügung stehenden finanziellen Ressourcen. Gleichzeitig kann die Höhe der Bildungsausgaben als Indikator für den Stellenwert gesehen werden, der Bildung in einer Gesellschaft beigemessen wird, stehen sie doch in unmittelbarem Zusammenhang zu individueller, gesellschaftlicher und wirtschaftlicher Entwicklung (Autorengruppe Bildungsberichterstattung, 2012). Um einen europäischen Vergleich der Bildungsausgaben zu ermöglichen, bietet es sich an, die jeweiligen Anteile am Bruttoinlandsprodukt (BIP) zu vergleichen.

Deutschland investiert lediglich 0.6 Prozent des BIP in primäre Bildung. Für die sekundäre und tertiäre Bildung stehen 3.3 Prozent des BIP zur Verfügung. Im Vergleich mit den europäischen Teilnehmerstaaten wird deutlich, dass die Bildungsausgaben für den Bereich der primären Bildung in Deutschland erheblich geringer ausfallen als in den meisten anderen europäischen Staaten der Vergleichsgruppe (UNESCO *Institute for Statistics*, 2012). Deutschland liegt gemeinsam mit der Slowakei und der Tschechischen Republik deutlich unter dem Mittelwert der Vergleichsgruppe EU (1.2%) (vgl. Tarelli et al., 2012). Wesentlich höhere Investitionen in die primäre Bildung leistet sich Dänemark (1.9%).

2.2 Klassengröße

Nach wie vor ist die Klassengröße, das heißt, die Anzahl der Schülerinnen und Schüler in einer Klasse, ein regelmäßig kontrovers diskutiertes Thema. Forschungsbefunde hinsichtlich des Zusammenhangs der Klassengröße und der Schülerleistung zeigen ein uneinheitliches Bild, allerdings lässt sich die These, dass sich die Klassengröße auf die Schülerleistungen auswirkt, auf Basis empirischer Befunde bisher nicht zuverlässig bestätigen (Arnold, 2005; Lankes & Carstensen, 2010). Insgesamt kann davon ausgegangen werden, dass die Größe einer Klasse bestimmte lernförderliche und damit auch leistungssteigernde Maßnahmen im Unterricht ermöglichen kann, jedoch nicht zwingend herbeiführt (Arnold, 2005; Helmke, 1997; Lankes & Carstensen, 2010; Ramseger, 2007; von Saldern, 2011). Von einem direkten Zusammenhang im Sinne von „je kleiner die

Klasse, desto größer die Fachleistung der Schülerinnen und Schüler“ kann aber nicht pauschal gesprochen werden.

In Deutschland beträgt die mittlere Klassengröße für vierte Klassen an Grundschulen nach Angaben der Lehrkräfte 21.6 Schülerinnen und Schüler. Hinsichtlich der Unterschiedlichkeit der Klassengröße in Deutschland wird deutlich, dass nach Angaben der Lehrkräfte lediglich 14.7 Prozent der Viertklässlerinnen und Viertklässler in Klassen mit mehr als 25 Schülerinnen und Schülern unterrichtet werden. Klassengrößen mit 31 Schülerinnen und Schülern stellen in Deutschland eine Ausnahme dar. In der Vergleichsgruppe EU variiert die mittlere Klassengröße zwischen 26.7 Schülerinnen und Schülern in England und 19.3 Schülerinnen und Schülern in Österreich. Der Mittelwert der VG_{EU} liegt bei 21.7 Kindern je Klasse. Damit liegt Deutschland leicht unterhalb des Mittelwerts der europäischen Vergleichsgruppe (siehe auch Tarelli et al., 2012).

2.3 Unterrichtszeit

In Grundschulen in Deutschland findet durchschnittlich an 195 Tagen im Jahr Unterricht statt. Dieser Wert liegt etwas über dem Mittelwert der VG_{EU} (187 Tage). Für die Jahresunterrichtszeit in der vierten Jahrgangsstufe in Deutschland ergibt sich ein Mittelwert von 863 Unterrichtsstunden pro Jahr. Deutschland liegt damit etwas über dem Durchschnitt der VG_{EU} (860 Stunden) (vgl. Tarelli et al., 2012).

2.4 Einschulungsalter

Die Kinder, die 2011 in der vierten Jahrgangsstufe waren, sind in der Regel im Jahr 2007 eingeschult worden. Nach retrospektiven Angaben der Eltern ist der Großteil der Schülerinnen und Schüler in Deutschland zu Beginn der Schulzeit 6 Jahre alt (77.8%). 16.1 Prozent der Kinder sind nach Angaben der Eltern bei der Einschulung 7 Jahre alt, 5.9 Prozent sind 5 Jahre oder jünger und 0.2 Prozent sind 8 Jahre oder älter.

2.5 Ausbaustand und außerschulische Angebote der Ganztagschulen in Deutschland

Der Ausbaustand der Ganztagschullandschaft kann sowohl als bildungspolitische als auch als äußere schulische Rahmenbedingung betrachtet werden. Traditionell ist das deutsche Schulsystem als Halbtagschule organisiert. In den letzten Jahren gab es jedoch vielfältige Bestrebungen, die Ganztagschullandschaft auszubauen. Daher wurden in den vergangenen Jahren erhebliche finanzielle Ressourcen von Bund und Ländern, beispielsweise durch das Programm „Zukunft, Bildung und Betreuung“ (IZBB), in den Ausbau der Ganztagslandschaft investiert (BMBF, 2008). Dies entspricht auch den Wünschen der Eltern: Wie die neue JAKO-O Elternbefragung zeigt, wünschen sich 70 Prozent der Eltern eine Ganztagschule und nur 28 Prozent eine Halbtagschule für ihr Kind (Valtin, 2012).

Holtappels (2009) nennt für diese Entwicklung vier schultheoretisch bedeutsame Begründungslinien: Zum einen wird die verstärkte Entwicklung von

Ganztagsschulen als Antwort auf die gewandelten Bildungsanforderungen angeführt, da die zusätzliche Zeit beispielsweise für die Vermittlung wichtiger fächerübergreifender Schlüsselkompetenzen genutzt werden kann. Zum anderen stellt die Etablierung von Ganztagsschulen einen soziokulturellen Beitrag dar, da sie zeitlich geregelte Betreuungsangebote gewährleistet und somit die Vereinbarkeit von Familie und Beruf unterstützt. Ein weiteres Argument für die Etablierung von Ganztagsschulen ist die Möglichkeit, mit Hilfe der zusätzlichen Zeit eine Lern- und Förderkultur zu entwickeln, die unter anderem durch verstärkte sozialpädagogische Hilfen und individuelle Förderung dazu beitragen soll, Schulversagerquoten zu reduzieren oder der anhaltenden Bildungsbenachteiligung von Kindern aus Familien mit einem niedrigen sozioökonomischen Status und Kindern mit Migrationshintergrund entgegenzutreten. Darüber hinaus führt Holtappels (2009) die sozialerzieherische Funktion von Ganztagsschulen an, welche die Stärkung der Schule bezüglich der Erziehungsaufgabe der Schülerinnen und Schüler impliziert, die im Zuge der gewandelten außerschulischen Sozialisationsbedingungen und der Pluralisierung der Lebensformen an Bedeutung gewonnen hat. Empirische Belege für die Wirksamkeit von Ganztagschulen auf die Leistung der Schülerinnen und Schüler liegen allerdings nicht vor. Strietholt, Manitiuis, Berkemeyer und Bos (2012) finden in ihrer Untersuchung an Ganztagsschulen beispielsweise weder ein höheres Leistungsniveau noch eine Reduktion von Bildungsungleichheiten.

Laut Definition der *Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland* (KMK) gilt eine Schule als Ganztagschule, wenn an mindestens drei Tagen in der Woche ein ganztägiges Angebot von mindestens sieben Stunden für die Schülerinnen und Schüler bereitgestellt wird, an diesen Tagen ein Mittagessen angeboten wird und die Ganztagsangebote in einem konzeptionellen Zusammenhang mit dem Unterricht stehen (KMK, 2011).

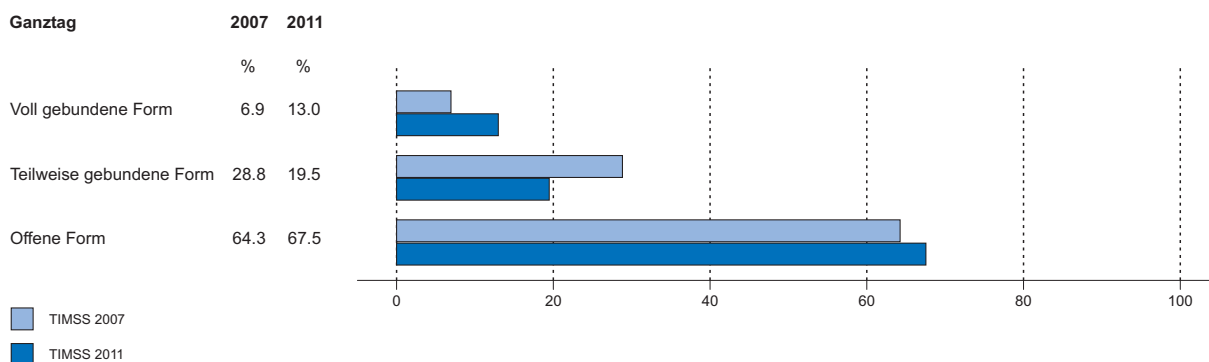
Konzeptionell werden drei Formen des Ganztagsbetriebs unterschieden, die sich hinsichtlich ihres Verbindlichkeitsgrads unterscheiden: Bei der *voll gebundenen* Form sind alle Schülerinnen und Schüler verpflichtet, an mindestens drei Tagen in der Woche an dem Ganztagsangebot teilzunehmen. Bei der *teilweise gebundenen* Form sind hingegen nur alle Schülerinnen und Schüler einer Klasse oder einer Klassenstufe zu einer Teilnahme verpflichtet. In der *offenen* Form nehmen lediglich einzelne Schülerinnen und Schüler auf Wunsch ihrer Eltern für mindestens ein Schulhalbjahr an dem ganztägigen Angebot teil (KMK, 2011). Die Definition und die beschriebenen Formen sind allerdings als Mindestkonsens zu betrachten, da sie sich landesspezifisch bei der Ausgestaltung sowie bei der Benennung unterscheiden.

Nach Angaben der KMK hat sich im Zuge des Ausbaus der Ganztagschullandschaft die Anzahl der öffentlichen schulischen Verwaltungseinheiten mit Ganztagsbetrieb in den vergangenen zehn Jahren mehr als verdoppelt. Jeder dritte Schulstandort in Deutschland bietet inzwischen einen Ganztagsbetrieb an (KMK, 2010). Am weitesten ausgebaut ist die Ganztagschullandschaft im Grundschulbereich. Im Schuljahr 2008/2009 besuchte jede fünfte Schülerin beziehungsweise jeder fünfte Schüler eine Grundschule mit Ganztagsangebot, während es im Schuljahr 2004/2005 noch 6.8 Prozent aller Grundschülerinnen und Grundschüler waren (KMK, 2010).

Laut Angaben der im Rahmen von TIMSS befragten Schulleitungen in Deutschland besuchen 2011 rund 46.8 Prozent der Schülerinnen und Schüler eine Ganztagsgrundschule. Bei TIMSS 2007 lag dieser Anteil lediglich bei 35.1 Prozent. Das entspricht einer Zunahme von 11.7 Prozentpunkten.

Inwieweit Unterschiede im Ausbau der Ganztagschulen hinsichtlich der drei Formen offen, teilweise gebunden und voll gebunden bestehen und wie sich die Entwicklung seit der vorherigen Erhebung im Jahr 2007 darstellt, ist in Abbildung 5.1 zu sehen. Seit dem Jahr 2007 hat bei den Kindern, die an einer Ganztagschule angemeldet sind, vor allem der Anteil an Schülerinnen und Schülern zugenommen, die eine voll gebundene Ganztagschule besuchen. Ebenfalls, wenn auch nur leicht um etwa 3 Prozentpunkte, ist der Anteil der Schülerinnen und Schüler angestiegen, die eine offene Form der Ganztagschule besuchen. Eine Abnahme ist hingegen für den Anteil der Schülerinnen und Schüler zu beobachten, die eine teilweise gebundene Form besuchen. Dieser Befund kann als ein Hinweis darauf betrachtet werden, dass dieses Modell als Übergang bei der Entwicklung von Ganztagschulen genutzt wurde. Die mit deutlichem Abstand am häufigsten vorhandene Form bleibt die offene Ganztagschule (67.5%).

Abbildung 5.1: Formen des Ganztags im Vergleich 2007 zu 2011 (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen in Prozent)



IEA: Trends in International Mathematics and Science Study

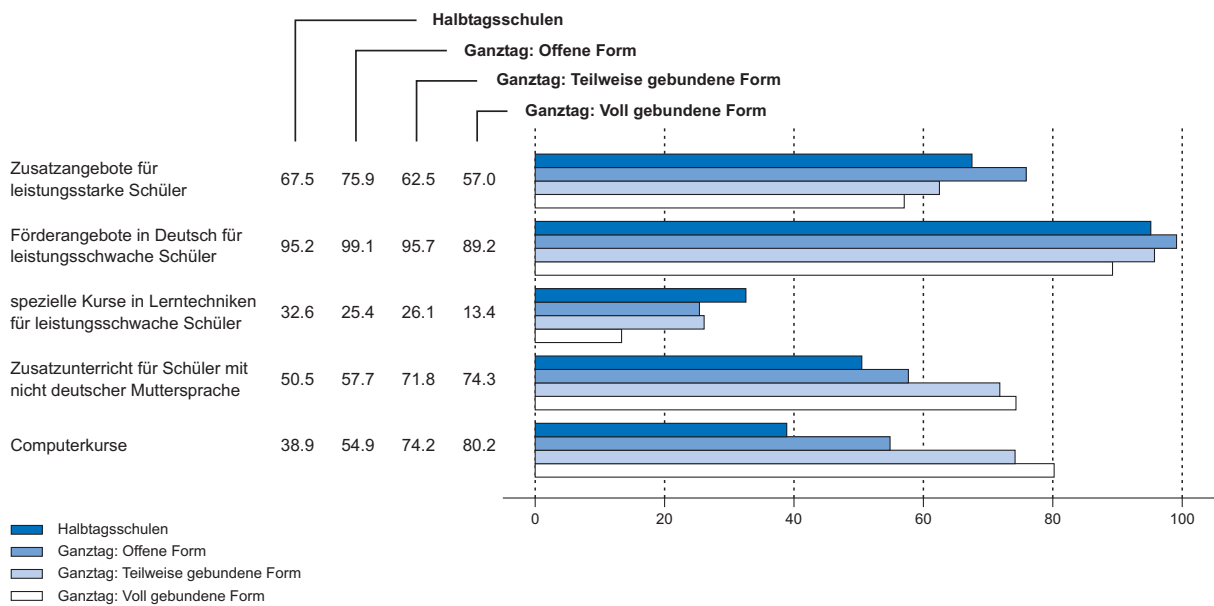
© TIMSS 2011

Dem Konzept von Ganztagschulen folgend soll die zusätzliche Zeit, die den Ganztagsgrundschulen im Gegensatz zu Halbtagschulen zur Verfügung steht, vor allem für erweiterte Lern- und Förderangebote für die Schülerinnen und Schüler genutzt werden. Wie im Folgenden ersichtlich wird, bieten viele Halbtagschulen diese erweiterten Angebote allerdings außerunterrichtlich an.

Inwieweit sich die Schulen hinsichtlich der Bereitstellung ihrer Angebote unterscheiden, zeigt Abbildung 5.2. Da Informationen über fachspezifische Angebote für Mathematik und Sachkunde nicht erhoben wurden, werden im Folgenden Angaben zu Angeboten berichtet, welche darauf abzielen, die Kompetenzen der Schülerinnen und Schüler fächerübergreifend zu fördern.

Mehr als drei Viertel der Schülerinnen und Schüler an offenen Ganztagschulen besuchen Schulen, an denen es nach Angaben der Schulleitungen Zusatzangebote für leistungsstarke Kinder gibt. Etwas geringer ist dieser Anteil an den Halbtagschulen (67.5%) und an den teilweise gebundenen Ganztagschulen (62.5%). Am geringsten fällt dieser Anteil mit 57.0 Prozent an voll gebundenen Ganztagsgrundschulen aus.

Nahezu alle Schülerinnen und Schüler, die Halbtagsgrundschulen oder teilweise gebundene und offene Ganztagschulen besuchen, haben Schulleitungen, die angeben, dass Förderangebote in Deutsch für leistungsschwache Kinder vorhanden sind. Lediglich an voll gebundenen Ganztagschulen liegt dieser Anteil unter 90 Prozent.

Abbildung 5.2: Erweiterte Lern- und Förderangebote in Ganz- und Halbtagsgrundschulen (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen in Prozent)

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Der Anteil der Schülerinnen und Schüler, denen nach Angaben der Schulleitungen spezielle Kurse in Lerntechniken für leistungsschwache Schülerinnen und Schüler angeboten werden, ist mit Abstand am geringsten an voll gebundenen Ganztagsgrundschulen (13,4%). Allerdings beträgt dieser an den offenen und teilweise gebundenen Schulformen jeweils auch nur etwa ein Viertel und bei der Halbtagschule etwa ein Drittel.

Über 70 Prozent der Schülerinnen und Schüler an voll und teilweise gebundenen Ganztagsgrundschulen und etwa 60 Prozent der Schülerinnen und Schüler an offenen Ganztagsgrundschulen besuchen nach Angaben der Schulleitungen Schulen mit Zusatzunterricht für Schülerinnen und Schüler mit nicht deutscher Muttersprache. Am geringsten fällt dieser Anteil an den Halbtagsgrundschulen aus (50,5%).

Die zahlenmäßig größten Unterschiede hinsichtlich der Verfügbarkeit erweiterter Lern- und Förderangebote zeigen sich bei Betrachtung des Angebots von Computerkursen. Über 80 Prozent der Schülerinnen und Schüler an voll gebundenen Ganztagsgrundschulen lernen an Schulen, an denen nach Angaben der Schulleitungen Computerkurse angeboten werden. Etwas geringer fällt dieser Anteil an teilweise gebundenen Grundschulen (74,2%) und wesentlich geringer an offenen Ganztags- (54,9%) und Halbtagsgrundschulen (38,9%) aus.

Insgesamt zeigt sich bei Betrachtung der vorhandenen zusätzlichen Lern- und Förderangebote kein einheitliches Bild hinsichtlich der Unterschiede zwischen den Halbtagsgrundschulen und den verschiedenen Formen der Ganztagsgrundschulen. Deutlich wird, dass auch Halbtagsgrundschulen über ein breites erweitertes Förderangebot verfügen. Entsprechend wird es wichtig sein vertiefend zu untersuchen, inwieweit eine unterschiedliche Nutzung stattfindet und ob sich Qualitätsunterschiede zeigen.

3 Schulinterne Merkmale

Schulinterne Bedingungen, beispielsweise Ausstattungsmerkmale, die im Folgenden näher betrachtet werden, stehen in einem komplexen Wirkungszusammenhang. Im Wesentlichen werden sie vom sozialen und kulturellen Kapital und dem Erziehungs- und Unterstützungsverhalten der Elternhäuser der Schülerinnen und Schüler sowie von den bildungspolitischen und äußeren schulischen Rahmenbedingungen beeinflusst. Die schulinternen Bedingungen beeinflussen hingegen indirekt über die Lernumgebung, in der Lernprozesse stattfinden, die Leistungsergebnisse der Schülerinnen und Schüler. Der genaue Wirkungszusammenhang ist allerdings komplex und Effekte können durch andere Rahmenbedingungen wie die regionale Lage der Schule, die Schülerzusammensetzung oder die elterliche Unterstützung mediiert sein. Eine gut ausgestattete Schule ist demnach kein Garant für gute schulische Leistungen.

3.1 Ausstattungsmerkmale

Ausstattungsmerkmale wie Schulbibliotheken, Labore für den naturwissenschaftlichen Sachunterricht sowie die Ausstattung mit Computern und Internet können wichtige Ressourcen für das Lernen in Mathematik und im naturwissenschaftlichen Sachunterricht darstellen und werden im Folgenden im europäischen Vergleich der Teilnehmerstaaten betrachtet.

In Deutschland besuchen rund 73 Prozent der Schülerinnen und Schüler Schulen, an denen eine Schulbibliothek vorhanden ist. Für Klassenbibliotheken liegt dieser Anteil in Deutschland bei 81.5 Prozent (weitere Angaben sowie Informationen über die Nutzung finden sich in Tarelli et al., 2012).

In Deutschland besuchen nur 12.8 Prozent der Schülerinnen und Schüler Schulen, welche über ein Labor für den naturwissenschaftlichen Sachunterricht verfügen. Damit liegt Deutschland unter dem europäischen Mittelwert von 22.5 Prozent. Im Vergleich zur relativ guten Ausstattung mit Schulbibliotheken gibt es im europäischen Vergleich der Teilnehmerstaaten insgesamt wenige Schülerinnen und Schüler, die Schulen besuchen, welche über ein Labor für naturwissenschaftlichen Sachunterricht verfügen. In Irland, Litauen und Nordirland stehen beispielsweise gar keine Labore für den naturwissenschaftlichen Sachunterricht zur Verfügung. In Dänemark besuchen hingegen nach Angaben der Schulleitungen 84.0 Prozent der Schülerinnen und Schüler Schulen, die über ein Labor verfügen.

Im Folgenden wird die Ausstattung mit neuen Medien fokussiert, da sich der Umgang mit neuen Medien mehr und mehr zu einer fächerübergreifenden Schlüsselkompetenz etabliert (Baumert, 2002; Schulz-Zander, Eickelmann & Goy, 2010). Zwar ist der Forschungsstand hinsichtlich des Zusammenhangs mit der Leistung der Schülerinnen und Schüler nicht eindeutig (siehe zusammenfassend Schulz-Zander et al., 2010), es lassen sich jedoch gesellschaftliche und pädagogische Begründungslinien für den Einsatz digitaler Medien im Unterricht in den Grundschulen aufzeigen, beispielsweise die „Vorbereitung der Kinder auf ein Leben in einer von Technologien geprägten Welt (...)“ (ebd., S. 91). Damit Schülerinnen und Schüler Medienkompetenz erwerben können, müssen Schulen dafür zunächst hinreichend ausgestattet sein.

Tabelle 5.1: Anzahl der Computer (absolute Zahlen) für Unterrichtszwecke und die Verfügbarkeit im Fachunterricht (in Prozent)

Teilnehmer		Anzahl der Computer für Unterrichtszwecke		Verfügbarkeit von Computern				Verfügbarkeit von Internetzugang			
		(4. Klasse)		Mathematikunterricht vorhanden ^B		Mathematikunterricht vorhanden ^B		Sachunterricht vorhanden ^B		Sachunterricht vorhanden ^B	
		Anzahl ^A	Anzahl ^A	2011	2007	2011	2007	2011	2007	2011	2007
2007	2011			%	%	%	%	%	%	%	%
	³ Belgien (Fläm. Gem.)	16.1	-	52.0	-	51.4	-	84.4	-	83.4	-
³	^{2,3} Dänemark	18.7	53.1	70.4	94.8	69.9	94.2	80.7	91.4	79.7	91.2
	Deutschland	15.1	11.9	57.5	53.6	46.6	37.6	60.9	64.2	55.1	48.7
^{1,3}	^{1,3} England	42.5	26.4	71.5	75.7	69.9	72.3	74.1	77.1	74.1	75.2
	Finnland	17.4	-	59.2	-	58.5	-	65.8	-	64.1	-
	Irland	11.6	-	55.4	-	52.6	-	62.2	-	59.9	-
²	³ Italien	21.4	19.0	25.4	30.8	16.5	15.6	31.5	25.0	24.9	17.3
²	² Litauen	14.0	11.4	43.9	39.0	38.4	23.5	49.5	37.4	44.1	18.6
¹	Malta	13.7	-	69.1	-	65.9	-	74.3	-	72.3	-
³	³ Niederlande	11.7	15.3	83.5	84.0	83.3	78.4	64.2	61.6	64.0	58.4
	³ Nordirland	24.1	-	75.7	-	75.6	-	78.1	-	78.1	-
²	² Österreich	6.2	7.0	61.7	69.5	51.7	43.2	72.6	74.2	64.0	51.4
	Polen	11.3	-	16.4	-	13.8	-	19.2	-	16.2	-
	Portugal	6.9	-	42.6	-	36.2	-	46.7	-	40.1	-
	Rumänien	17.0	-	23.9	-	17.2	-	27.8	-	20.5	-
	Schweden	11.6	11.6	59.7	66.9	57.8	65.6	67.7	76.8	65.9	75.3
	Slowakei	26.8	16.2	37.6	47.0	35.9	38.7	44.6	51.3	43.0	45.0
	Slowenien	21.1	20.4	31.5	39.1	27.9	35.6	40.7	53.1	37.5	45.7
²	² Spanien	21.8	-	35.9	-	32.6	-	39.5	-	38.2	-
²	² Tschechische Republik	22.1	22.2	49.8	58.9	47.7	47.0	52.8	53.6	52.1	49.9
	Ungarn	20.0	14.8	33.9	23.2	27.1	12.6	36.0	24.3	30.2	20.2
	VG EU	17.7	-	50.3	-	46.5	-	55.9	-	52.7	-

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Die Angaben stammen aus dem Schulleitungsfragebogen.

B = Die Angaben stammen aus dem Lehrerfragebogen.

- = Für diese Teilnehmer liegen für den Studienzyklus 2007 keine Ergebnisse vor.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Wie viele Computer in den Schulen vorhanden sind und auch für eine Nutzung im Fachunterricht zur Verfügung stehen, ist in Tabelle 5.1 dargestellt.

Wie aus Tabelle 5.1 ersichtlich wird, stehen in Deutschland nach Angaben der Schulleitungen für die Jahrgangsstufe 4 durchschnittlich etwa 15 Computer für Unterrichtszwecke zur Verfügung. In Relation zur Jahrgangsstufengröße betrachtet (vgl. Abschnitt 2.2) teilen sich in Deutschland damit durchschnittlich etwa vier Kinder einen Computer. Betrachtet man die Anzahl in Relation zur Klassengröße findet sich die geringste Computerausstattung in Österreich und in Portugal. Schülerinnen und Schüler in England besuchen nach Angaben der Schulleitung Schulen, die mit etwa durchschnittlich 43 Computern und in Relation zur Schüleranzahl in der Jahrgangsstufe 4 am besten ausgestattet sind.

Im Vergleich zu TIMSS 2007 lässt sich für Deutschland und die meisten weiteren Staaten der VG_{EU} feststellen, dass sich die Ausstattung mit Computern leicht verbessert hat, wobei der größte Anstieg mit durchschnittlich 16 zusätzlichen Computern pro Schule in England zu beobachten ist.

Bezüglich der Medienausstattung im Mathematikunterricht zeigt sich, dass 57.5 Prozent der Schülerinnen und Schüler Klassen besuchen, in denen nach Angaben der Lehrkräfte mindestens ein Computer zur Verfügung steht. Im Sachunterricht ist der Anteil etwas höher (60.9%). Im Mathematikunterricht stehen den Schülerinnen und Schülern in Deutschland durchschnittlich mehr Computer zur Verfügung als in der VG_{EU} (50.3%), ebenso wie im Sachunterricht (55.9%).

Die Verfügbarkeit von Computern im Mathematik- und Sachunterricht hat im Jahr 2011 im Vergleich zum Jahr 2007 in den meisten Staaten der VG_{EU}, für die bereits aus dem vorangegangenen Erhebungszyklus Werte vorliegen, abgenommen. Lediglich in Deutschland, Litauen und Ungarn ist der Anteil der Schülerinnen und Schüler, denen im Mathematikunterricht nach Angaben der Lehrkräfte Computer zur Verfügung stehen, in den letzten Jahren angestiegen, wobei mit elf Prozentpunkten der stärkste Anstieg in Ungarn zu verzeichnen ist. Im Sachunterricht ist ein Anstieg für Italien, Litauen, die Niederlande und Ungarn zu beobachten.

Der überwiegende Teil der Schülerinnen und Schüler in Deutschland besucht nach Angaben der Lehrkräfte Klassen, in denen die Computer für den Mathematik- und Sachunterricht auch über einen Internetanschluss verfügen. Im Gegensatz zur Computerverfügbarkeit ist der Anteil der Schülerinnen und Schüler, die Klassen besuchen, in denen ein Computer mit Internetzugang zur Verfügung steht, in den meisten Staaten der VG_{EU} im Vergleich zu 2007 angestiegen.

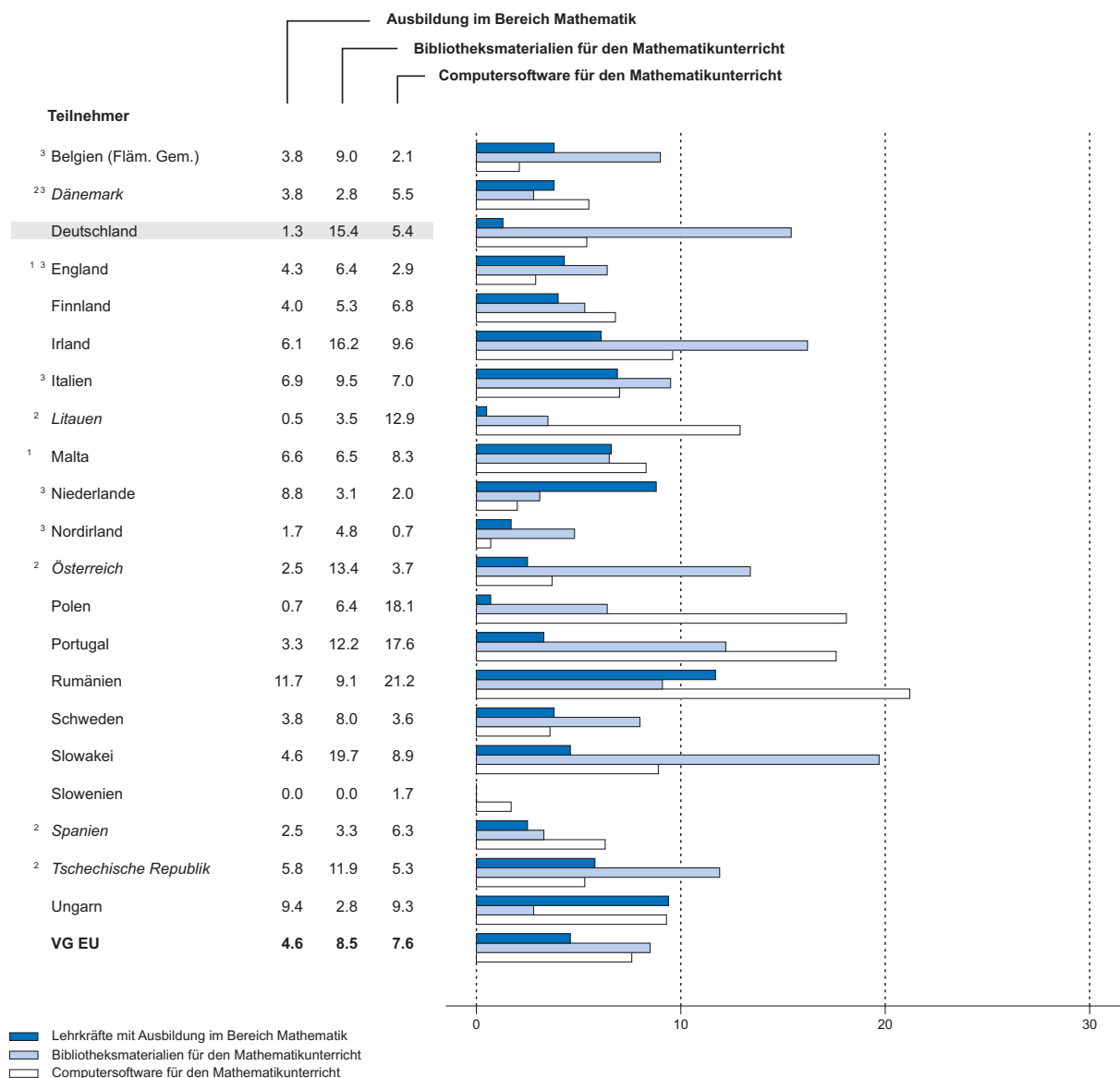
3.2 Probleme mit der Ausstattung

Neben Informationen zu den vorhandenen Ressourcen ist es auch wichtig zu berücksichtigen, inwieweit die Schülerinnen und Schüler an Schulen unterrichtet werden, in denen es Beeinträchtigungen des Unterrichts durch Knappheit oder Unzulänglichkeit gibt. Im Folgenden werden daher Ausstattungsprobleme jeweils für Mathematik und den naturwissenschaftlichen Sachunterricht betrachtet.

In Abbildung 5.3 ist der prozentuale Anteil der Schülerinnen und Schüler angegeben, die laut Angaben der Schulleitungen in den Bereichen ‚Lehrkräfte mit Ausbildung in Mathematik‘, ‚Bibliotheksmaterialien für den Mathematikunterricht‘ sowie ‚Computersoftware für den Mathematikunterricht‘ angeben, dass der Unterricht durch einen Mangel in diesen Bereichen sehr beeinträchtigt ist.

Nach Angaben der Schulleitungen sind lediglich 1.3 Prozent der Schülerinnen und Schüler in Deutschland von einer Beeinträchtigung des Unterrichts aufgrund eines Mangels an Lehrkräften mit einer Ausbildung im Bereich Mathematik betroffen. Lediglich in Slowenien (0.0%), Litauen (0.5%) und Polen (0.7%) ist dieser Anteil geringer. Der Mittelwert der europäischen Teilnehmerstaaten liegt bei 4.6 Prozent. Vergleichsweise hohe Beeinträchtigungen lassen sich in Rumänien (11.7%), Ungarn (9.4%) und den Niederlanden (8.8%) finden.

Wie aus Abbildung 5.3 ersichtlich ist, zeigt sich in Bezug auf die Bibliotheksmaterialien für Schülerinnen und Schüler in Deutschland ein vergleichsweise hohes Defizit für den Bereich Mathematik, obwohl Schulen in

Abbildung 5.3: Ausstattungsprobleme im Mathematikunterricht (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen in Prozent)


■ Lehrkräfte mit Ausbildung im Bereich Mathematik
■ Bibliotheksmaterialien für den Mathematikunterricht
■ Computersoftware für den Mathematikunterricht

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

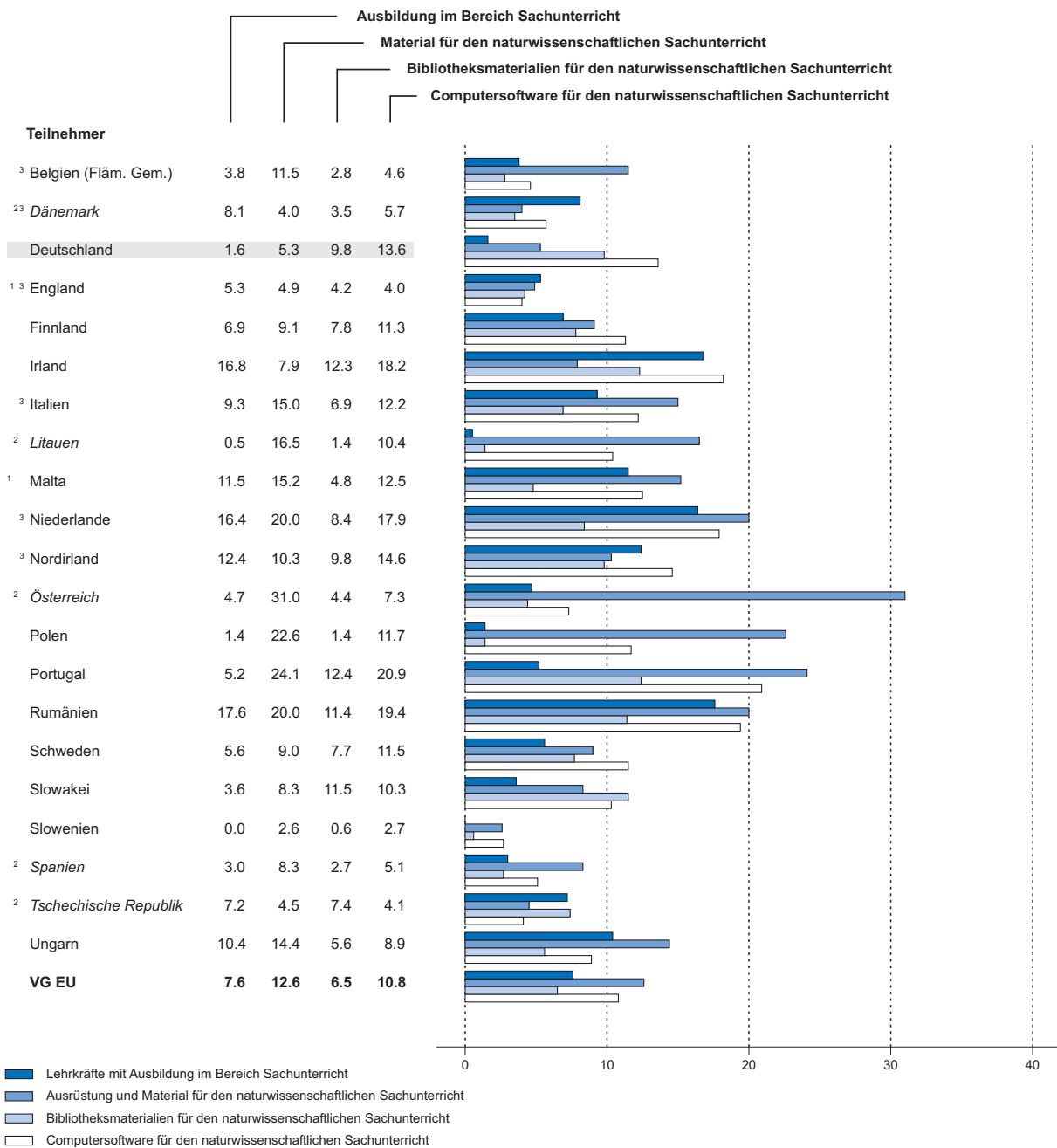
IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Deutschland relativ gut mit Schul- und Klassenbibliotheken ausgestattet sind (vgl. Tarelli et al., 2012). Nach Angaben der Schulleitungen sind über 15 Prozent der Schülerinnen und Schüler in Deutschland von einer starken Beeinträchtigung des Unterrichts durch fehlende Bibliotheksmaterialien im Bereich Mathematik betroffen. Lediglich in der Slowakei (19.7%) und Irland (16.2%) sind anteilig mehr Schülerinnen und Schüler durch einen Mangel in diesem Bereich beeinträchtigt. In Slowenien sind hingegen keine Schülerinnen und Schüler beeinträchtigt. Der europäische Mittelwert der Teilnehmerstaaten liegt in diesem Bereich bei 8.5 Prozent.

Von einem Mangel hinsichtlich der Computersoftware für den Mathematikunterricht sind in Deutschland nach Angaben der Schulleitungen 5.4 Prozent der Schülerinnen und Schüler betroffen. Der Mittelwert der europäischen Teilnehmer liegt bei 7.6 Prozent.

Abbildung 5.4: Ausstattungsprobleme im naturwissenschaftlichen Sachunterricht (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen in Prozent)



Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

In Abbildung 5.4 ist der prozentuale Anteil der Schülerinnen und Schüler angegeben, für die der naturwissenschaftliche Sachunterricht laut Angaben der Schulleitungen in den Bereichen ‚Lehrkräfte mit Ausbildung im Bereich Sachunterricht‘, ‚Ausrüstung und Material‘, ‚Bibliotheksmaterialien‘ sowie ‚Computersoftware‘ durch einen Mangel sehr beeinträchtigt ist.

Nach Angaben der Schulleitungen sind lediglich 1.6 Prozent der Schülerinnen und Schüler in Deutschland davon betroffen, dass es an Lehrkräften mit einer Ausbildung im Bereich Sachunterricht mangelt. Ebenso wie im Bereich Mathematik ist dieser Anteil lediglich in Slowenien (0.0%), Litauen (0.5%) und

Polen (1.4%) geringer (vgl. auch Abschnitt 6.2). Der Mittelwert der europäischen Teilnehmerstaaten liegt bei 7.6 Prozent. Hohe Beeinträchtigungen werden in Rumänien (17.6%), Irland (16.8%) und den Niederlanden (16.4%) berichtet, wobei der Anteil höher ist als für Mathematik.

Von einem starken Mangel an Ausrüstung und Material für den naturwissenschaftlichen Sachunterricht sind nach Angaben der Schulleitungen in Deutschland rund 5 Prozent der Schülerinnen und Schüler betroffen. Dieser Anteil liegt unter dem europäischen Mittelwert der Teilnehmerstaaten (12.6%). Hohe Beeinträchtigungen durch einen Mangel an Materialien gibt es nach Angaben der Schulleitungen in Österreich (31.0%), Portugal (24.1%) und Polen (22.6%). In Slowenien (2.6%) und Dänemark (4.0%) ist der geringste Anteil an Schülerinnen und Schülern in der VG_{EU} von einem Mangel im Bereich der Ausrüstung für den Sachunterricht beeinträchtigt.

In Deutschland sind nach Angaben der Schulleitungen rund 10 Prozent der Schülerinnen und Schüler durch eine Knappheit der Bibliotheksmaterialien für den naturwissenschaftlichen Sachunterrichts sehr beeinträchtigt. In Portugal (12.4%), Irland (12.3%), der Slowakei (11.5%) und Rumänien (11.4%) sind nach Einschätzung der Schulleitungen mehr Schülerinnen und Schüler betroffen. In Slowenien (0.6%) und Polen (1.4%) ist nach Angaben der Schulleitungen im Vergleich der Teilnehmer in der VG_{EU} der geringste Anteil von Schülerinnen und Schülern von einem Mangel in diesem Bereich im Unterricht beeinträchtigt. Wie auch bei den Bibliotheksmaterialien für den Mathematikunterricht liegt der Wert für Deutschland unter dem Durchschnitt der VG_{EU} (6.5%).

Durch einen Mangel hinsichtlich der Computersoftware für den Bereich des naturwissenschaftlichen Sachunterrichts sind nach Angaben der Schulleitungen in Deutschland 13.6 Prozent der Schülerinnen und Schüler stark im Unterricht beeinträchtigt. Der Mittelwert der europäischen Teilnehmer liegt bei 10.8 Prozent und ist demnach etwas höher als im Bereich der Beeinträchtigungen für die Bibliotheksmaterialien in Mathematik (7.6%).

Zusammenfassend kann festgestellt werden, dass nach Einschätzung der Schulleitungen für nur wenige Schülerinnen und Schüler von einer Beeinträchtigung des Unterrichts aufgrund von Ausstattungsmerkmalen auszugehen ist. Handlungsbedarf ergibt sich nach den Angaben der Schulleitungen in Deutschland insbesondere hinsichtlich der Ausstattung der Bibliotheken mit Büchern für den Mathematikunterricht und der Verfügbarkeit von Computersoftware für den naturwissenschaftlichen Sachunterricht.

4 Unterrichtsmerkmale

Bei Betrachtung der Ergebnisse internationaler Schulleistungsstudien stellt sich die Frage nach den Ursachen für die zu beobachtenden Unterschiede. Rahmenbedingungen und organisatorischen Merkmalen kommt hier eine deutlich geringere Bedeutung zu als Merkmalen des Unterrichts (Hattie, 2009). Auch wenn Leistungsstudien nicht primär das Ziel verfolgen, Unterricht zu erforschen, so wird doch stets die Gelegenheit genutzt, durch eine Befragung der Lehrkräfte Anhaltspunkte über die Unterrichtsgestaltung in den Teilnehmerstaaten zu erhalten.

Im Folgenden werden zunächst methodische Vorgehensweisen der Lehrkräfte im Unterricht beschrieben. Anschließend wird auf die Nutzung digitaler Medien im Unterricht eingegangen, bevor abschließend dargestellt wird, inwieweit die Lehrkräfte selber den Computer nutzen, um ihren Unterricht vorzubereiten.

4.1 Methodische Vorgehensweisen der Lehrkräfte im Unterricht

Für die Darstellung von grundlegenden methodischen Vorgehensweisen werden sowohl die Angaben der Mathematiklehrkräfte als auch die der Lehrkräfte für den naturwissenschaftlichen Sachunterricht genutzt, um ein fächerübergreifendes Bild der Lehrerhandlungen zu skizzieren (vgl. Abbildung 5.5). In Deutschland werden 31.8 Prozent der Schülerinnen und Schüler von Lehrkräften unterrichtet, die angeben, dass sie jede beziehungsweise fast jede Stunde zusammenfassen, was die Schülerinnen und Schüler im Unterricht gelernt haben sollten. Lediglich in Dänemark (19.2%), Schweden (23.5%) und Finnland (24.2%) ist dieser Anteil geringer. Hingegen werden 93.7 Prozent der Schülerinnen und Schüler in Litauen und 90.9 Prozent der Schülerinnen und Schüler in Ungarn von Lehrkräften unterrichtet, die in fast jeder Stunde den Unterrichtsstoff zusammenfassen. Diese Staaten liegen damit im Gegensatz zu Deutschland weit über dem europäischen Durchschnittswert der Teilnehmerstaaten von 63.3 Prozent.

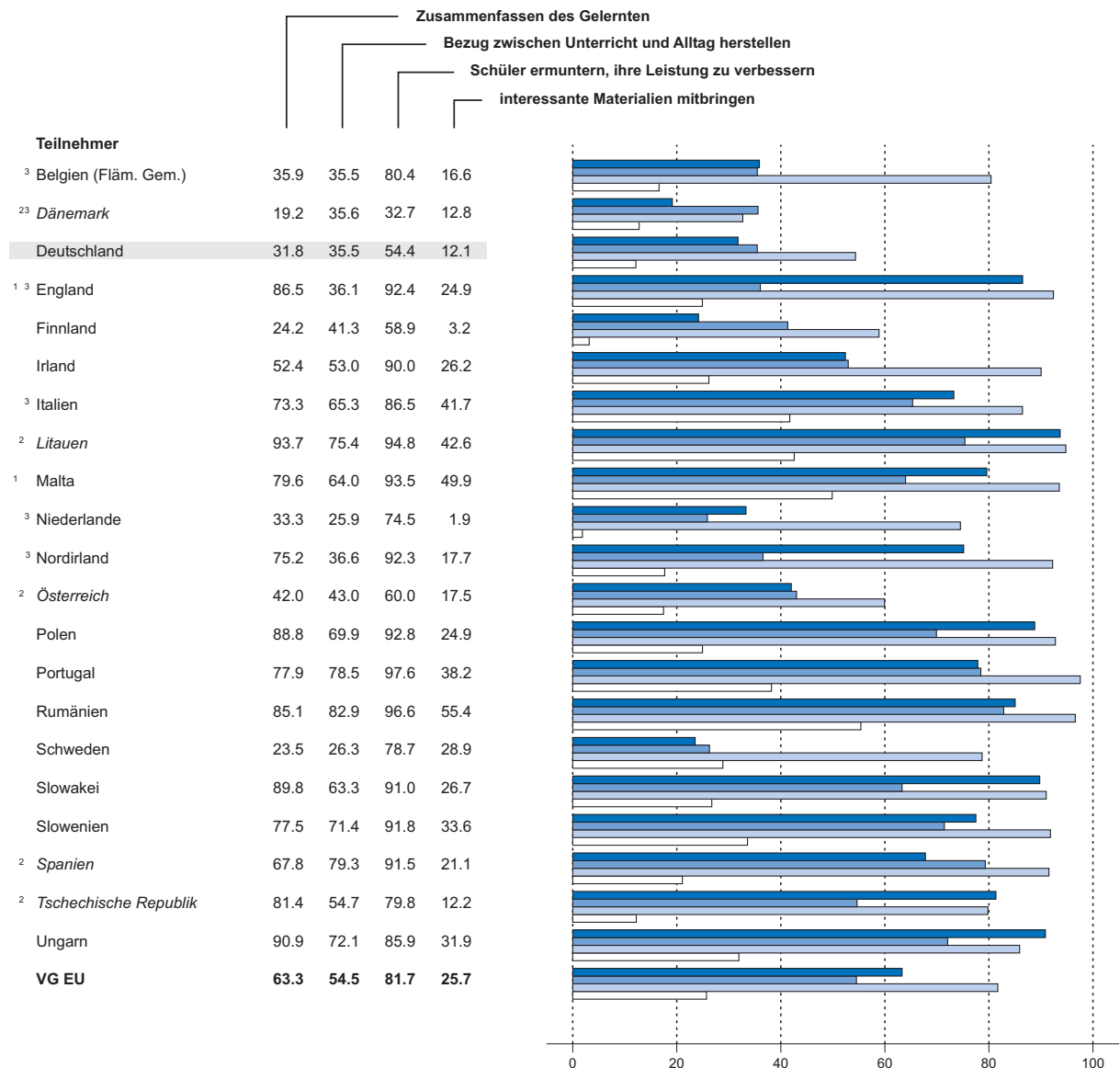
Ebenfalls deutlich unter dem Durchschnitt der europäischen Teilnehmerstaaten (54.6%) liegt der Anteil der Schülerinnen und Schüler in Deutschland, die von Lehrkräften unterrichtet werden, die in jeder Stunde einen Bezug zwischen Unterricht und dem Alltag der Schülerinnen und Schüler herstellen (35.5%). Lediglich in den Niederlanden (25.9%) und in Schweden (26.3%) werden anteilig weniger Schülerinnen und Schüler von Lehrkräften unterrichtet, die in fast jeder Stunde diesen Bezug herstellen. Hingegen ist dieser Anteil mit 82.9 Prozent am größten in Rumänien.

54.4 Prozent der Schülerinnen und Schüler in Deutschland werden von Lehrkräften unterrichtet, die in jeder beziehungsweise fast jeder Stunde die Schülerinnen und Schüler ermutigen, ihre Leistungen zu verbessern. Lediglich in Dänemark fällt dieser Anteil geringer aus (32.7%). In etwa der Hälfte der Staaten der europäischen Vergleichsgruppe werden hingegen über 90 Prozent der Schülerinnen und Schüler von Lehrkräften unterrichtet, die die Schülerinnen und Schüler in jeder Stunde ermutigen, ihre Leistungen zu verbessern, wobei dieser Anteil in Portugal am größten ist (97.6%).

In Deutschland werden 12.1 Prozent der Schülerinnen und Schüler von Lehrkräften unterrichtet, die angeben, dass sie jede beziehungsweise fast jede Stunde interessante Materialien in die Klasse mitbringen. Lediglich in den Niederlanden (1.9%) und Finnland (3.2%) ist dieser Anteil geringer. Hingegen werden im Vergleich der europäischen Teilnehmerstaaten anteilig die meisten Schülerinnen und Schüler in Rumänien von Lehrkräften unterrichtet, die angeben, jede beziehungsweise fast jede Stunde interessante Materialien mit in die Klasse zu bringen (55.4%). Der Durchschnitt der europäischen Teilnehmer liegt bei 25.7 Prozent in diesem Bereich und damit deutlich über dem Wert für Deutschland.

Zusammenfassend lässt sich in allen vier berichteten Vorgehensweisen feststellen, dass den Angaben der Lehrkräfte nach deutlich weniger Schülerinnen und Schüler in Deutschland von Lehrkräften unterrichtet werden, die in jeder beziehungsweise fast jeder Stunde Zusammenfassungen des Gelernten geben, einen Bezug zwischen den Unterrichtsinhalten und dem Alltag herstellen, die Schülerinnen und Schüler ermutigen, ihre Leistungen zu verbessern, und interessante Materialien mit in den Unterricht bringen, als der Durchschnitt der europäischen Vergleichsgruppe.

Abbildung 5.5: Methodische Vorgehensweisen der Lehrkräfte (Anteile der Schülerinnen und Schüler nach Angaben der Lehrkräfte in Prozent)



■ Zusammenfassen, was die Schüler im Unterricht gelernt haben sollten (jede/fast jede Stunde)
■ einen Bezug zwischen Unterricht und Alltag der Schüler herstellen (jede/fast jede Stunde)
■ alle Schüler ermuntern, ihre Leistung zu verbessern (jede/fast jede Stunde)
■ interessante Materialien in die Klasse mitbringen (jede/fast jede Stunde)

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

4.2 Computernutzung im Mathematik- und naturwissenschaftlichen Sachunterricht

In Deutschland nutzen nach Angaben der Lehrkräfte 5,8 Prozent der Schülerinnen und Schüler, die im Mathematikunterricht einen Computer zur Verfügung haben, diesen mindestens wöchentlich, um sich mit mathematischen Grundsätzen und Konzepten zu befassen. Am geringsten ist dieser Anteil in Polen (2,4%) und Ungarn (2,8%). Hingegen nutzt rund ein Drittel der Schülerinnen und Schüler in Nordirland, Malta und den Niederlanden den Computer mindestens einmal wöchentlich, um sich mit mathematischen Grundsätzen und Konzepten zu be-

fassen. Insgesamt liegt Deutschland unter dem Mittelwert der europäischen Teilnehmerstaaten von 11.4 Prozent.

Rund 17 Prozent der Schülerinnen und Schüler in Deutschland, denen ein Computer im Mathematikunterricht zur Verfügung steht, nutzen diesen mindestens wöchentlich, um mathematische Fertigkeiten und Prozeduren zu üben. Am geringsten ist der Anteil in Polen (3.3%) und mit Abstand am stärksten ausgeprägt in den Niederlanden (69.0%). Auch in diesem Bereich liegt Deutschland, wenn auch knapp, unter dem Durchschnitt der VG_{EU} von 20.2 Prozent.

Die Recherche nach Informationen als ein weiterer wichtiger Aspekt der Computernutzung wird von etwa 6 Prozent der Schülerinnen und Schüler in Deutschland mindestens wöchentlich im Mathematikunterricht praktiziert. Am geringsten ist der Anteil in Finnland (1.8%) sowie in Slowenien (3.2%) und am stärksten ausgeprägt hingegen in Nordirland (26.6%) und den Niederlanden (26.3%). Deutschland liegt auch hier unter dem europäischen Durchschnitt von 10.1 Prozent.

Hingegen nutzen nach Angaben der Lehrkräfte mehr als die Hälfte (54.1%) der Schülerinnen und Schüler in Deutschland, denen im naturwissenschaftlichen Sachunterricht ein Computer zur Verfügung steht, diesen mindestens wöchentlich im Fachunterricht, um nach Informationen zu suchen. Mit Abstand am seltensten nutzen Schülerinnen und Schüler in Polen (16.0%) den Computer im naturwissenschaftlichen Sachunterricht, um nach Informationen zu suchen. Der größte Anteil an Schülerinnen und Schülern, die nach Angaben der Lehrkräfte den Computer wöchentlich nutzen, findet sich in der Flämischen Gemeinschaft Belgiens (78.3%). Der europäische Durchschnitt liegt für diesen Bereich der Nutzung bei 49.8 Prozent, so dass die Schülerinnen und Schüler in Deutschland etwas darüber liegen.

Für den Bereich ‚Naturwissenschaftliche Verfahren und Experimente‘ geben die Lehrkräfte in Deutschland an, dass 14.8 Prozent der Schülerinnen und Schüler mindestens wöchentlich den Computer nutzen, um solche durchzuführen. In Polen ist dieser Anteil mit 7.4 Prozent der Schülerinnen und Schüler am geringsten und mit 46.9 Prozent in Irland am höchsten. Der Durchschnittswert der VG_{EU} liegt bei 23.4 Prozent.

4.3 Computernutzung von Lehrkräften

Inwieweit Lehrkräfte Computer im Unterricht einsetzen, hängt nicht ausschließlich mit der Verfügbarkeit zusammen, sondern auch damit, ob sie sich einen souveränen Einsatz zutrauen und diesbezüglich Unterstützung an ihrer Schule erfahren.

In Deutschland werden nach Angaben der Lehrkräfte rund zwei Drittel der Schülerinnen und Schüler, die den Computer im Unterricht nutzen, von Lehrkräften unterrichtet, die der Frage danach, ob sie sich beim Einsatz von Computern im Unterricht wohl fühlen, zustimmen. Im Durchschnitt werden rund drei Viertel der Schülerinnen und Schüler der europäischen Vergleichsgruppe, die den Computer im Unterricht nutzen, von Lehrkräften unterrichtet, die sich wohl dabei fühlen. Deutschland liegt damit unter dem Mittelwert der europäischen Teilnehmerstaaten.

Bei Betrachtung der Unterstützungsmöglichkeiten für Lehrkräfte bei technischen Problemen zeigt sich in Deutschland, dass nur 41.3 Prozent der Schülerinnen und Schüler von Lehrkräften unterrichtet werden, die der Frage zustimmen,

dass ihnen eine Mitarbeiterin oder ein Mitarbeiter bei technischen Problemen zur Unterstützung zur Verfügung steht. Lediglich in Spanien (35.4%) und Italien (20%) ist dieser Anteil geringer. In Nordirland fällt der Anteil hingegen am höchsten aus (83.0%). Der Mittelwert der VG_{EU} liegt bei 56.3 Prozent.

Auch bezüglich der Integration von Computern in den Unterricht werden in Deutschland anteilig weniger Schülerinnen und Schüler von Lehrkräften unterrichtet (43.2%), die zustimmen, dass sie Unterstützung in diesem Bereich erhalten, als in der europäischen Vergleichsgruppe. Wie bei der technischen Unterstützung gibt es in der europäischen Vergleichsgruppe lediglich zwei Staaten, in denen der Anteil der Schülerinnen und Schüler, die von Lehrkräften unterrichtet werden, die sich in diesem Bereich unterstützt fühlen, geringer ist (Spanien: 39.2%, Italien: 21.5%). Am höchsten ist der Anteil erneut in Nordirland (90.4%).

Insgesamt zeigt sich, dass die Einschätzungen der Lehrkräfte in Deutschland darüber, inwiefern sie sich einen souveränen Einsatz von Computern im Unterricht zutrauen und wie sie sich technisch sowie bei der Integration in den Unterricht unterstützt fühlen, geringer ausfallen als in der VG_{EU} , wobei die Lage in Deutschland besonders defizitär hinsichtlich des Fehlens externer Unterstützung anzusehen ist.

Inwieweit die Lehrkräfte den Computer nach eigenen Angaben tatsächlich im und für den Unterricht nutzen, ist in Abbildung 5.6 dargestellt.

Fast alle Schülerinnen und Schüler in Deutschland (97.2%) werden, wie in den meisten Teilnehmerstaaten der VG_{EU} , von Lehrkräften unterrichtet, die einen Computer zur Vorbereitung des Unterrichts nutzen. In Spanien ist dieser Anteil am geringsten (64.3%), in England (99.4%) und Dänemark (99.2%) am höchsten.

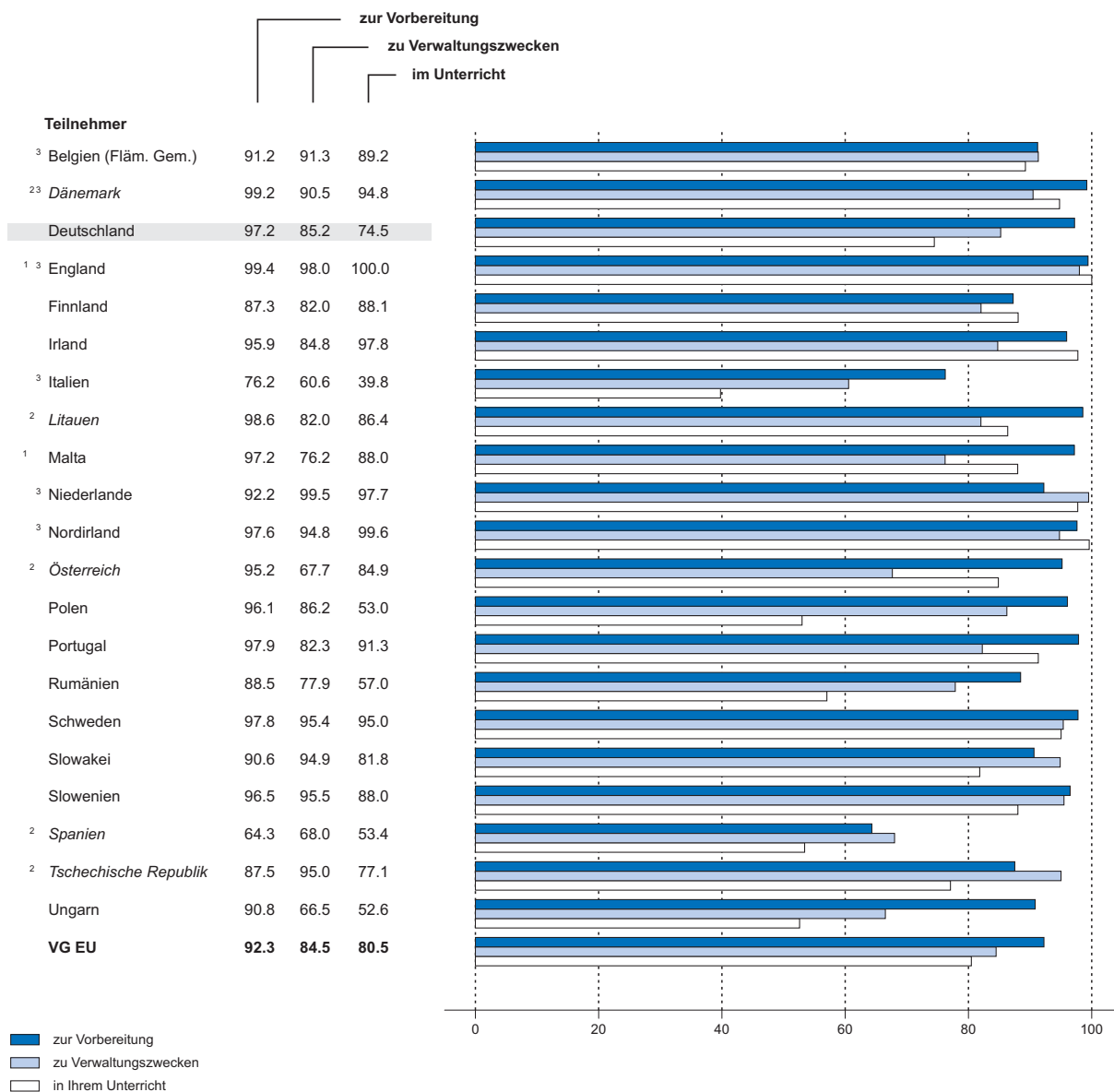
Etwas weniger Schülerinnen und Schüler werden in Deutschland von Lehrkräften unterrichtet, die den Computer zu Verwaltungszwecken nutzen (85.2%). Damit liegt Deutschland etwa im Bereich des Durchschnitts der europäischen Vergleichsgruppe (84.5%).

Hinsichtlich des Einsatzes von Computern im Unterricht zeigt sich folgendes Bild: In Deutschland werden drei Viertel der Schülerinnen und Schüler von Lehrkräften unterrichtet, die den Computer im Unterricht einsetzen. Damit liegt Deutschland unter dem Mittelwert der VG_{EU} (80.5%). Mit Abstand am geringsten ist dieser Anteil in Italien (39.8%). In England werden nach Angaben der Lehrkräfte sogar alle Schülerinnen und Schüler von Lehrkräften unterrichtet, die einen Computer im Unterricht einsetzen.

Beim Vergleich der Ausprägungen der drei Nutzungsformen wird deutlich, dass die meisten Schülerinnen und Schüler von Lehrkräften unterrichtet werden, die den Computer zur Unterrichtsvorbereitung und zu Verwaltungszwecken nutzen. Neben der direkten Nutzung im Unterricht ist davon auszugehen, dass besonders der Einsatz von Computern bei der Unterrichtsvorbereitung für die Kompetenzentwicklung der Schülerinnen und Schüler bedeutsam sein kann. Beispielsweise können Lehrkräfte den Computer nutzen, um Arbeitsmaterialien für den Unterricht zu erstellen oder auch um den Austausch mit anderen Lehrkräften zu erleichtern.

Computer werden im Unterricht hingegen deutlich seltener eingesetzt, wobei bei allen Angaben offen ist, wie oft und mit welcher Qualität Computer genutzt werden.

Abbildung 5.6: Computernutzung der Lehrkräfte (Anteile der Schülerinnen und Schüler nach Angaben der Lehrkräfte in Prozent)



Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

5 Klassenmerkmale

Bei der Zusammensetzung von Klassen wird eine starke Heterogenität der Schülerschaft vielfach als besondere Herausforderung für den schulischen Alltag und die Lern- und Vermittlungsprozesse angesehen (Gröhlich, Scharenberg & Bos, 2009). Im Folgenden sollen ausgewählte Angaben von Schulleitungen die Situation an Grundschulen in Deutschland im europäischen Vergleich skizzieren. Anschließend werden Einschätzungen der Eltern zu ähnlichen Kompositionsmerkmalen dargestellt. Für vertiefende Analysen zum soziokulturellen Hintergrund von Viertklässlerinnen und Viertklässlern sei auf die Kapitel 6 und 7 in diesem Band verwiesen.

Die erste Spalte der Tabelle 5.2 zeigt die Einschätzungen der Schulleitungen zu den ökonomischen Verhältnissen, aus denen die Schülerinnen und Schüler an ihrer Schule stammen. Um Schulen, in denen die wirtschaftliche Situation der Schülerfamilien eine besondere Herausforderung für den Schulalltag darstellen könnte, von Schulen mit günstigeren Bedingungen zu unterscheiden, werden die Anteile der Viertklässlerinnen und Viertklässler berichtet, die an Schulen unterrichtet werden, in denen mehr als die Hälfte der Schülerinnen und Schüler aus wirtschaftlich benachteiligten Verhältnissen stammen. Dabei muss einschränkend festgehalten werden, dass es sich bei dieser Angabe um eine subjektive Einschätzung der Schulleitungen handelt, was Vergleiche erschwert. Die Angaben verstehen sich eher als Orientierung für die Interpretation unterschiedlicher Leistungsergebnisse, die in diesem Band berichtet werden.

Es zeigt sich, dass in Deutschland 11.6 Prozent der Viertklässlerinnen und Viertklässler an Grundschulen unterrichtet werden, an denen nach Angaben der Schulleitungen mindestens die Hälfte der Schülerschaft aus wirtschaftlich benachteiligten Elternhäusern stammt. In der VG_{EU} erweist sich die Situation in den einzelnen Teilnehmerstaaten erwartungsgemäß als sehr unterschiedlich. Während die Anteile in Finnland (4.2%), der Flämischen Gemeinschaft in Belgien (4.7%) und Dänemark (5.7%) eher gering sind, fällt der Anteil in Rumänien (48.0%) sehr hoch aus.

Ein weiterer Aspekt sind die sprachlichen Voraussetzungen der Schülerinnen und Schüler. Für Schulen, an denen eine Vielzahl von Schülerinnen und Schülern nicht über ausreichend sprachliche Kompetenzen verfügt, kann diese Situation zur besonderen Herausforderung werden, sofern keine oder unzureichend kompensierende Maßnahmen vorhanden sind. In der dritten und vierten Spalte der Tabelle 5.2 wird anhand der Auskünfte der befragten Schulleitungen aufgezeigt, wie hoch die Anteile von Schülerinnen und Schülern in den jeweiligen Teilnehmerstaaten sind, die an Schulen lernen, in denen der Anteil an Kindern, deren Muttersprache die jeweilige Testsprache ist, als vergleichsweise hoch beziehungsweise niedrig einzuschätzen ist.

In Deutschland besuchen 13.4 Prozent der Schülerinnen und Schüler Schulen, an denen für weniger als die Hälfte der Kinder Deutsch die Muttersprache ist. Dieser Anteil liegt etwas über dem Mittelwert der VG_{EU} (12.5%). In einigen Teilnehmerstaaten sind die Anteile deutlich geringer, so zum Beispiel in Polen, wo es nicht vorkommt, dass an einer Schule für mehr als 50 Prozent der Schülerinnen und Schüler die Testsprache nicht auch die Muttersprache ist. Höhere Anteile weisen zum Beispiel England (21.9%) und Malta (82.4%) auf. Allerdings bildet Malta hier eine Ausnahme: Die Testsprache war Englisch, die Unterrichtssprache ist vielfach maltesisch (Firman & Camilleri, 2012).

Eine weitere Perspektive ergibt sich, wenn man die Anteile an Schülerinnen und Schülern betrachtet, die an Schulen unterrichtet werden, an denen mehr als 90 Prozent der Kinder die jeweilige Testsprache als Muttersprache sprechen. In Deutschland besucht knapp die Hälfte der Viertklässlerinnen und Viertklässler eine Schule mit entsprechendem Anteil an Kindern. Geringere Anteile ergeben sich für Österreich (32.7%) und Malta (5.8%).

Insgesamt lässt sich für Deutschland festhalten, dass Zusammensetzungen von Schülerschaften mit einem überwiegenden Anteil aus ökonomisch benachteiligten Verhältnissen oder mit einem hohen Anteil an nicht oder nicht ausschließlich deutschsprachigen Kindern eher die Ausnahme darstellen, wobei hier eine vertiefende Analyse in der Bundesrepublik Deutschland notwendig wäre, um die Situation in den Ländern, Regionen beziehungsweise einzelnen Stadtteilen zu differenzieren.

Anders sieht es bei den vorschulischen Kenntnissen der Schülerinnen und Schüler in Deutschland aus, was allerdings im Aufbau des deutschen Bildungssystems begründet ist (siehe Abschnitt 1 und Tarelli et al., 2012). Als Indikatoren dienten das Schreiben von Zahlen und das eigenständige Zählen, wobei diese Fähigkeiten noch keinen Aufschluss darüber erlauben, ob Kinder tatsächlich über ein Zahlenverständnis verfügen. In Deutschland besuchen 40.1 Prozent der Schülerinnen und Schüler Schulen, an denen nach Angaben der Schulleitungen weniger als 25 Prozent der Schülerschaft über die Fähigkeit verfügt, Zahlen zu schreiben, wenn sie das erste Schuljahr beginnen. In Nordirland (76.3%), Österreich (57.1%), Portugal (48.5%), der Slowakei (52.5%), der Tschechischen Republik (43.6%) und Ungarn (61.6%) ist der Anteil von Schülerinnen und Schülern an Schulen, an denen weniger als 25 Prozent der Kinder zum Zeitpunkt der Einschulung bereits Zahlen schreiben können, höher als in Deutschland. Die geringsten Anteile in diesem Bereich finden sich in Dänemark (3.4%) und Spanien (0.5%).

Hingegen liegt der Anteil der Schülerinnen und Schüler in Deutschland, die Schulen besuchen, an denen mehr als die Hälfte der Kinder zum Zeitpunkt der Einschulung Zahlen schreiben kann, nach Angaben der Schulleitungen bei 27.6 Prozent. Geringer als in Deutschland fällt dieser Anteil in den Staaten aus, die bereits höhere Anteile an Schülerinnen und Schülern aufweisen, von denen weniger als 25 Prozent Zahlen schreiben können.

Für den Bereich des eigenständigen Zählens bis 100 geben die Schulleitungen an, dass 73.6 Prozent der Schülerinnen und Schüler in Deutschland Schulen besuchen, an denen weniger als 25 Prozent eigenständig zählen können, wenn sie eingeschult werden. Im Vergleich zur europäischen Vergleichsgruppe (58.6%) fällt dieser Anteil in Deutschland deutlich höher aus. Der Anteil der Schülerinnen und Schüler in Deutschland, die Schulen besuchen, an denen über die Hälfte der Schülerinnen und Schüler nach Angaben der Schulleitungen eigenständig Zählen kann, wenn sie in die Grundschule kommen, liegt bei 7.6 Prozent.

Die Schülerinnen und Schüler in Deutschland liegen damit sowohl im Bereich Zahlen schreiben als auch beim eigenständigen Zählen deutlich über (weniger als 25%) beziehungsweise unter (mehr als 50%) den Mittelwerten der europäischen Vergleichsgruppe.

Auch die Eltern der Schülerinnen und Schüler wurden zu ähnlichen Aspekten wie die Schulleitungen befragt. Die Ergebnisse werden ebenfalls zur Beschreibung der Schülerkomposition genutzt (siehe Tabelle 5.3). Eine Vergleichbarkeit zu den Schulleiterangaben ist allerdings aufgrund der unterschiedlichen Indikatoren und Antwortformate nur sehr eingeschränkt möglich.

Ein Indikator für die ökonomischen Verhältnisse der Elternhäuser ist der Berufsstand der Eltern. Dieser wurde von Wendt, Stubbe und Schwippert (2012) ausführlich beschrieben und wird im Folgenden genutzt, um die Anteile der Schülerinnen und Schüler zu identifizieren, die aus Elternhäusern kommen, in denen eher manuelle Tätigkeiten (zum Beispiel Hilfsarbeiter oder Fachkräfte in der Landwirtschaft) ausgeführt werden, und solchen, in denen Führungstätigkeiten (zum Beispiel Akademiker und Techniker) nachgegangen wird.

In Deutschland haben 14.0 Prozent der Schülerinnen und Schüler Eltern, die beruflich manuellen Tätigkeiten nachgehen. Lediglich in Nordirland (13.1%), Finnland (9.9%) und Schweden (7.0%) fällt dieser Anteil geringer aus. Mit Abstand am größten ist der Anteil in Rumänien (43.2%). Der überwiegende Teil der Schülerinnen und Schüler in Deutschland kommt hingegen aus Elternhäusern, deren berufliche Beschäftigungen als Führungstätigkeiten beschrieben werden

Tabelle 5.2: Aspekte der Schülerkomposition aus Sicht der Schulleitungen (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen in Prozent)

Teilnehmer	Ökonomische Verhältnisse wirtschaftlich benachteiligte Familien mehr als 50%	Prozentualer Anteil der Schüler, die die Testsprache als Muttersprache haben		Vorschulische Kenntnisse ^A Zahlen schreiben		Vorschulische Kenntnisse ^A Eigenständig zählen	
		weniger als 50%	mehr als 90%	weniger als 25%	mehr als 50%	weniger als 25%	mehr als 50%
³ Belgien (Fläm. Gem.)	4.7	12.4	51.6	35.1	38.8	90.4	4.8
^{2,3} Dänemark	5.7	0.6	95.0	3.4	85.3	14.1	63.3
Deutschland	11.6	13.4	49.3	40.1	27.6	73.6	7.6
^{1,3} England	25.2	21.9	56.4	28.3	56.4	38.9	35.6
Finnland	4.2	0.8	84.7	10.9	74.4	29.8	37.1
Irland	26.5	3.3	63.6	-	-	-	-
³ Italien	12.7	5.9	63.9	35.9	36.4	79.8	8.5
² Litauen	26.2	3.8	88.0	18.7	48.2	40.5	29.7
¹ Malta	9.1	82.4	5.8	20.8	50.3	56.6	22.8
³ Niederlande	4.5	9.8	74.9	9.8	67.8	67.4	11.2
³ Nordirland	16.3	4.4	88.1	76.3	2.6	91.1	1.9
² Österreich	7.6	15.5	32.7	57.1	8.7	84.5	4.0
Polen	12.8	0.0	100.0	17.9	55.5	60.2	14.6
Portugal	19.8	1.6	92.2	48.5	25.0	74.5	8.0
Rumänien	48.0	4.2	87.8	9.4	72.9	27.0	34.5
Schweden	8.5	15.0	56.5	13.7	72.0	20.5	50.5
Slowakei	10.0	3.9	89.4	52.5	21.8	82.8	5.5
Slowenien	12.3	1.8	70.1	31.7	36.4	65.5	7.3
² Spanien	10.0	16.2	60.0	0.5	92.8	28.7	45.0
² Tschechische Republik	6.8	1.4	96.4	43.6	20.5	77.9	3.5
Ungarn	29.1	0.5	96.4	61.6	11.3	68.8	3.5
VG EU	14.8	12.5	71.6	30.8	45.2	58.6	20.7

Kursiv gesetzt sind die Teilnehmer, für die voneiner eingeschränkten Vergleichbarkeit ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Für Irland liegen hierzu keine Angaben vor.

können, wobei auch Techniker und Akademiker subsumiert wurden (69.5%). Dieser Anteil liegt leicht über dem Mittelwert der europäischen Vergleichsgruppe (63.3%).

Im Vergleich zur Einschätzung der Schulleitungen, wie viele Kinder die Testsprache als Muttersprache erlernt haben, wurden die Schülerinnen und Schüler danach gefragt, ob sie die Testsprache zu Hause mit ihren Eltern sprechen. Dies ist in Deutschland bei über 80 Prozent der Schülerinnen und Schüler der Fall. Hingegen spricht unter ein Prozent der Schülerschaft die Testsprache nie zu Hause.

Hinsichtlich der Angaben zu vorschulischen Kenntnissen, welche am ehesten mit den Angaben der Schulleitungen vergleichbar sind, zeigt sich, dass nach Elternangaben 67.2 Prozent der Schülerinnen und Schüler in Deutschland alle

Zahlen bis 10 schreiben können, wenn sie eingeschult werden. In Anbetracht der Schulleiterangabe darüber, dass über 40 Prozent der Schülerinnen und Schüler Schulen besuchen, in denen weniger als ein Viertel der Kinder bei Schulbeginn Zahlen bis 10 schreiben können, zeigt sich bezüglich der vorschulischen Erfahrungen der Kinder bei den Eltern eine andere Wahrnehmung. Lediglich 6.7 Prozent der Schülerinnen und Schüler konnten nach Angaben der Eltern bei der Einschulung gar keine Zahlen schreiben.

Eine ähnliche Diskrepanz zwischen den Einschätzungen zu den vorschulischen Kenntnissen zeigt sich auch im Bereich des eigenständigen Zählens, wobei erneut die Eltern die Kompetenzen ihrer Kinder höher einschätzen als die Schulleitungen. In Deutschland können nach Elternangaben 35.0 Prozent der

Tabelle 5.3: Aspekte der Schülerkomposition aus Sicht der Eltern und Schülerinnen und Schüler (Anteile der Schülerinnen und Schüler in Prozent)

Teilnehmer	Ökonomische Verhältnisse ^B		Prozentualer Anteil der Schüler, die die Testsprache sprechen ^C		Vorschulische Kenntnisse ^B			
	manuelle Tätigkeiten ^A	Führungs-kräfte	nie	immer	Zahlen schreiben keine Zahlen	alle 10 Zahlen	Eigenständig zählen überhaupt nicht	bis 100 oder weiter
³ Belgien (Fläm. Gem.)	-	-	2.9	75.0	-	-	-	-
²³ Dänemark	-	-	1.2	81.1	-	-	-	-
Deutschland	14.0	69.5	0.9	80.2	6.7	67.2	0.1	35.0
^{1 3} England	-	-	1.7	81.0	-	-	-	-
Finnland	9.9	83.7	0.6	89.2	1.7	77.8	0.9	54.9
Irland	17.2	72.7	2.4	84.2	12.5	48.4	0.7	19.5
³ Italien	27.7	49.0	2.9	79.4	7.9	71.4	1.3	29.7
² Litauen	28.2	53.4	0.8	82.8	4.5	82.1	1.5	46.8
¹ Malta	24.3	58.5	24.9	17.0	2.7	83.0	0.8	39.1
³ Niederlande	-	-	1.9	80.9	-	-	-	-
³ Nordirland	13.1	80.0	1.1	91.1	16.2	40.7	1.1	14.0
² Österreich	16.3	63.2	3.4	76.8	9.4	60.6	0.4	34.5
Polen	32.0	50.1	0.3	90.9	4.7	70.2	1.6	28.9
Portugal	23.3	60.3	0.9	89.9	5.9	75.2	0.5	30.5
Rumänien	43.2	30.2	1.2	89.4	6.2	78.0	3.0	50.1
Schweden	7.0	89.7	1.6	80.0	2.1	77.6	0.4	58.6
Slowakei	21.2	61.8	3.2	79.8	10.3	60.5	4.1	24.8
Slowenien	14.3	74.4	-	-	6.7	67.2	1.5	24.2
² Spanien	22.1	60.9	10.1	67.7	2.6	87.1	0.9	47.6
² Tschechische Republik	19.5	65.4	0.6	86.9	4.0	68.4	0.3	34.6
Ungarn	26.2	53.3	0.5	97.3	8.6	60.5	0.5	47.5
VG EU	20.0	63.3	3.1	80.0	6.6	69.2	1.1	36.5

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Hierunter fallen außerdem Angestellte und Kleinunternehmer.

B = Angaben der Eltern.

C = Angaben der Schülerinnen und Schüler.

- = Für die Teilnehmerstaaten liegen keine Werte vor.

Schülerinnen und Schüler bis 100 und weiter zählen, wenn sie eingeschult werden. Hingegen geben die Schulleitungen an, dass rund drei Viertel der Schülerinnen und Schüler Schulen besuchen, an denen weniger als 25 Prozent der Kinder eigenständig bis 100 und weiter zählen können.

6 Personelle Merkmale: Lehrkräfte und Schulleitungen

Im letzten Teil dieses Kapitels stehen die Lehrkräfte und Schulleitungen im Fokus der Betrachtung. Die dargestellten Befunde sollen die personelle Situation an Grundschulen im europäischen Vergleich skizzieren. Betrachtet werden (6.1) das Geschlechterverhältnis und das Alter der Lehrkräfte, (6.2) die Ausbildung der Lehrkräfte sowie (6.3) die Situation der Fortbildungen der Lehrkräfte und Schulleitungen.

6.1 Geschlechterverhältnis und Alter der Lehrkräfte

Hinsichtlich des Geschlechterverhältnisses zeigt sich, dass 85.0 Prozent der Viertklässlerinnen und Viertklässler in Deutschland von weiblichen Lehrkräften unterrichtet werden. Auch im europäischen Vergleich überwiegt mit 84.0 Prozent der Frauenanteil im Primarbereich (weitere Befunde siehe Tarelli et al., 2012).

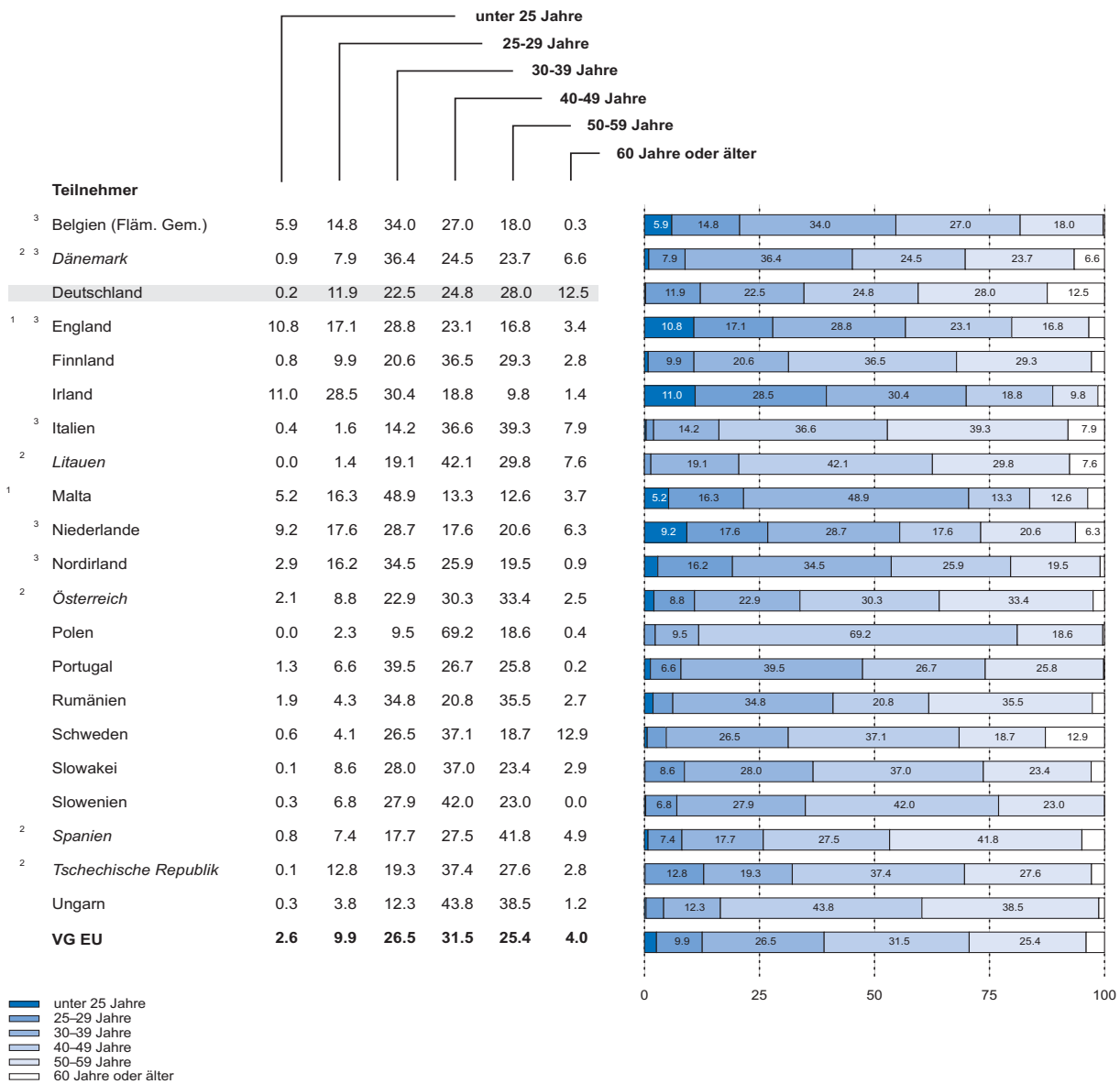
Inwieweit das Alter von Lehrpersonen im Zusammenhang mit Unterrichtsqualität oder Leistungen von Schülerinnen und Schülern steht, lässt sich pauschal nicht beantworten. Es kann gleichermaßen argumentiert werden, dass Schülerinnen und Schüler von den Erfahrungen älterer Lehrkräfte profitieren oder dass bei jüngeren Lehrpersonen ein hohes Potential hinsichtlich moderner Unterrichtsmethoden vermutet werden kann.

Betrachtet man die Angaben der Lehrkräfte in TIMSS 2011 (siehe Abbildung 5.7), so zeigt sich, dass 40.5 Prozent der Schülerinnen und Schüler an Grundschulen in Deutschland von Lehrpersonen unterrichtet werden, die 50 Jahre oder älter sind. 12.5 Prozent aller Viertklässlerinnen und Viertklässler in Deutschland werden von Lehrkräften unterrichtet, die bereits 60 Jahre oder älter sind. Lediglich in Spanien (46.7%) und Italien (37.2%) werden anteilig mehr Schülerinnen und Schüler von Lehrkräften unterrichtet, die über 50 Jahre alt sind, wobei der Anteil der Schülerinnen und Schüler, die von über 60 Jahre alten Lehrkräften unterrichtet werden, in diesen beiden Staaten geringer ausfällt als in Deutschland.

Etwa die Hälfte der Schülerinnen und Schüler in Deutschland wird von Lehrkräften unterrichtet, die zwischen 30 und 49 Jahre alt sind (47.3%). In Spanien (45.2%) und den Niederlanden (46.3%) fällt dieser Anteil geringer aus. In Polen werden hingegen 78.7 Prozent der Schülerinnen und Schüler von Lehrkräften unterrichtet, die zwischen 30 und 49 Jahre alt sind. Der Durchschnitt der europäischen Vergleichsgruppe liegt in dieser Alterskohorte bei 58.0 Prozent.

12.2 Prozent der Viertklässlerinnen und Viertklässler in Deutschland werden von Lehrkräften unterrichtet, die unter 30 Jahre alt sind. In Irland ist dieser Anteil mit 39.5 Prozent am höchsten und in Litauen (1.4%), Italien (2.0%), Polen (2.3%), Ungarn (4.2%) und Schweden (4.7%) am geringsten. Der Mittelwert in der europäischen Vergleichsgruppe liegt in diesem Bereich bei 12.5 Prozent. Gravierende Veränderungen hinsichtlich der Altersstruktur haben sich in den letzten Jahren nicht ergeben (Tarelli et al., 2012).

Abbildung 5.7: Altersverteilung der Lehrkräfte (Anteile der Schülerinnen und Schüler nach Angaben der Lehrkräfte in Prozent)



Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

6.2 Ausbildung der Lehrkräfte

Die Ausbildung der Lehrkräfte ist im europäischen Raum nicht einheitlich strukturiert und somit auch nur bedingt miteinander vergleichbar. Besonders für den naturwissenschaftlichen Sachunterricht muss bei der Interpretation der Ergebnisse beachtet werden, dass es sich in Deutschland um ein integratives Fach handelt, in dem neben naturwissenschaftlichen auch sozialwissenschaftliche und technische Bezüge hergestellt werden (vgl. Kapitel 4). Daher wurden die Lehrkräfte in Deutschland nicht danach gefragt, ob sie Naturwissenschaften studiert haben, sondern danach, ob sie die Fächer Biologie, Chemie, Geographie, Physik oder Sachunterricht als Schwerpunkt während ihres Studiums hatten.

Das Informationsnetzwerk der Europäischen Union (*Eurydice*) unterscheidet grundsätzlich zwei Typen der Lehrerausbildung: zum einen grundspezifische Modelle, in denen fach- und berufsbezogene Inhalte parallel vermittelt werden, und zum anderen konsekutive Systeme, in denen ein fachorientierter Bachelor und ein berufsspezifischer Master aufeinanderfolgen (Blömeke, 2006). Der Bologna-Prozess, der 1999 unter anderem mit dem Ziel initiiert wurde, die universitäre Lehrerbildung bis zum Jahre 2010 zu vereinheitlichen, um mehr Vergleichbarkeit und Mobilität zu gewährleisten, ist allerdings bis heute nicht vollständig abgeschlossen. In Deutschland haben sich fast alle Bundesländer, mit Ausnahme des Saarlandes und Mecklenburg-Vorpommern, für die Einführung einer gestuften Studienstruktur entschieden, wobei sich der Ausbaustand derzeit noch stark unterscheidet. Lediglich in Berlin, Brandenburg, Bremen, Hamburg, Niedersachsen, Rheinland-Pfalz, Sachsen und Schleswig-Holstein ist das System vollständig umgestellt.

Das Hochschulrahmengesetz in Deutschland sieht derzeit eine Regelstudienzeit von 3 bis 4 Jahren vor, was im *European Credit Transfer and Accumulation System* (ECTS) 180 bis 240 Leistungspunkten beziehungsweise *Creditpoints* entspricht, wobei sich ein Leistungspunkt in 25 bis 30 Zeitstunden umrechnen lässt (HRK, 2011).

Aus den Daten der TIMSS-Lehrerbefragung geht hervor, dass 90,5 Prozent der Schülerinnen und Schüler in Deutschland von Lehrkräften unterrichtet werden, die auch das Grundschullehramt studiert haben. Damit gehört Deutschland, ebenso wie beispielsweise Österreich, Slowenien und die Niederlande, in der Vergleichsgruppe EU zu den 60 Prozent der Staaten, in denen die Schülerinnen und Schüler zu über 90 Prozent von Lehrkräften unterrichtet werden, die Grundschullehramt auch als Schwerpunkt studiert haben. Lediglich in Italien (13,6%) und Dänemark (44,6%) werden die Schülerinnen und Schüler von Lehrkräften unterrichtet, von denen weniger als die Hälfte einen entsprechenden Schwerpunkt während des Studiums hatte (Aktionsrat Bildung, 2011).

In Deutschland werden 52,4 Prozent der Schülerinnen und Schüler im Mathematikunterricht von Lehrkräften unterrichtet, die Mathematik auch als Schwerpunkt während ihres Studiums hatten. Lediglich in Dänemark ist dieser Anteil höher (56,4%). Der Anteil der Schülerinnen und Schüler, die in Deutschland im naturwissenschaftlichen Sachunterricht von Lehrkräften unterrichtet werden, die Biologie, Geographie, Chemie, Physik oder Sachunterricht auch als Schwerpunkt während ihres Studiums hatten, beträgt 61,6 Prozent.

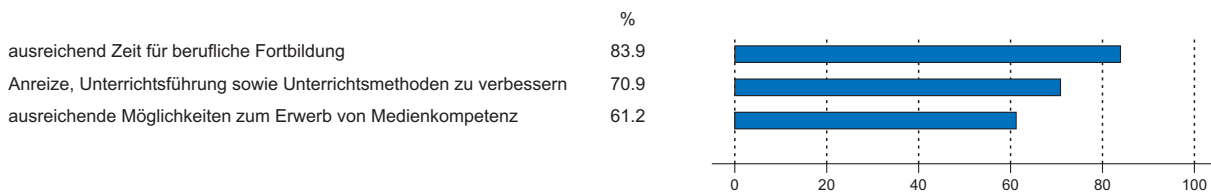
6.3 Fortbildung der Lehrkräfte und Schulleitungen

Vor dem Hintergrund europäischer Standardisierungsbemühungen und dem nationalen Wunsch, zukünftiges Lehrpersonal mit zentralen, organisatorischen, personalen und unterrichtlichen Kompetenzen auszustatten, die es ihnen ermöglichen, die an sie gestellten komplexen beruflichen Anforderungen (vgl. KMK, 1994, 2000) erfolgreich zu bewältigen, wurden in der letzten Dekade Anstrengungen unternommen, die Lehrerausbildung grundlegend neu zu strukturieren und zu konzipieren. Dies erfolgte sowohl in struktureller als auch in inhaltlicher Hinsicht (Bastian, Keuffer & Lehberger, 2005). Die KMK stellt in ihrem Beschluss von 2004 zu den Standards für die Lehrerbildung die Weiterbildung von Lehrkräften als Teil des professionellen Verständnisses des Lehrerberufs heraus: „*Lehrerinnen und Lehrer entwickeln ihre Kompetenzen ständig weiter und*

nutzen wie in anderen Berufen auch Fort- und Weiterbildungsangebote, um die neuen Entwicklungen und wissenschaftlichen Erkenntnisse in ihrer beruflichen Tätigkeit zu berücksichtigen“ (KMK, 2004).

Alle Bemühungen, das Fortbildungsangebot zu optimieren, nützen jedoch wenig, wenn an der Einzelschule grundlegende Rahmenbedingungen für die Teilnahme an Fortbildungsveranstaltungen nicht gewährleistet werden können. Die Einschätzung der Lehrkräfte darüber, ob an ihren Schulen ausreichend Möglichkeiten für ihre berufliche Weiterbildung bestehen, ist in Abbildung 5.8 dargestellt. Dabei wurden die Antworten zu der ‚starken‘ und der ‚eher starken‘ Zustimmung zu einer Kategorie zusammengefasst. Der Großteil der Schülerinnen und Schüler in Deutschland wird von Lehrkräften unterrichtet, die zu der Einschätzung kommen, dass ihnen ihre Schule genügend Zeit für berufliche Fortbildungen zur Verfügung stellt (83.9%). Weniger Schülerinnen und Schüler werden von Lehrkräften unterrichtet, die angeben, dass ihnen die Schule Anreize bietet, die Unterrichtsführung und die Unterrichtsmethoden zu verbessern (70.9%). Am geringsten fällt hingegen mit 61.2 Prozent der Anteil der Schülerinnen und Schüler aus, deren Lehrkräfte angeben, dass ihnen die Schule ausreichend Möglichkeiten zum Erwerb von Medienkompetenz bietet, obwohl gerade in diesem Bereich ein großer Bedarf besteht (siehe Abschnitt 4.3).

Abbildung 5.8: Fortbildungsmöglichkeiten der Lehrkräfte (Anteile der Schülerinnen und Schüler nach Angaben der Lehrkräfte in Prozent)



IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Zu welchen Inhalten die Lehrkräfte nach eigenen Angaben in den letzten zwei Jahren an Fortbildungsveranstaltungen im Bereich Mathematik teilgenommen haben, ist in Tabelle 5.4 dargestellt. In Deutschland werden die Schülerinnen und Schüler von Lehrkräften unterrichtet, die in den letzten zwei Jahren mehr Fortbildungen zu den Inhaltsbereichen ‚mathematische Inhalte‘ (54.6%), ‚Mathematikdidaktik‘ (43.5%), ‚Lehrplan‘ (33.0%) und zum Thema ‚Eingehen auf die individuellen Bedürfnisse der Schülerinnen und Schüler‘ (46.1%) besucht haben als im Durchschnitt der europäischen Teilnehmerstaaten. Leicht unter dem Durchschnitt der europäischen Vergleichsstaaten liegt der Anteil der Schülerinnen und Schüler in Deutschland, die von Lehrkräften unterrichtet werden, die in den letzten zwei Jahren mindestens eine Fortbildung im Bereich der Leistungsfeststellung absolviert haben (27.4%). Hingegen werden die Schülerinnen und Schüler in Deutschland von Lehrkräften unterrichtet, die im Vergleich zu den anderen europäischen Teilnehmerstaaten insgesamt am wenigsten Fortbildungen zur Integration von Informationstechnologien besucht haben (5.3%). In Anbetracht der Ergebnisse im Abschnitt 4.3 ‚Computernutzung von Lehrkräften‘ wird jedoch deutlich, dass gerade in diesem Bereich Handlungsbedarf besteht.

Die Viertklässlerinnen und Viertklässler in England, Nordirland und Rumänien werden von Lehrkräften unterrichtet, die in allen genannten Bereichen in den

letzten zwei Jahren mehr Fortbildungen absolviert haben als der Durchschnitt der europäischen Teilnehmerstaaten. Hingegen liegen die Lehrkräfte in Finnland und der Tschechischen Republik in allen genannten Bereichen unter dem Durchschnitt der europäischen Teilnehmerländer. Für den naturwissenschaftlichen Sachunterricht liegen allerdings keine Informationen vor.

Tabelle 5.4: Fortbildungsinhalte der Mathematiklehrkräfte in den letzten zwei Jahren (Anteile der Schülerinnen und Schüler nach Angaben der Lehrkräfte in Prozent)

Teilnehmer	Mathematische Inhalte	Mathematikdidaktik	Lehrplan	Integration von Informationstechnologien	Leistungs feststellung	Eingehen auf individuelle Bedürfnisse der Schüler
³ Belgien (Fläm. Gem.)	11.7	11.0	20.0	20.7	5.9	33.6
^{2,3} Dänemark	28.6	32.5	12.5	19.5	24.1	25.2
Deutschland	54.6	43.5	33.0	5.3	27.4	46.1
^{1,3} England	53.8	71.1	46.4	30.1	59.0	51.9
Finnland	9.3	19.7	3.5	9.3	3.4	28.6
Irland	31.6	32.4	34.5	31.3	24.9	33.4
³ Italien	27.9	37.9	27.0	22.3	20.7	38.3
² Litauen	33.1	31.4	50.9	66.3	48.1	46.3
¹ Malta	18.4	20.6	16.8	32.0	22.8	42.3
³ Niederlande	22.0	27.2	11.0	20.4	18.0	50.1
³ Nordirland	54.7	64.3	61.6	55.2	61.2	41.5
² Österreich	75.4	54.9	32.8	14.7	27.2	58.2
Polen	61.2	31.1	48.8	33.6	24.2	70.7
Portugal	58.3	54.3	61.0	36.0	25.2	31.0
Rumänien	54.1	50.1	54.2	33.9	60.7	50.0
Schweden	52.9	59.7	57.3	10.5	43.6	32.5
Slowakei	11.5	20.0	44.6	47.2	16.8	22.4
Slowenien	31.8	23.3	44.6	44.0	42.9	53.0
² Spanien	15.5	24.5	18.6	39.6	14.2	33.2
² Tschechische Republik	16.4	26.4	7.6	22.0	11.1	25.3
Ungarn	27.8	44.8	12.8	22.4	22.2	36.9
VG EU	35.7	37.2	33.3	29.3	28.8	40.5

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

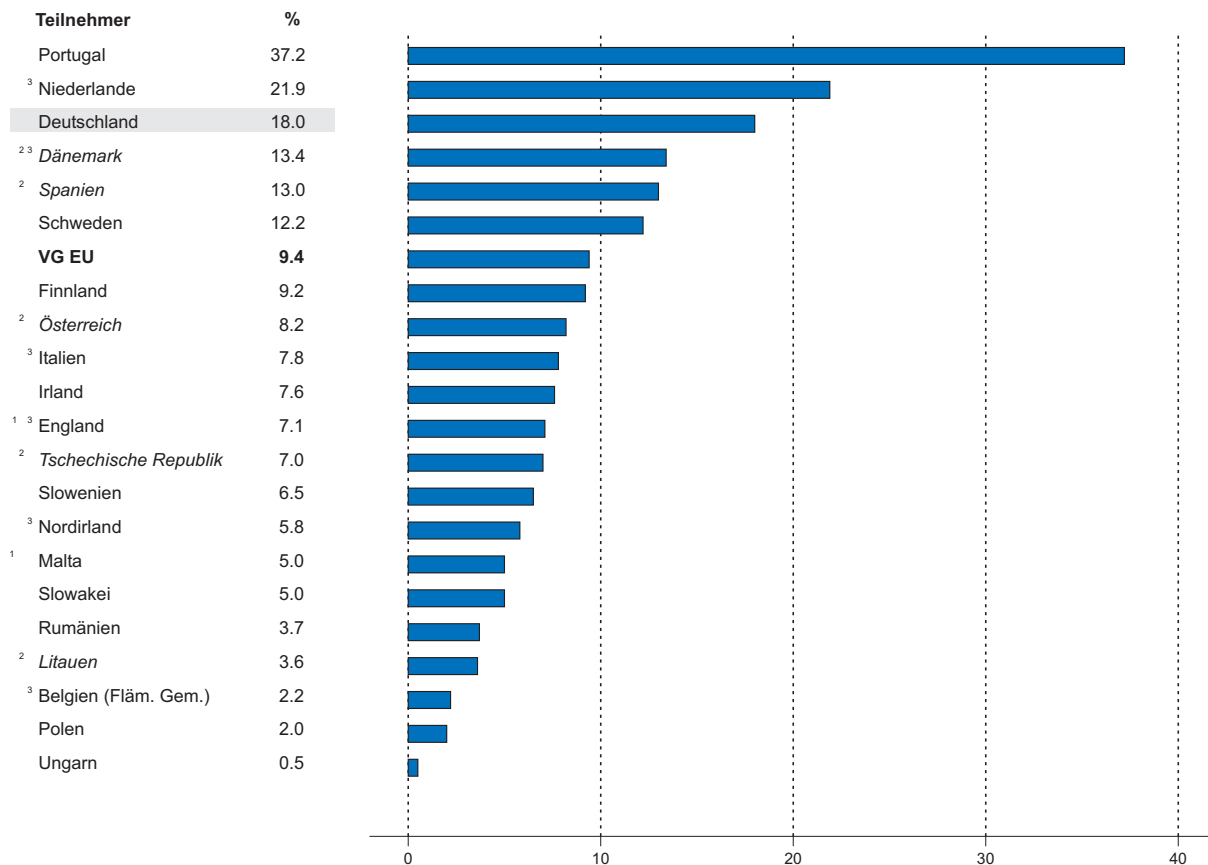
3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

Im Zuge der erweiterten Anforderungen an Schulleitungen werden auch Fortbildungsveranstaltungen, die speziell auf die Bedürfnisse von Schulleitungen abgestimmt sind, immer wichtiger. Gleichzeitig ist fraglich, inwieweit das Angebot qualitativ hochwertiger Fortbildungen, die speziell auf die Rolle der Schulleitung vorbereiten oder Weiterqualifizierung im Hinblick auf die sich wandelnden Ansprüche anbieten, als ausreichend anzusehen ist (Huber, 2004, 2009). Darüber hinaus stellt sich die Frage, ob Schulleitungen von den Angeboten Gebrauch machen (vgl. Abbildung 5.9).

In Deutschland besuchen 18.0 Prozent der Schülerinnen und Schüler Schulen, an denen die Schulleitung angibt, im vergangenen Jahr keine Zeit für Fortbildungen, die speziell für Schulleitungen angelegt sind, investiert zu haben. Lediglich in Portugal (37.2%) und den Niederlanden (21.9%) ist dieser Anteil

höher. Der geringste Anteil an Schülerinnen und Schülern, deren Schulleitungen angeben, keine Zeit für Schulleiterfortbildungen aufgewendet zu haben, liegt mit 0.5 Prozent in Ungarn vor, gefolgt von Polen mit 2.2 Prozent. Der Durchschnitt der europäischen Teilnehmerstaaten liegt bei 9.4 Prozent.

Abbildung 5.9: Zeit für Schulleiterfortbildung (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen, Kategorie „keine Zeit“)



Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3= Die Teilnahmequoten auf Schul- und/oder Schülerseite erreichen nicht die internationalen Vorgaben.

7 Zusammenfassung

Die in TIMSS realisierte Fragebogenerhebung ermöglicht es, repräsentative Befunde zur Situation von Grundschulen in den Teilnehmerstaaten darzustellen. In diesem Kapitel wurden einige ausgewählte Rahmenbedingungen, in denen das Lehren und Lernen an Grundschulen stattfindet, in den Blick genommen. Dabei war das primäre Anliegen, einen Eindruck zu Unterschieden und Gemeinsamkeiten zwischen den Teilnehmerstaaten zu vermitteln, um vor diesem Hintergrund die in diesem Band berichteten Leistungsergebnisse einordnen zu können. Als Vergleichsgruppe wurden die Teilnehmerstaaten ausgewählt, die Mitglieder der EU sind. Des Weiteren zeigen die Ergebnisse auch auf, in welchen Bereichen aus Sicht der Lehrkräfte beziehungsweise Schulleitungen Handlungs- und Verbesserungsbedarf besteht.

Bildungspolitische Merkmale und äußere schulische Rahmenbedingungen

Auf bildungspolitischer Ebene zeigt sich, dass Deutschland im Vergleich zur VG_{EU} anteilig zum BIP wenig in die primäre Bildung investiert. Es gibt allerdings durchaus Bereiche in der Grundschule, in denen Investitionen getätigt werden könnten, wie eine flächendeckende Ausstattung mit Bibliotheken, in denen vor allem mehr Bücher für den Mathematikunterricht zugänglich sind, die mediale Ausstattung oder Fortbildungsmöglichkeiten hinsichtlich der Einbeziehung von Computern in den Fachunterricht.

In Bezug auf den Ausbau der Ganztagschullandschaft zeigt sich eine eindeutige Tendenz: Besuchten im Jahr 2007 etwa ein Drittel der Schülerinnen und Schüler eine Ganztagschule, sind es 2011 bereits 46.8 Prozent, was einer Zunahme von 11.7 Prozentpunkten entspricht. Dabei ist zu beobachten, dass sich die teilweise gebundene Form nicht etabliert hat. Die Verfügbarkeit von erweiterten Lern- und Förderangeboten ist hingegen sowohl bei den Halbtags- als auch bei den Ganztagsgrundschulen als gut zu bewerten, wobei es keine gravierenden Unterschiede in der Angebotsbereitstellung gibt.

Schulinterne Merkmale

Bei den schulinternen Merkmalen zeigt sich, dass Schulen in Deutschland im Vergleich zur VG_{EU} im Bereich der digitalen Medien relativ gut ausgestattet sind, das Verhältnis von einem Computer für vier Kinder allerdings noch optimiert werden könnte. Die Verfügbarkeit von Computern hat sich im naturwissenschaftlichen Sachunterricht verbessert, ist im Mathematikunterricht jedoch etwas gesunken. Auch hier besteht allerdings Potential, die Nutzung zu intensivieren. Durchaus positiv zu bewerten ist, dass der Großteil der verfügbaren Geräte im Mathematik- und Sachunterricht auch über einen Zugang zum Internet verfügt.

Probleme mit der Ausstattung im Mathematik- und Sachunterricht bestehen im Mangel an fachspezifischen Büchern – und dies obwohl an den meisten Schulen Schul- oder Klassenbibliotheken vorhanden sind. Es zeigt sich auch, dass ein Mangel in der Ausstattung mit Software für den naturwissenschaftlichen Sachunterricht besteht. Zudem besuchen lediglich 12.8 Prozent der Schülerinnen und Schüler in Deutschland Schulen, welche über ein Labor für den naturwissenschaftlichen Sachunterricht verfügen.

Unterrichtsmerkmale

Grundlegende Vorgehensweisen im Unterricht wie das Zusammenfassen zentraler Lerninhalte und das Herstellen von Bezügen zum Alltag werden im Mathematik- und Sachunterricht in Deutschland nach Angaben der Lehrkräfte weniger praktiziert als in den meisten Staaten der VG_{EU}.

In Deutschland nutzen fast alle Mathematik- und Sachunterrichtslehrkräfte den Computer zur Vorbereitung des Unterrichts. Allerdings setzen nur drei Viertel der Lehrkräfte den Computer im Unterricht ein, was damit zusammenhängen könnte, dass sich einige Lehrkräfte hier keinen souveränen Einsatz zutrauen. Auf diese Vermutung deutet auch der Befund hin, dass vergleichsweise viele Lehrkräfte angeben, dass ein großer Unterstützungsbedarf bei technischen Angelegenheiten, bei der Computernutzung im Unterricht besteht.

Klassenmerkmale

Insgesamt lässt sich für Deutschland festhalten, dass extreme Zusammensetzungen von Schülerschaften aus ökonomisch benachteiligten Verhältnissen oder mit hohem Anteil an nicht muttersprachlichen Kindern eher die Ausnahme darstellen.

len. Länder-, regional-, beziehungsweise stadtteilspezifische Problemlagen bleiben davon unbenommen natürlich bestehen. Zwischen den Einschätzungen der Schulleitungen und den Elternangaben zu den vorschulischen Kenntnissen im Bereich Mathematik zeigen sich unterschiedliche Wahrnehmungen: Nach Schulleiterangaben besuchen über 40 Prozent der Schülerinnen und Schüler Schulen, in denen weniger als ein Viertel der Schülerschaft zu Einschulungsbeginn Zahlen schreiben kann. Nach Elternangaben beherrschen allerdings über 62 Prozent der Kinder das Schreiben von Zahlen.

Personelle Merkmale

Die Altersstruktur in den Kollegien an Schulen in Deutschland zeigt, dass es in einigen Jahren, auch unter Berücksichtigung demographischer Entwicklungen, zu einem akuten Lehrkräftemangel kommen könnte. Über 40 Prozent der Schülerinnen und Schüler an Grundschulen in Deutschland werden von Lehrpersonen unterrichtet, die 50 Jahre oder älter sind, wobei 12.5 Prozent der Viertklässlerinnen und Viertklässler von Lehrkräften unterrichtet werden, die bereits 60 Jahre oder älter sind.

Im Vergleich zur VG_{EU} werden in Deutschland deutlich mehr Lehrkräfte speziell für die Primarstufe ausgebildet. Ebenso hatten im Vergleich zur VG_{EU} deutlich mehr Lehrkräfte in Deutschland das Fach (Mathematik oder Sachunterricht), welches sie unterrichten, auch als Schwerpunkt während des Studiums. Dennoch gibt es einen Anteil von 48.6 Prozent der Schülerinnen und Schüler in Deutschland, die in Mathematik von Lehrkräften unterrichtet werden, die das Fach nicht als Schwerpunkt während ihres Studiums hatten. Für den naturwissenschaftlichen Sachunterricht fällt dieser Anteil etwas geringer aus (38.4%).

Die Lehrkräfte in Deutschland sind der Ansicht, dass ihnen genügend Zeit für berufliche Weiterbildung zur Verfügung steht. Verbesserungsbedarf besteht bei den Angeboten zum Erwerb von Medienkompetenzen. In diesem Bereich wurden auch kaum Fortbildungsveranstaltungen in den letzten zwei Jahren besucht.


Zudem zeigt sich, dass 18.0 Prozent der Schülerinnen und Schüler Schulen besuchen, an denen die Schulleitung angibt, im vergangenen Jahr keine Zeit für Fortbildungen speziell für Schulleitungen investiert zu haben, was jedoch in Anbetracht der gewandelten Anforderungen an Schulleitungen zunehmend wichtiger wird.

Literatur

- Aktionsrat Bildung. (2011). *Bildungsreform 2000 – 2010 – 2020. Jahresgutachten 2011* (1. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Arnold, G. (2005). *Kleine Klassen – große Klasse? Eine empirische Studie zur Bedeutung der Klassengröße in Schule und Unterricht*. Bad Heilbrunn: Klinkhardt.
- Autorengruppe Bildungsberichterstattung. (2012). *Bildung in Deutschland 2012. Ein indikatorengestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf*. Zugriff am 11.10.2012 unter http://www.bildungsbericht.de/daten2012/bb_2012.pdf.
- Bastian, J., Keuffer, J. & Lehberger, R. (2005). *Lehrerbildung in der Entwicklung. Das Bachelor-Master-System: Modelle – kritische Hinweise – Erfahrungen*. Weinheim: Beltz.
- Baumert, J. (2002). Deutschland im internationalen Bildungsvergleich. In N. Killius, J. Kluge & L. Reisch (Hrsg.), *Die Zukunft der Bildung* (S. 100–150). Frankfurt a. M.: Suhrkamp.
- Blömeke, S. (2006). Struktur der Lehrerausbildung im internationalen Vergleich. Ergebnisse einer Untersuchung zu acht Ländern. *Zeitschrift für Pädagogik*, 52 (3), 393–416.

- BMBF – Bundesministerium für Bildung und Forschung. (2008). *Investitionsprogramm „Zukunft Bildung und Betreuung“*. Ganztagschulen. Zeit für mehr. Zugriff am 28.10.2012 unter http://www.bmbf.de/pub/ganztagschulen-zeit_fuer_mehr.pdf.
- Bos, W., Tarelli, I., Bremerich-Vos, A. & Schwippert, K. (Hrsg.). (2012). *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Firman, C. & Camilleri, R. (2012). Malta. In I. V. S. Mullis, M. O. Martin, C. A. Minnich, K. T. Drucker & M. A. Ragan (Hrsg.), *PIRLS 2011 Encyclopedia. Education policy and curriculum in reading, Volumes 1 and 2* (S. 395–406). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Gröhlich, C., Scharenberg, K. & Bos, W. (2009). Wirkt sich Leistungsheterogenität in Schulklassen auf den individuellen Lernerfolg in der Sekundarstufe aus? *Journal for educational research online*, 1 (1), S. 86–105.
- Hattie, J. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge, Taylor & Francis.
- Helmke, A. (1997). Individuelle Bedingungsfaktoren der Schulleistungen: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 203–216). Weinheim: Beltz.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (1. Aufl.). Seelze-Velber: Kallmeyer.
- Holtappels, H. G. (2009). Qualitätsmodelle – Theorie und Konzeptionen. In I. Kamski, H. G. Holtappels & T. Schnetzer (Hrsg.), *Qualität von Ganztagschule. Konzepte und Orientierungen für die Praxis* (S. 11–25). Münster: Waxmann.
- HRK – Hochschulrektorenkonferenz. (2011). *Statistische Daten zu Bachelor- und Masterstudiengängen. Wintersemester 2011/2012* (Statistiken zur Hochschulpolitik, 2/2011, 1. Aufl.). Bonn: Hochschulrektorenkonferenz.
- Huber, S. G. (2004). Schulleiterqualifizierung: Die Angebote im Ausland sind umfangreich – in Deutschland besteht Nachholbedarf. *Schulverwaltung. Bayern. Zeitschrift für Schulleitung und Schulaufsicht*, 27 (3), 98–101.
- Huber, S. G. (2009). Schulleitung – Anforderungen und Professionalisierung aus internationaler Perspektive. *Lehren und Lernen*, 35 (8/9), 12–21.
- Klieme, E., Jude, N., Baumert, J. & Prenzel, M. (2010). PISA 2000 – 2009: Bilanz der Veränderungen im Schulsystem. 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. 277–300). Münster: Waxmann.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (1994). *Empfehlungen zur Arbeit in der Grundschule (Beschluss der Kultusministerkonferenz vom 02.07.1970 in der Fassung vom 06. 05. 1994)*. Zugriff am 28.10.2012 unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1970/1970_07_02_Empfehlungen_Grundschule.pdf.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2000). *Rahmenvereinbarung zur Ausbildung und Prüfung von Erziehern/Erzieherinnen (Beschluss der Kultusministerkonferenz vom 28.01.2000)*. Zugriff am 28.10.2012 unter <http://www.berufsbildung.schulministerium.nrw.de/cms/upload/fs/download/sozial/kmkrahm.pdf>.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften (Beschluss der Kultusministerkonferenz vom 16.12.2004)*. Zugriff am 28.10.2012 unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2010). *Allgemein bildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik Deutschland – Statistik 2004 bis 2008*. Zugriff am 28.10.2012 unter http://www.kmk.org/fileadmin/pdf/Statistik/GTS_2008.pdf.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2011). *Definitionskatalog zur Schulstatistik 2011*. Zugriff am 28.10.2012 unter <http://www.kmk.org/fileadmin/pdf/Statistik/Defkat2011.pdf>.
- Lankes, E.-M. & Carstensen, C. H. (2010). Klassengrößen. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, I. Tarelli & R. Valtin

- (Hrsg.), *IGLU 2006 – die Grundschule auf dem Prüfstand. Vertiefende Analysen zu Rahmenbedingungen schulischen Lernens*. (S. 121–142). Münster: Waxmann.
- Ramseger, J. (2007). Kleine Klassen – bessere Leistungen? Zum Widerspruch zwischen pädagogischer Alltagserfahrung und empirischer Schulforschung. *Erziehung und Wissenschaft*, 59 (3), 30–31.
- Saldern, M. v. (2011). *Klassengröße. Über ein vernachlässigtes Merkmal*. Norderstedt: Books on Demand.
- Schorch, G. (2007). *Studienbuch Grundschulpädagogik. Die Grundschule als Bildungsinstitution und pädagogisches Handlungsfeld* (3. Aufl.). Bad Heilbrunn: Klinkhardt.
- Schulz-Zander, R., Eickelmann, B. & Goy, M. (2010). Mediennutzung, Medieneinsatz und Lesekompetenz. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, I. Tarelli & R. Valtin (Hrsg.), *IGLU 2006 – die Grundschule auf dem Prüfstand. Vertiefende Analysen zu Rahmenbedingungen schulischen Lernens*. (S. 91–119). Münster: Waxmann.
- Statistisches Bundesamt. (2011). *Bildung und Kultur. Allgemeinbildende Schulen. Schuljahr 2009/2010*. Fachserie 11, Reihe 1. Zugriff am 28.10.2012 unter https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/AllgemeinbildendeSchulen2110100107004.pdf?__blob=publicationFile.
- Statistisches Bundesamt. (2012). *Allgemeinbildende und berufliche Schulen. Schüler/innen nach Schularten*, Statistisches Bundesamt. Zugriff am 24.09.2012 unter <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Schulen/Tabellen/AllgemeinBildendeBeruflicheSchulenSchulartenSchueler.html>.
- Strietholt, R., Manitus, V., Berkemeyer, N. & Bos, W. (2012). *Bildung und Bildungsungleichheit an Halb- und Ganztagschulen* (Manuskript zur Publikation eingereicht).
- Tarelli, I., Lankes, E.-M., Drossel, K. & Gegenfurtner, A. (2012). Lehr- und Lernbedingungen an Grundschulen im internationalen Vergleich. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Tarelli, I., Wendt, H., Bos, W. & Zylowski, A. (2012). Ziele, Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU 2011). In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- UNESCO Institute for Statistics (2012). *Education Finance*. Zugriff am 28.10.2012 unter http://stats.uis.unesco.org/unesco/TableViewer/tableView.aspx?ReportId=3560&IF_Language=eng.
- Valtin, R. (2006). Grundschule – die Schule der Nation. Überlegungen zum Bildungsauftrag der Grundschule. In: *Lehren und Lernen*, ZS für Schule und Innovation in Baden-Württemberg. 32. Jg., 2, S. 4-11.
- Valtin, R. (2012). Auf dem Weg zu einer besseren Schule bildungspolitische Folgerungen aus der 2. JAKO-O-Bildungsstudie. In D. Killus & K.-J. Tillmann (Hrsg.), *Eltern ziehen Bilanz. Ein Trendbericht zu Schule und Bildungspolitik in Deutschland*. (S.169-184). Münster: Waxmann
- Wendt, H., Stubbe, T. C. & Schwippert, K. (2012). Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.



Kapitel VI

Geschlechtsspezifische Unterschiede in mathematischen und naturwissenschaftlichen Kompetenzen

Theresia Brehl, Heike Wendt und Wilfried Bos

1 Einleitung

Historisch betrachtet war schulische Bildung über Jahrtausende weitgehend Jungen vorbehalten – eine allmähliche Institutionalisierung der Mädchen- und Frauenbildung vollzog sich in Deutschland erst ab Ende des 18. Jahrhunderts (Stanat & Bergmann, 2009). Eine für Mädchen und Jungen gleiche elementare Schulbildung wurde in Deutschland erst mit der Verankerung der Grundschule als einer Schule für alle Kinder zu Beginn der Weimarer Republik realisiert (Geissler, 2011, S. 354). Auch heute noch werden in der Bundesrepublik Deutschland alle Kinder während der Grundschulzeit gemeinsam unterrichtet.

Dieses Kapitel stellt aktuelle Ergebnisse zu geschlechtsspezifischen Unterschieden in mathematischen und naturwissenschaftlichen Kompetenzen im Kontext des aktuellen Forschungsstandes dar. Dabei werden die Befunde auch im Unterschied zu TIMSS 2007 dargestellt. Außerdem werden das fachbezogene Selbstkonzept und die fachbezogenen Einstellungen von Mädchen und Jungen in Mathematik und Naturwissenschaften beschrieben. Schließlich wird untersucht, inwieweit das Geschlecht, das fachbezogene Fähigkeitsselbstkonzept in Mathematik und im Sachunterricht, die Noten in Mathematik und im Sachunterricht sowie das Bildungsniveau und der Berufsstatus der Eltern mit den erzielten Kompetenzen von Grundschulkindern zusammenhängen.

2 Bisherige Forschungsergebnisse

Die Bildungsforschung beschäftigt sich seit mehreren Jahrzehnten mit der Frage nach Geschlechterdisparitäten und inwieweit es gelingt, allen Kindern und Jugendlichen in Deutschland unabhängig von ihrem Geschlecht gute Entwicklungsmöglichkeiten zu bieten.

Zusammenfassend kann festgestellt werden, dass Mädchen heute nicht mehr als Bildungsverliererinnen gelten können, sondern ähnliche, mitunter auch bessere Leistungen als Jungen erzielen. In Bezug auf die Bildungsbeteiligung in der Grundschule zeigt sich, dass Mädchen im Vergleich zu Jungen häufiger vorzeitig und seltener verspätet eingeschult werden. So besuchen 64 Prozent der sechsjährigen Mädchen, aber nur 58 Prozent der sechsjährigen Jungen bereits die Schule (Autorengruppe Bildungsberichterstattung, 2006). Dies wird darauf zurückgeführt, dass Mädchen in fast allen Einschulungskriterien besser abschneiden und insbesondere im sprachlichen Bereich bereits früh den Jungen gegenüber einen Entwicklungsvorsprung aufweisen (Stanat & Bergmann, 2009). Des Weiteren lassen sich beim Übergang von der Grundschule in die Sekundarstufe I Geschlechterdifferenzen zugunsten der Mädchen finden. So besuchten 2009/10 mehr Mädchen (52.9%) als Jungen (47.1%) das Gymnasium, während in Hauptschulen das Verhältnis umgekehrt war (55.5% Jungen, 44.5% Mädchen) (Statistisches Bundesamt, 2012, S. 136). Geschlechterdifferenzen zeigen sich auch in der Wahrscheinlichkeit für Klassenwiederholungen, die bei Jungen in allen Jahrgangsstufen außer der Grundschule durchweg höher liegt (Autorengruppe Bildungsberichterstattung, 2006).

Insgesamt wird bei Betrachtung der Bildungsbeteiligung deutlich, dass in einigen Bereichen des allgemeinen Schulsystems Unterschiede zugunsten der Mädchen bestehen. Den Bildungsbeteiligungen an den unterschiedlichen Sekundarschulformen kommt dabei eine besondere Bedeutung für die weitere Bildungskarriere der Schülerinnen und Schüler zu.

Bereits am Ende der Grundschulzeit lassen sich geschlechtsbezogene Disparitäten in den erzielten Kompetenzen feststellen, wenngleich diese im Vergleich zum weiteren Verlauf der Bildungskarriere relativ gering ausfallen. In der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) 2006 zeigten sich bereits am Ende der Grundschulzeit höhere Lesekompetenzen der Mädchen im Vergleich zu den Jungen (Hornberg, Valtin, Potthoff, Schwippert & Schulz-Zander, 2007). In den mathematischen und naturwissenschaftlichen Leistungsbereichen bestehen hingegen signifikante Geschlechterunterschiede zugunsten der Jungen (Prenzel, Geiser, Langeheine & Lobemeier, 2003; Prenzel, Geiser, Langeheine & Lobemeier, 2004; Walther, Geiser, Langeheine & Lobemeier, 2003; Walther, Geiser, Langeheine & Lobemeier, 2004; Walther, Selter, Bonsen & Bos, 2008; Wittwer, Saß & Prenzel, 2008). In diesem Zusammenhang konnte in IGLU-E 2001 (Bos, Lankes, Prenzel, Schwippert & Walther, 2004), der nationalen Erweiterungsstudie zu IGLU 2001, mit der erstmalig umfassend mathematische und naturwissenschaftliche Kompetenzen von Viertklässlerinnen und Viertklässlern in der Bundesrepublik Deutschland untersucht wurden, gezeigt werden, dass in beiden Kompetenzdomänen Mädchen auf den unteren Kompetenzstufen überrepräsentiert und auf den oberen Kompetenzstufen unterrepräsentiert waren (Prenzel et al., 2003; Walther et al., 2003). In TIMSS 2007 (Bos et al., 2008) wurden die Ergebnisse von IGLU-E 2001 für Deutschland in weiten Teilen bestätigt. Auch hier zeigte sich auf der Gesamtskala, dass Jungen in Deutschland sowohl in Mathematik als auch in den Naturwissenschaften signifikant höhere Leistungsmittelwerte erzielten als Mädchen. Ein moderater Geschlechtereffekt ließ sich zudem insbesondere auf den beiden untersten Kompetenzstufen erkennen. Darüber hinaus war es in dieser Studie möglich, die Leistungsdifferenzen valider Kompetenzmessungen international zu vergleichen (Walther et al., 2008; Wittwer et al., 2008). Der internationale Vergleich verdeutlicht, dass sowohl die Höhe als auch die Richtung der Geschlechterunterschiede

über die verschiedenen Teilnehmerstaaten hinweg unterschiedlich ausfallen, wobei folgende Ergebnisse hervorgehoben werden können:

- Im internationalen Vergleich konnten für elf Teilnehmerstaaten bezüglich der Mathematikkompetenzen signifikante Leistungsvorsprünge zugunsten der Jungen und für drei Teilnehmerstaaten zugunsten der Mädchen gefunden werden.
- Für die Naturwissenschaften zeigte sich lediglich im internationalen Mittelwert ein signifikanter Leistungsvorsprung der Mädchen und in sechs Teilnehmerstaaten ein signifikanter Leistungsvorsprung der Jungen, wobei für die meisten Teilnehmerstaaten kein signifikanter Geschlechterunterschied zu finden war. In keinem der anderen EU- oder OECD-Teilnehmerstaaten war die Geschlechterdifferenz in der naturwissenschaftlichen Kompetenz größer als in Deutschland.

Neben Kompetenzunterschieden konnten in TIMSS 2007 aber auch geringe Differenzen in fachbezogenen Selbstkonzepten sowie in den fachbezogenen Einstellungen gefunden werden:

- In TIMSS 2007 zeigte sich, dass Jungen sowohl eine signifikant positivere Einstellung zur Mathematik haben als auch ein signifikant höheres mathematikbezogenes Selbstkonzept.
- Außerdem gehört Deutschland zu den Teilnehmerstaaten, in denen Jungen ein signifikant höheres Fähigkeitsselfkonzept in den Naturwissenschaften haben als Mädchen. Dagegen zeigen sich in Deutschland keine geschlechtsspezifischen Unterschiede in den positiven Einstellungen zu den Naturwissenschaften.

Die Befunde aus dem *Programme for International Student Assessment* (PISA) deuten darauf hin, dass sich die Leistungsunterschiede zwischen Mädchen und Jungen im Laufe der Schullaufbahn verstärken (Frey, Heinze, Mildner, Hochweber & Asseburg, 2010; Naumann, Artelt, Schneider & Stanat, 2010; Rönnebeck, Schöps, Prenzel, Mildner & Hochweber, 2010; Schöps, Walter, Zimmer & Prenzel, 2006).

Für eine Einordnung der Ergebnisse sei darauf hingewiesen, dass die Unterschiede in den Leistungen innerhalb der Geschlechtergruppen in der Regel größer sind als die Leistungsunterschiede zwischen Mädchen und Jungen.

3 Erklärungsansätze zu Geschlechterdisparitäten im fachspezifischen Kompetenzerwerb

In der Forschungsliteratur gibt es eine Vielzahl an Erklärungsansätzen zur Entstehung von Geschlechterunterschieden in der Kompetenzentwicklung, welche im Folgenden nur in aller Kürze erläutert werden können (für eine Übersicht vgl. Halpern, 2012). Insgesamt können grob drei Gruppen von Ansätzen unterschieden werden, die in der Regel schulfachübergreifende Gültigkeit beanspruchen: *kognitive*, *biologische* und *psychosoziale* Ansätze (vgl. Stanat & Bergmann, 2009).

Kognitive Ansätze erklären Leistungsunterschiede zwischen Mädchen und Jungen, indem sie fachbezogene Leistungsdisparitäten auf Unterschiede in basalen kognitiven Fähigkeiten zurückführen (Bonsen, Lintorf & Bos, 2008; Halpern, 2004; Linn & Petersen, 1985; Vandenberg & Kuse, 1978). *Biologische*

Erklärungsansätze argumentieren zum einen aus evolutionsbiologischer Sicht und zum anderen aus genetischer Sicht, dass Frauen und Männer unterschiedliche kognitive Fähigkeiten ausbilden beziehungsweise unterschiedlich ausgeprägte Spezialisierungen bei kognitiven Aktivitäten aufweisen (Geary, 1996; Nyborg, 1983; Stanat & Bergmann, 2009). Die kognitiven Ansätze greifen dabei zum Teil auf die Annahme *biologischer Unterschiede* zwischen den Geschlechtern zurück. Diese beiden Ansätze werden durch die Ergebnisse der internationalen Vergleichsstudien nicht bestätigt.

Neben kognitiven und biologischen Ansätzen gelten in der empirischen Bildungsforschung psychosoziale Erklärungsansätze als geeignet, um Leistungsunterschiede zwischen Mädchen und Jungen zu erklären. *Psychosoziale Ansätze* beruhen auf der Annahme, dass geschlechtstypische Eigenschaften, Einstellungen und Verhaltensweisen in der Interaktion mit dem sozialen Umfeld erlernt werden, wobei Unterschiede zwischen den Geschlechtern durch Sozialisationsprozesse ausgeformt und modifiziert werden (Böhme & Roppelt, 2012). Diesen Annahmen liegt das Erwartungs-Wert-Modell (Eccles, 1983; Eccles, 2007; Wigfield & Eccles, 2002) zugrunde, welches eine Systematisierung derjenigen Faktoren vornimmt, von denen angenommen werden kann, dass sie unter anderem die Einschätzung der eigenen Leistungen (Selbstkonzeptgenese) beeinflussen und eine Vorhersage von leistungsthematischem Verhalten erlauben (Heckhausen & Heckhausen, 2010).

Dem Erwartungs-Wert-Modell zufolge bedingen verschiedene sozialisationsbedingte Aspekte wie kulturelle oder demographische Charakteristika von Familien, in Kombination mit spezifischen Eigenschaften der Eltern, sowohl die Entwicklung von Kindern als auch deren Überzeugungen, Verhaltensweisen, Motivationen und Interessensmuster sowie das tatsächliche Verhalten (Eccles, 2006) und damit auch die Leistung der Kinder (Ehmke & Siegle, 2008).

Anzunehmen ist, dass Überzeugungen und Einstellungen durch Geschlechterstereotype beeinflusst werden, welche sozial geteiltes Wissen über die charakteristischen Merkmale von Frauen und Männern enthalten (Eckes, 2008). Dem Erwartungs-Wert-Modell folgend beeinflussen sowohl deskriptive Anteile (Annahmen darüber, wie Frauen und Männer sind) als auch präskriptive Anteile (Annahmen darüber, wie Frauen und Männer sein sollten) beispielsweise Einstellungen und Selbstkonzepte der Schülerinnen und Schüler und spiegeln sich schließlich auch in dem Lernverhalten und den Lernergebnissen wider (Hannover, 2010; Möller & Trautwein, 2009).

Um Aspekte einer affektiven Bewertung eines Faches vor allem im motivationalen Sinn zu erfassen, wurden in TIMSS 2011 mit der Administration der Schülerfragebögen auch Einstellungen als fachspezifische positive Einstellungen erfasst. Nach einer Klassifizierung von Motivationsformen nach Deci und Ryan (2000) zählen diese Einstellungen zur intrinsischen Motivation. Intrinsische Motivation kann dabei als ein „relativ dauerhaftes, dispositionales Merkmal einer Person verstanden werden, das sich in der Auseinandersetzung mit einem Gegenstandsbereich (z.B. Schulfach) entwickelt und als mehr oder weniger starke Wertschätzung dieses Bereichs zum Ausdruck kommt“ (Schiefele, 2009, S. 163–164). Dies bedeutet konkret, dass Schülerinnen und Schüler beispielsweise Mathematik oder den Sachunterricht mögen oder das jeweilige Fach interessant finden.

Neben der Einstellung wurde in TIMSS 2011 auch das fachspezifische Selbstkonzept erfasst, welches allgemein als organisiertes Wissen über die eigene Person definiert werden kann (Hellmich & Günther, 2011; Moschner, Dickhäuser

& Rost, 2006). Hier geht es konkret ausgedrückt darum, wie Kinder ihre eigenen Fähigkeiten in Mathematik und im Sachunterricht einschätzen. Die Bedeutung des Selbstkonzepts liegt unter anderem darin begründet, dass ein Zusammenhang zwischen dem Selbstkonzept und der Leistung besteht (Hansford & Hattie, 1982). Hier stellt sich allerdings die Frage nach der Kausalität: Beeinflusst die Leistung das Fähigkeitsselbstkonzept oder das Fähigkeitsselbstkonzept die Leistung (Kammermeyer & Martschinke, 2003)? Eine Symbiose dieser beiden Vorstellungen bildet das *Reciprocal-Effects*-Modell, welches postuliert, dass sich Leistung und Fähigkeitsselbstkonzept gegenseitig beeinflussen (Marsh & Craven, 2006; Marsh, Trautwein, Lüdtke, Köller & Baumert, 2005). Aufgrund dieses Zusammenhangs ist es sinnvoll, in Verbindung mit dem Selbstkonzept stets die Leistung mit zu berücksichtigen und dabei möglichst Vorleistungen miteinzubeziehen. Da es sich bei TIMSS 2011 um eine Querschnittuntersuchung handelt, können jedoch lediglich Noten als Vorläuferkompetenzen in die Analysen einbezogen werden (Lintorf, 2012).

Doch neben Einstellung und Selbstkonzept haben auch familiäre Merkmale, beispielsweise das Bildungsniveau der Eltern und der sozioökonomische Status, einen Effekt auf die Leistungen von Schülerinnen und Schülern. Dabei lassen sich die zur Operationalisierung des sozialen Status eingesetzten Instrumente auf der Grundlage der Arbeiten von Pierre Bourdieu einordnen (Bourdieu, 1983), welcher zwischen ökonomischem, kulturellem und sozialem Kapital unterscheidet (siehe Kapitel 7).

Zusammenfassend kann davon ausgegangen werden, dass es sich bei der Entstehung von Geschlechterdisparitäten um ein Zusammenspiel verschiedener individueller Persönlichkeitsmerkmale auf der einen Seite und außerschulischer Voraussetzungen des Kompetenzerwerbs auf der anderen Seite handelt.

Vor dem Hintergrund der erläuterten Forschungsergebnisse und theoretischen Erklärungsansätze stellen sich folgende Fragen:

- Welche Leistungsdifferenzen ergeben sich zwischen Jungen und Mädchen in den Bereichen Mathematik und Naturwissenschaften am Ende der vierten Jahrgangsstufe? Wie unterscheiden sich die Ergebnisse von TIMSS 2007 und TIMSS 2011?
- Unterscheiden sich Jungen und Mädchen in Deutschland in ihrer positiven Einstellung zur Mathematik und zum Sachunterricht sowie in ihrem mathematikbezogenen und sachunterrichtsbezogenen Selbstkonzept? Wie unterscheiden sich die Ausprägungen im Jahr 2011 von den Ergebnissen aus dem Jahr 2007?
- Wie stellen sich Leistungsunterschiede von Jungen und Mädchen in Mathematik und Naturwissenschaften unter Berücksichtigung des fachbezogenen Selbstkonzepts, der fachbezogenen Noten sowie des Bildungsniveaus und des Berufsstatus der Eltern dar?

Im folgenden Abschnitt werden zunächst die Ergebnisse von Jungen und Mädchen in Mathematik und anschließend in den Naturwissenschaften dargestellt. In den Abbildungen und Tabellen werden für den internationalen Vergleich alle Teilnehmerstaaten abgebildet, die Mitglieder der *Europäischen Union* (EU) und/oder der *Organisation for Economic Cooperation and Development* (OECD) sind, sowie zusätzlich alle Staaten, die in den jeweiligen Gesamtskalen in Mathematik und Naturwissenschaften signifikant höhere Leistungsmittelwerte erzielen konnten als Deutschland.

4 Ergebnisse zu geschlechtsspezifischen Leistungsunterschieden in Mathematik

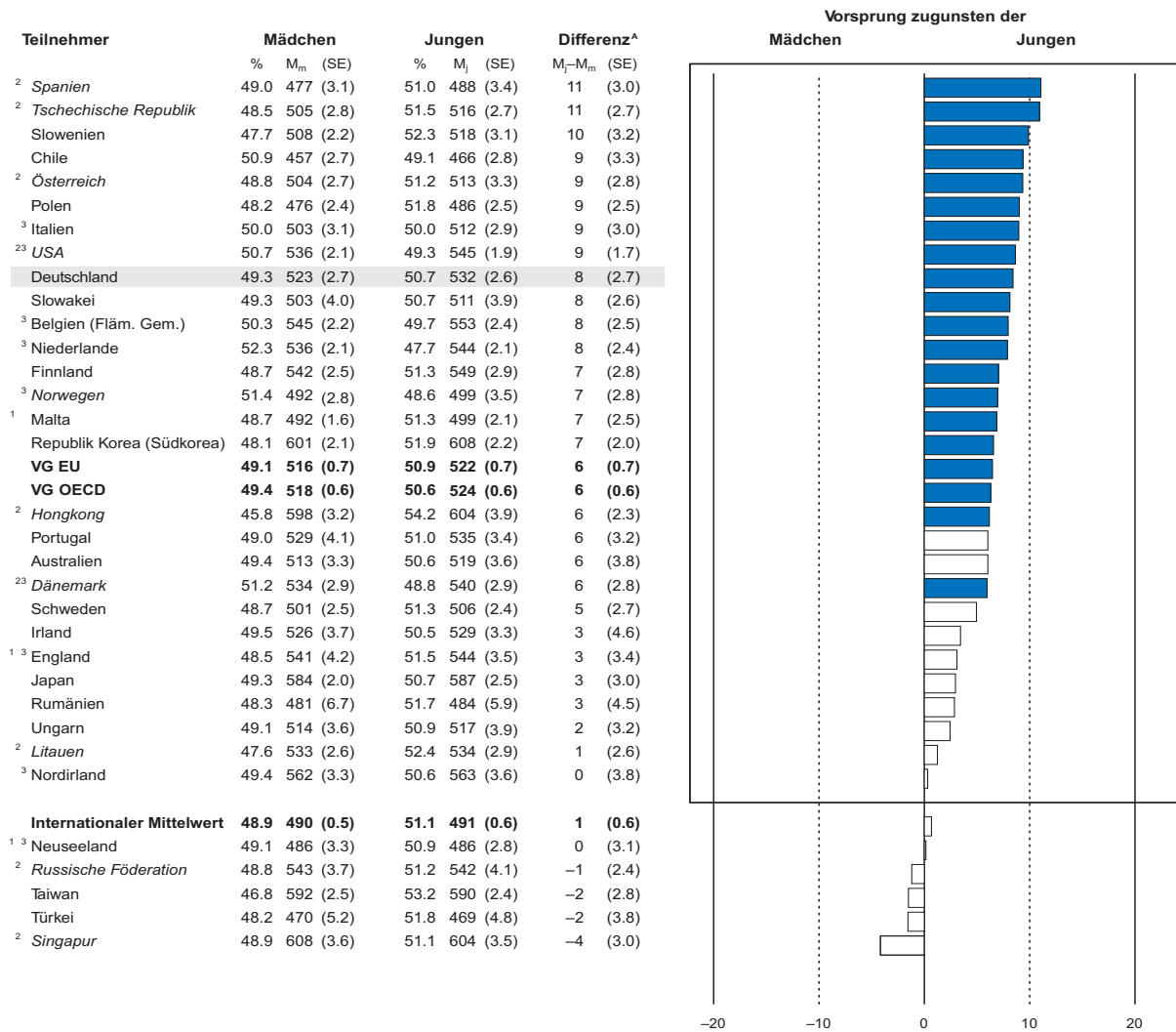
Geschlechtsspezifische Unterschiede in der Mathematikleistung. In der Abbildung 6.1 sind die Ergebnisse der Mathematikgesamtskala für Mädchen und Jungen getrennt dargestellt. Der linke Teil der Abbildung weist tabellarisch die jeweiligen Anteile von Mädchen beziehungsweise Jungen in den einzelnen Teilnehmerstaaten sowie ihre mittleren Mathematikleistungen und die dazugehörigen Standardfehler aus. Zusätzlich sind die jeweiligen Leistungsdifferenzen mit Standardfehlern aufgeführt. Diese Differenzen sind darüber hinaus im rechten Teil der Abbildung als Balkendiagramm illustriert. Statistisch signifikante Unterschiede zwischen Mädchen und Jungen sind durch eine Blaufärbung der Differenzbalken dargestellt. Teilnehmerstaaten, in denen sich die Mittelwertdifferenz nicht signifikant von der Differenz für Deutschland unterscheidet, sind durch eine schwarze Umrandung gekennzeichnet.

Betrachtet man die Ergebnisse, so zeigt sich, dass in allen Teilnehmerstaaten, in denen es einen signifikanten Unterschied in der Mathematikleistung von Mädchen und Jungen gibt, dieser zugunsten der Jungen ausfällt. In Deutschland erreichen Mädchen einen Leistungsmittelwert von 523 Punkten, während Jungen eine mittlere mathematische Kompetenz von 532 Punkten erzielen. Der Unterschied von 8 Punkten ist signifikant. Für 27 Teilnehmer ergeben sich vergleichbare Leistungsunterschiede, wobei diese nicht für alle Teilnehmerstaaten signifikant sind. Der mittlere Unterschied in den Vergleichsgruppen EU (VG_{EU}) und OECD (VG_{OECD}) liegt bei 6 Punkten. Diese Werte unterscheiden sich nicht signifikant von der Leistungsdifferenz zwischen Mädchen und Jungen in Deutschland. Im internationalen Durchschnitt wird eine Leistungsdifferenz von einem Punkt erreicht, die signifikant kleiner ausfällt als die Differenz für Deutschland.

Für die Teilnehmer Portugal, Australien, Schweden, Irland, England, Japan, Rumänien, Ungarn, Litauen, Nordirland, Neuseeland, die Russische Föderation, Taiwan, die Türkei und Singapur zeigen sich keine signifikanten Unterschiede zwischen Mädchen und Jungen hinsichtlich der Mathematikkompetenz.

Geschlechtsspezifische Unterschiede nach mathematischen Inhaltsbereichen. Die TIMSS-Rahmenkonzeption zur Erfassung mathematischer Kompetenzen sieht vor, neben der mittleren Gesamtleistung auch Teilbereiche der mathematischen Kompetenz zu berücksichtigen (vgl. Kapitel 3). Dies ermöglicht eine genauere Analyse der Stärken und Schwächen der Schülerinnen und Schüler. In dem hier vorliegenden Abschnitt werden zunächst die Leistungen von Mädchen und Jungen in den durch die TIMSS-Leistungstests erfassten Inhaltsbereichen *Arithmetik*, *Geometrie/Messen* und *Umgang mit Daten* beschrieben, die in der Abbildung 6.2 dargestellt sind. Im oberen Teil der Abbildung sind für die einzelnen Inhaltsbereiche jeweils die mittleren Leistungswerte und die dazugehörigen Standardfehler für Mädchen und Jungen und zusätzlich die jeweilige Differenz zwischen Mädchen und Jungen mit dem entsprechenden Standardfehler aufgeführt. Analog zu Abbildung 6.1 sind im unteren Teil der Abbildung die Differenzen illustriert. Leistungsunterschiede, die statistisch signifikant ausfallen, sind durch eine blaue Färbung der Balken gekennzeichnet. Die Sortierung der Teilnehmer orientiert sich an der in Abbildung 6.1 dargestellten Größe der mittleren Leistungsdifferenz.

Abbildung 6.1: Mathematikleistung nach Geschlecht – Gesamtskala



□ Kein statistisch signifikanter Unterschied zum Differenzwert von Deutschland ($p > .05$).
 ■ Statistisch signifikante Unterschiede ($p < .05$).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 A= Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Wie in der Abbildung zu sehen ist, ergibt sich ein heterogenes Bild für die drei Inhaltsbereiche. Der Leistungsvorsprung, den die Jungen auf der Gesamtskala Mathematik verzeichnen, ist nicht in allen Bereichen gegeben. In Deutschland erreichen Mädchen im Inhaltsbereich *Arithmetik* 514 Punkte, während der Mittelwert für die Jungen in diesem Bereich bei 526 Punkten liegt. Die Differenz von 12 Punkten zugunsten der Jungen ist signifikant. Für die überwiegende Zahl der Teilnehmer zeigt sich ein vergleichbares Ergebnis. Bei allen Teilnehmern, für die signifikante Differenzen für den Inhaltsbereich *Arithmetik* zu verzeichnen sind, fallen diese zugunsten der Jungen aus. In 13 Teilnehmerstaaten und Regionen ergibt sich kein signifikanter Unterschied. Die mittlere Differenz für die Vergleichsgruppe EU liegt bei 8 Punkten, ebenso wie die mittlere Differenz

für die Vergleichsgruppe OECD. International ergibt sich für den Inhaltsbereich *Arithmetik* eine signifikante Differenz von 3 Punkten zugunsten der Jungen.

Für den Inhaltsbereich *Geometrie/Messen* ergeben sich insgesamt für die hier dargestellten Teilnehmer weniger signifikante Leistungsvorsprünge. In Deutschland erreichen Mädchen einen Mittelwert von 532 Punkten und Jungen einen Mittelwert von 540 Punkten. Ebenso wie in Deutschland ergeben sich auch für die Teilnehmer Chile, Österreich, Italien, die USA, die Slowakei, die Flämische Gemeinschaft in Belgien und Hongkong für den Inhaltsbereich *Geometrie/Messen* signifikant bessere Leistungen für Jungen. Der Leistungsvorsprung der Jungen liegt für die Vergleichsgruppen EU (VG_{EU}) und OECD (VG_{OECD}) bei 4 Punkten. International fällt der Leistungsvorsprung von 2 Punkten signifikant und zugunsten der Mädchen aus. Dabei muss beachtet werden, dass der internationale Mittelwert auch die Teilnehmerstaaten und Regionen einschließt, die in der Abbildung nicht berücksichtigt sind.

Im Inhaltsbereich *Umgang mit Daten* zeigen sich für die einzelnen Teilnehmerstaaten sehr unterschiedliche Ergebnisse. In Deutschland werden in diesem Bereich keine signifikanten Unterschiede zwischen Mädchen und Jungen deutlich. Dies ist auch in 26 weiteren Teilnehmerstaaten und Regionen der Fall. Dagegen schneiden in Spanien, der Tschechischen Republik, Österreich sowie den USA die Jungen signifikant besser ab als die Mädchen. In Neuseeland und Taiwan erzielen die Mädchen signifikant bessere Leistungen in diesem Inhaltsbereich. Für die Vergleichsgruppe EU (VG_{EU}) ergibt sich eine Differenz von 3 Punkten und für die Vergleichsgruppe OECD (VG_{OECD}) von 2 Punkten. International erreichen die Mädchen einen signifikanten Vorsprung von 4 Punkten vor den Jungen.

Insgesamt erreichen in Österreich und den USA die Jungen in allen drei Inhaltsbereichen signifikant höhere Leistungswerte als die Mädchen. In Portugal, Australien, Irland, England, Rumänien, Ungarn, Litauen, Nordirland, der Russischen Föderation, der Türkei und Singapur zeigen sich hingegen in den drei Inhaltsbereichen keine signifikanten Leistungsunterschiede.

Geschlechtsspezifische Unterschiede nach kognitiven Anforderungsbereichen.

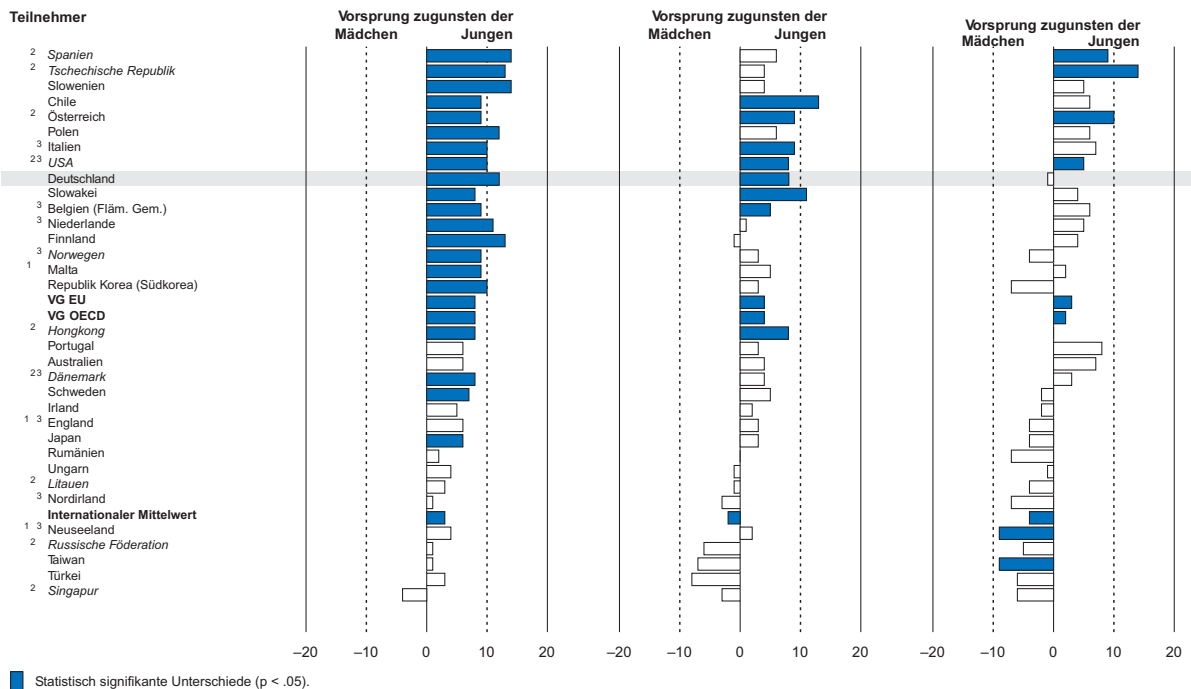
Neben den zuvor beschriebenen Inhaltsbereichen lassen sich nach der TIMSS-Rahmenkonzeption drei kognitive Anforderungsbereiche differenzieren: das *Reproduzieren*, das *Anwenden* und das *Problemlösen*, wobei die Ergebnisse für Jungen und Mädchen im folgenden Abschnitt berichtet werden. Die Ergebnisse im internationalen Vergleich werden aus Platzgründen in diesem Kapitel nicht dargestellt, können aber, mit Ausnahme der Vergleichsgruppen, im internationalen Ergebnisbericht für TIMSS 2011 nachgelesen werden (Mullis, Martin, Foy & Arora, 2012).

Für den Anforderungsbereich *Reproduzieren* ergibt sich für die Mädchen in Deutschland eine mittlere Leistung von 518 Punkten und für die Jungen von 529 Punkten, wobei die Differenz signifikant ist. Die mittlere Differenz zwischen Jungen und Mädchen für die Vergleichsgruppen EU (VG_{EU}) und OECD (VG_{OECD}) liegt bei 6 Punkten. Im internationalen Mittelwert ergeben sich für den Anforderungsbereich *Reproduzieren* keine Unterschiede zwischen Mädchen und Jungen.

Im Anforderungsbereich *Anwenden* haben die Jungen in Deutschland mit 531 Punkten 6 Punkte Vorsprung vor den Mädchen. Im EU-Durchschnitt ergibt sich eine mittlere Differenz zugunsten der Jungen von 7 Punkten und im

Abbildung 6.2: Leistungen in mathematischen Inhaltsbereichen nach Geschlecht im internationalen Vergleich

Teilnehmer	Arithmetik			Geometrie/Messen			Umgang mit Daten		
	M _m (SE)	M _j (SE)	Differenz ^A M _j - M _m (SE)	M _m (SE)	M _j (SE)	Differenz ^A M _j - M _m (SE)	M _m (SE)	M _j (SE)	Differenz ^A M _j - M _m (SE)
² Spanien	479 (3,2)	494 (3,6)	14 (3,2)	473 (3,2)	479 (3,7)	6 (3,4)	474 (3,9)	484 (4,3)	9 (4,1)
² Tschechische Republik	502 (3,0)	515 (3,0)	13 (3,3)	511 (3,4)	515 (3,7)	4 (3,7)	512 (4,3)	526 (4,2)	14 (6,0)
Slowenien	496 (2,6)	510 (3,4)	14 (3,0)	524 (2,8)	528 (3,1)	4 (3,7)	530 (2,7)	535 (3,7)	5 (3,8)
Chile	457 (3,4)	466 (3,3)	9 (3,8)	449 (3,6)	462 (3,5)	13 (4,0)	463 (3,6)	468 (3,9)	6 (5,6)
² Österreich	502 (2,7)	511 (3,1)	9 (2,9)	507 (3,5)	516 (4,7)	9 (4,6)	510 (3,0)	520 (4,1)	10 (3,3)
Polen	474 (2,7)	486 (2,7)	12 (3,1)	472 (2,9)	478 (3,4)	6 (3,4)	486 (5,4)	491 (3,7)	6 (7,1)
³ Italien	505 (3,1)	515 (3,0)	10 (2,9)	508 (3,5)	517 (3,5)	9 (3,5)	491 (3,1)	498 (4,1)	7 (3,8)
^{2,3} USA	538 (2,1)	548 (2,2)	10 (1,7)	531 (2,4)	539 (2,4)	8 (2,2)	542 (2,2)	547 (1,8)	5 (1,9)
Deutschland	514 (2,7)	526 (2,7)	12 (2,9)	532 (3,1)	540 (2,9)	8 (2,9)	546 (3,9)	545 (3,6)	-1 (4,9)
Slowakei	507 (4,1)	515 (3,7)	8 (2,7)	494 (4,7)	506 (4,3)	11 (3,0)	502 (5,6)	506 (4,2)	4 (3,3)
³ Belgien (Fläm. Gem.)	547 (2,5)	556 (2,4)	9 (2,4)	549 (2,4)	555 (2,1)	5 (2,2)	533 (4,0)	539 (4,0)	6 (5,3)
³ Niederlande	538 (2,3)	549 (2,1)	11 (2,9)	523 (2,2)	525 (4,1)	1 (3,0)	557 (2,9)	562 (4,4)	5 (4,6)
Finnland	538 (2,5)	552 (2,8)	13 (2,6)	544 (3,5)	543 (3,7)	-1 (4,1)	549 (4,7)	553 (4,7)	4 (6,3)
³ Norwegen	484 (3,2)	493 (4,0)	9 (3,6)	505 (3,9)	508 (3,8)	3 (4,8)	496 (4,3)	492 (4,7)	-4 (6,2)
¹ Malta	493 (1,9)	502 (2,8)	9 (3,0)	484 (2,1)	489 (2,0)	5 (2,9)	497 (2,5)	499 (3,6)	2 (5,3)
Republik Korea (Südkorea)	600 (2,3)	610 (2,2)	10 (2,0)	606 (3,0)	608 (2,0)	3 (3,6)	607 (3,9)	599 (3,5)	-7 (6,3)
VG EU	515 (0,7)	524 (0,7)	8 (0,7)	517 (0,8)	521 (0,8)	4 (0,9)	520 (0,9)	523 (0,9)	3 (1,1)
VG OECD	516 (0,6)	524 (0,6)	8 (0,6)	520 (0,7)	524 (0,7)	4 (0,7)	524 (0,8)	526 (0,8)	2 (1,0)
² Hongkong	600 (3,2)	608 (4,0)	8 (3,0)	600 (3,3)	609 (4,1)	8 (3,1)	593 (3,5)	593 (4,8)	0 (4,3)
Portugal	519 (4,5)	525 (3,6)	6 (3,3)	546 (4,5)	550 (4,9)	3 (3,7)	544 (3,5)	552 (3,7)	8 (4,7)
Australien	505 (3,6)	511 (4,1)	6 (4,3)	532 (3,5)	536 (4,1)	4 (4,7)	512 (4,0)	519 (3,8)	7 (4,7)
^{2,3} Dänemark	530 (2,7)	538 (2,9)	8 (2,8)	546 (3,5)	550 (3,6)	4 (3,8)	530 (4,3)	533 (4,2)	3 (6,2)
Schweden	497 (2,8)	504 (2,7)	7 (3,2)	497 (3,0)	502 (2,8)	5 (3,3)	525 (3,9)	522 (3,6)	-2 (4,6)
Irland	530 (3,8)	535 (3,3)	5 (4,7)	519 (4,4)	521 (3,8)	2 (5,3)	524 (3,6)	522 (5,0)	-2 (6,7)
^{1,3} England	536 (4,3)	542 (3,8)	6 (3,5)	544 (4,6)	547 (4,1)	3 (4,0)	551 (6,3)	547 (4,9)	-4 (6,4)
Japan	581 (2,0)	587 (2,2)	6 (2,8)	588 (2,4)	591 (2,6)	3 (3,1)	591 (2,3)	588 (4,4)	-4 (3,8)
Rumänien	496 (6,2)	498 (5,9)	2 (4,2)	468 (7,1)	469 (6,1)	0 (6,5)	460 (7,6)	453 (6,9)	-7 (5,2)
Ungarn	513 (3,4)	517 (3,8)	4 (3,3)	521 (4,2)	520 (4,1)	-1 (4,3)	510 (4,9)	510 (4,8)	-1 (4,7)
² Litauen	536 (2,9)	539 (2,9)	3 (3,2)	531 (3,7)	530 (3,3)	-1 (3,7)	528 (3,0)	524 (4,1)	-4 (4,1)
³ Nordirland	566 (3,3)	567 (3,8)	1 (3,9)	561 (3,8)	559 (4,3)	-3 (4,7)	558 (3,8)	552 (4,1)	-7 (5,1)
Internationaler Mittelwert	493 (0,5)	496 (0,6)	3 (0,6)	485 (0,6)	483 (0,7)	-2 (0,7)	486 (0,7)	482 (0,7)	-4 (0,8)
^{1,3} Neuseeland	481 (3,2)	485 (3,1)	4 (3,8)	482 (2,9)	484 (3,1)	2 (3,4)	496 (3,3)	487 (3,6)	-9 (4,2)
² Russische Föderation	544 (3,4)	545 (3,5)	1 (2,3)	545 (4,1)	538 (4,9)	-6 (3,3)	535 (5,3)	530 (4,6)	-5 (5,6)
Taiwan	599 (2,7)	600 (2,5)	1 (3,2)	576 (2,9)	570 (2,6)	-7 (3,5)	605 (2,8)	596 (3,8)	-9 (4,3)
Türkei	475 (5,4)	478 (4,4)	3 (4,0)	451 (5,7)	443 (5,2)	-8 (4,2)	481 (6,0)	475 (5,2)	-6 (4,3)
² Singapur	621 (3,7)	617 (3,8)	-4 (3,2)	591 (3,9)	588 (4,0)	-3 (3,0)	591 (4,1)	584 (4,4)	-6 (5,1)



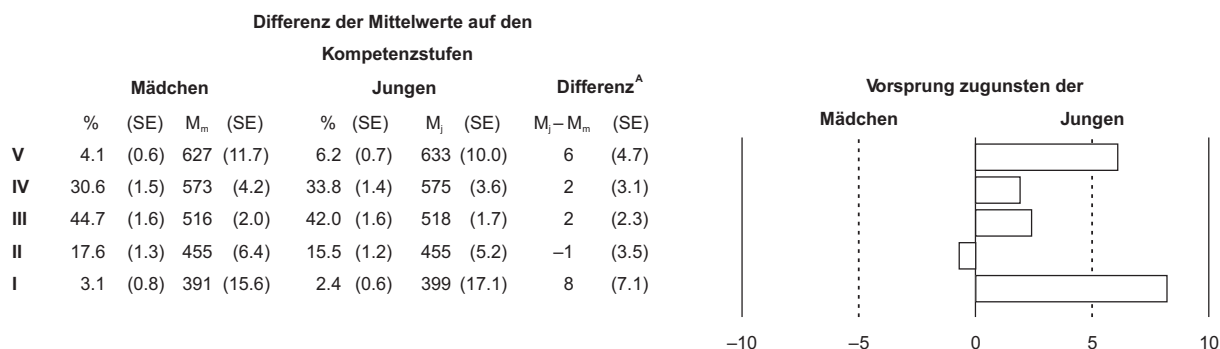
■ Statistisch signifikante Unterschiede ($p < .05$).
 Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.
 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerenebene erreichen nicht die internationalen Vorgaben.
 A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

OECD-Durchschnitt von 6 Punkten. Im internationalen Mittelwert zeigt sich eine Differenz von einem Punkt.

In Deutschland ergibt sich im Anforderungsbereich *Problemlösen* ein Leistungsunterschied von 11 Punkten zugunsten der Jungen. Für die Vergleichsgruppen EU (VG_{EU}) und OECD (VG_{OECD}) zeigt sich jeweils ein Vorsprung der Jungen von 7 Punkten. International beträgt der Leistungsvorsprung der Jungen 2 Punkte.

Mittlere Leistung auf den Kompetenzstufen. In der Abbildung 6.3 sind die mittleren Leistungen sowie die prozentualen Anteile von Jungen und Mädchen auf den Kompetenzstufen dargestellt. Die mittleren Leistungen, die Mädchen und Jungen auf den jeweiligen Kompetenzstufen erreichen, unterscheiden sich nicht signifikant voneinander. Betrachtet man hingegen die beiden oberen Kompetenzstufen und dort die Geschlechterdifferenzen, so zeigt sich, dass sich signifikant mehr Jungen auf den Kompetenzstufen IV und V befinden.

Abbildung 6.3: Mittlere Leistungen und prozentuale Verteilungen von Jungen und Mädchen auf den Kompetenzstufen der Mathematik



■ Statistisch signifikante Unterschiede ($p < .05$).

A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

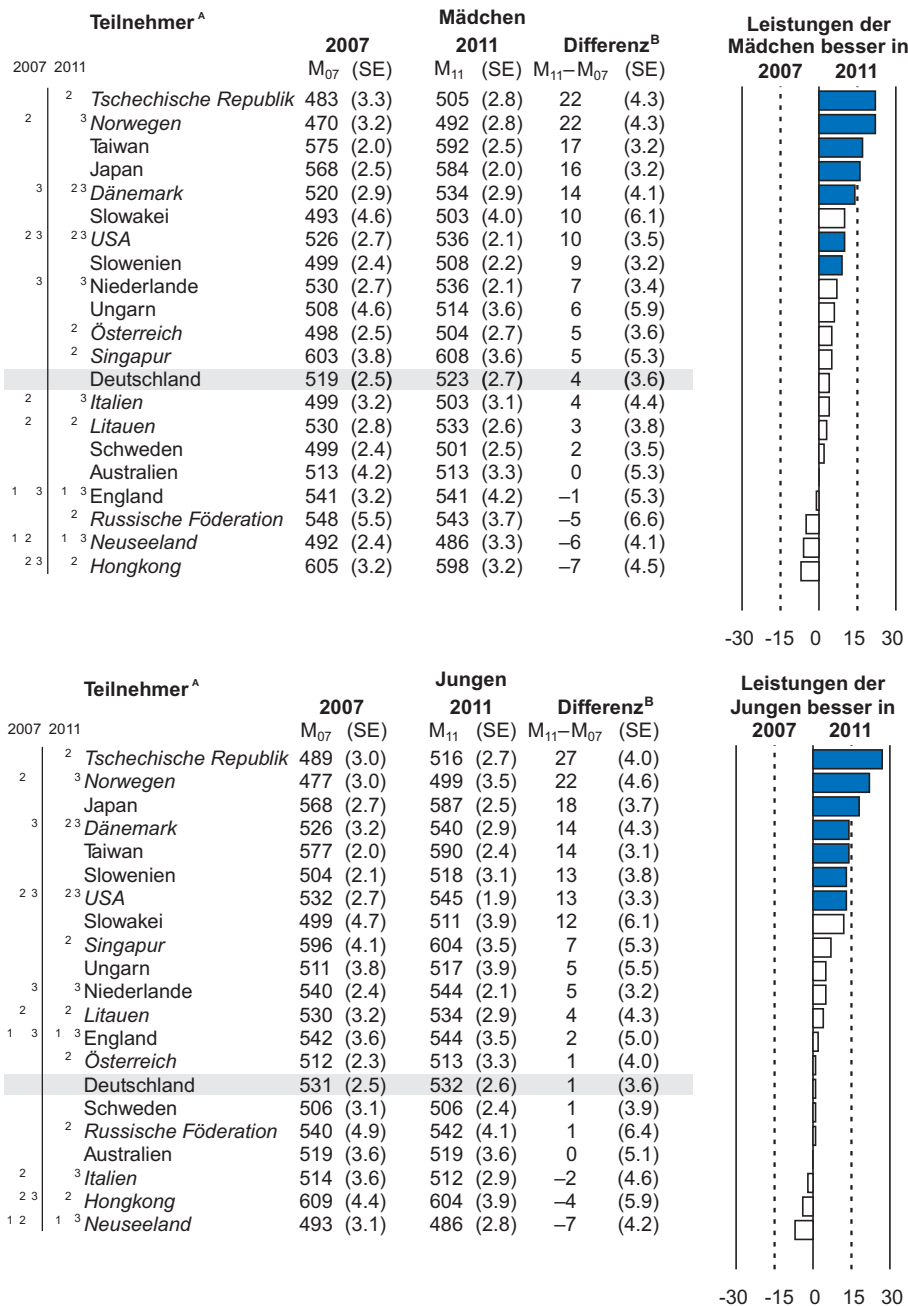
© TIMSS 2011

Vergleich der Kompetenzen in TIMSS 2007 und TIMSS 2011. Im Folgenden sollen noch die Veränderungen der geschlechtsspezifischen Leistungsunterschiede im Vergleich von TIMSS 2007 und TIMSS 2011 dargestellt werden. In Abbildung 6.4 sind für die hier berichteten Teilnehmerstaaten die mittleren mathematischen Leistungswerte und die Standardfehler für die Erhebungen 2007 und 2011 für Mädchen und Jungen sowie die Leistungsdifferenzen zwischen den beiden Messzeitpunkten angegeben. Ergänzend ist die Größe der jeweiligen Differenzen illustriert und signifikante Unterschiede farblich hervorgehoben. Die Ergebnisse für die Mädchen sind der linken Seite der Abbildung zu entnehmen, während die Werte für die Jungen rechts dargestellt sind.

Im internationalen Vergleich der beiden Messzeitpunkte zeigt sich, dass die getesteten Mädchen in der Tschechischen Republik, Norwegen, Taiwan, Japan, Dänemark, den USA und Slowenien 2011 eine signifikant höhere Leistung erreicht haben. Für 14 Teilnehmer, darunter auch Deutschland, ergeben sich keine signifikanten Veränderungen in den Mathematikleistungen der Mädchen.

In der Tschechischen Republik, Norwegen, Japan, Dänemark, Taiwan, Slowenien und den USA schneiden die Jungen in 2011 signifikant besser ab als

Abbildung 6.4: Vergleich der Leistungsunterschiede in Mathematik von Jungen und Mädchen in TIMSS 2007 und TIMSS 2011 (internationaler Vergleich)



■ Statistisch signifikante Unterschiede (p < .05).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülererebene erreichen nicht die internationalen Vorgaben.

A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

in 2007. Ähnlich wie in 13 weiteren Teilnehmerstaaten zeigt sich in Deutschland keine signifikante Veränderung der Leistungsergebnisse der Jungen.

Insgesamt wird deutlich, dass sich weder die mittleren Mathematikleistungen der Jungen noch die der Mädchen in Deutschland signifikant verändert haben. Außerdem kann festgestellt werden, dass es keinem Mitgliedstaat der EU oder OECD gelungen ist im Vergleich zu TIMSS 2007 Leistungsunterschiede in Mathematik zwischen den Geschlechtern signifikant zu verringern.

Werden allerdings die Leistungsdifferenzen zwischen Jungen und Mädchen in den einzelnen Inhaltsbereichen und Anforderungsbereichen betrachtet, ergeben sich einige Unterschiede zwischen TIMSS 2007 und TIMSS 2011. So zeigt sich, dass sich die Leistungen der Jungen in Deutschland vor allem im Inhaltsbereich *Geometrie/Messen* und im Anforderungsbereich *Reproduzieren* signifikant verbessert haben, während Mädchen im Inhaltsbereich *Umgang mit Daten* sowie im Anforderungsbereich *Reproduzieren* 2011 signifikant höhere Leistungen erzielen als 2007. Die Leistungsentwicklungen von Jungen und Mädchen müssen demnach nicht zwangsläufig parallel verlaufen.

5 Ergebnisse zu geschlechtsspezifischen Leistungsunterschieden in den Naturwissenschaften

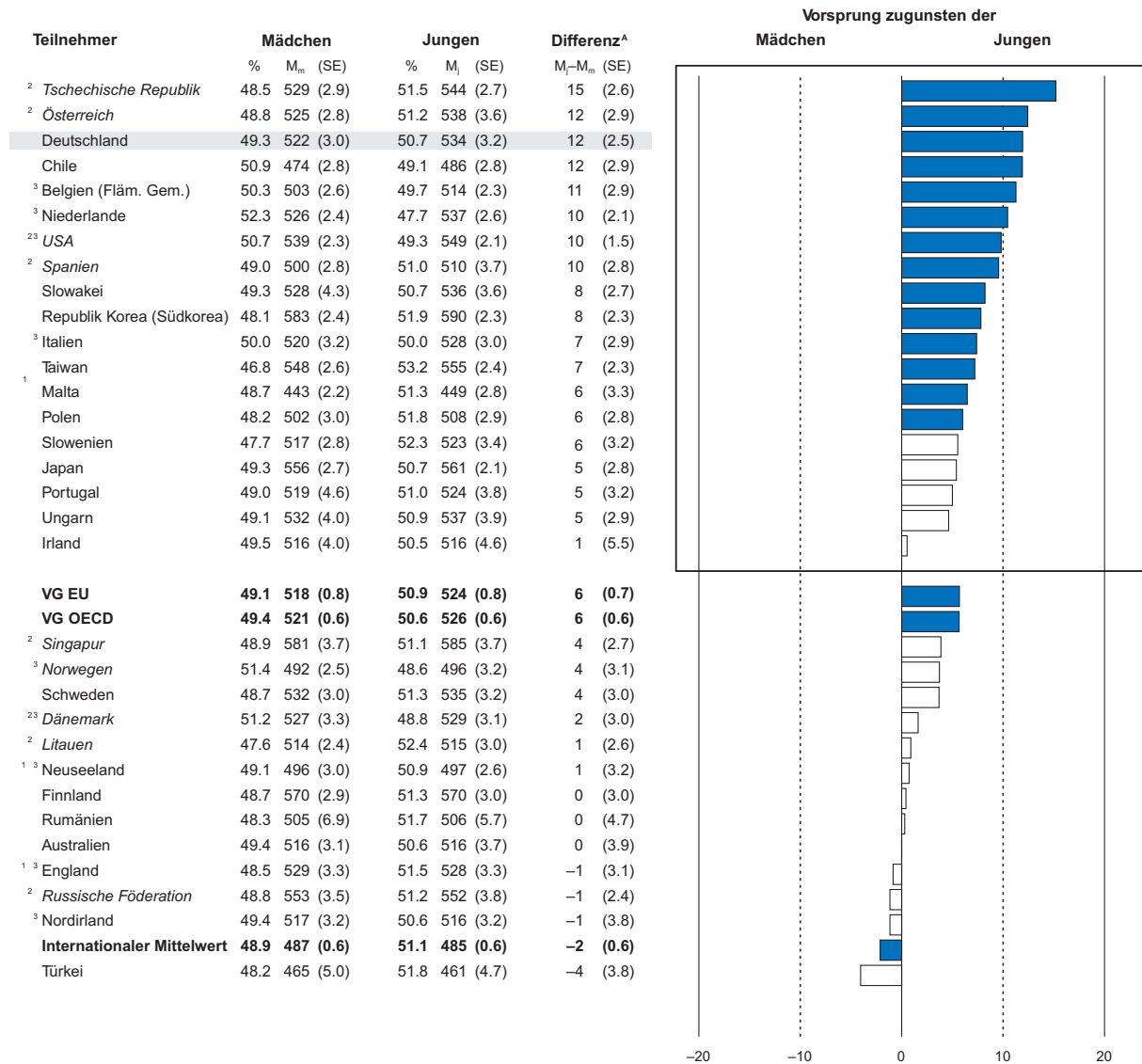
Nachdem die Leistungen von Jungen und Mädchen in Mathematik dargestellt wurden, werden im folgenden Abschnitt die geschlechtsspezifischen Leistungsunterschiede in den Naturwissenschaften betrachtet. Dabei werden die Ergebnisse identisch zum vorangegangenen Abschnitt dargestellt, um eine Vergleichbarkeit der beiden Domänen zu ermöglichen.

Geschlechtsspezifische Unterschiede in der naturwissenschaftlichen Leistung. Die Abbildung 6.5 zeigt analog zur Abbildung 6.1 die Ergebnisse der Gesamtskala Naturwissenschaften getrennt für Mädchen und Jungen.

In Deutschland erreichen Mädchen einen Leistungsmittelwert von 522 Punkten, während Jungen eine mittlere naturwissenschaftliche Leistung von 534 Punkten erzielen. Der signifikante Leistungsunterschied von 12 Punkten in Deutschland (sowie in Österreich und Chile) ist im internationalen Vergleich nur in der Tschechischen Republik mit 15 Punkten höher ausgeprägt. Für 18 Teilnehmer ergeben sich von Deutschland nicht signifikant abweichende Leistungsunterschiede zwischen Jungen und Mädchen, wobei diese nicht für alle Teilnehmerstaaten signifikant ausfallen. Der mittlere Unterschied der Vergleichsgruppen EU (VG_{EU}) und OECD (VG_{OECD}) liegt bei jeweils 6 Punkten. Diese Werte unterscheiden sich signifikant von der Leistungsdifferenz zwischen Mädchen und Jungen in Deutschland. Im internationalen Durchschnitt beträgt die Leistungsdifferenz 2 Punkte. Diese fällt signifikant kleiner aus als die Differenz für Deutschland, wobei sich im internationalen Durchschnitt ein Leistungsvorsprung der Mädchen zeigt. Dabei ist zu beachten, dass in den internationalen Mittelwert die Ergebnisse weiterer Staaten eingegangen sind, die hier nicht berichtet werden.

Für die Teilnehmer Slowenien, Japan, Portugal, Ungarn, Irland, Singapur, Norwegen, Schweden, Dänemark, Litauen, Neuseeland, Finnland, Rumänien, Australien, England, die Russische Föderation, Nordirland und die Türkei zeigen sich keine signifikanten Unterschiede zwischen Mädchen und Jungen hinsichtlich der naturwissenschaftlichen Kompetenzen.

Abbildung 6.5: Naturwissenschaftsleistung nach Geschlecht – Gesamtskala



□ Kein statistisch signifikanter Unterschied zum Differenzwert von Deutschland ($p > .05$).

■ Statistisch signifikante Unterschiede ($p < .05$).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A= Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

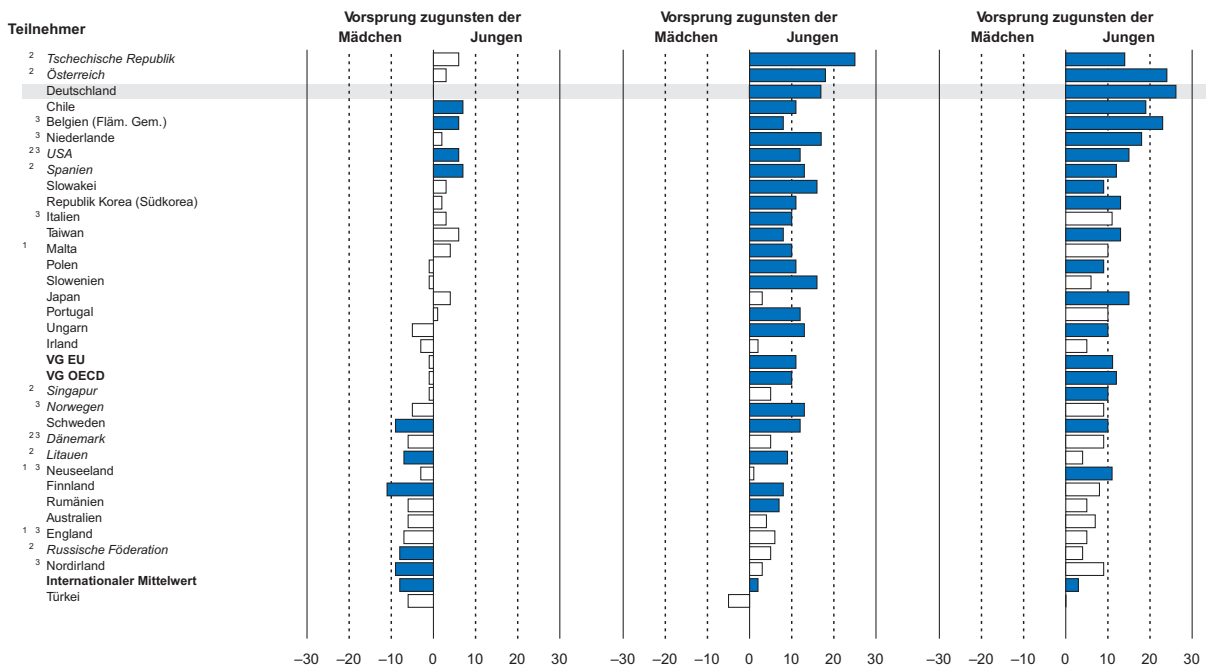
© TIMSS 2011

Geschlechtsspezifische Unterschiede nach Inhaltsbereichen. Um die Stärken und Schwächen von Schülerinnen und Schülern genauer bestimmen zu können, werden die Ergebnisse analog zu den Mathematikleistungen nach naturwissenschaftlichen Inhaltsbereichen und Anforderungsbereichen differenziert dargestellt. Dabei stellt die Abbildung 6.6 die Leistungen von Jungen und Mädchen in den Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geographie* dar und ist identisch zur Abbildung 6.2 aufgebaut. Die Sortierung der Teilnehmer orientiert sich an der in Abbildung 6.5 dargestellten Größe der mittleren Leistungsdifferenzen.

Wie in der Abbildung 6.6 zu sehen ist, ergibt sich ein heterogenes Bild für die drei Inhaltsbereiche und der Leistungsvorsprung, den die Jungen auf der Gesamtskala Naturwissenschaften verzeichnen konnten, ist nicht in allen Bereichen gegeben.

Abbildung 6.6: Leistungen in naturwissenschaftlichen Inhaltsbereichen nach Geschlecht im internationalen Vergleich

Teilnehmer	Biologie			Physik/Chemie			Geographie		
	Mädchen	Jungen	Differenz ^A	Mädchen	Jungen	Differenz ^A	Mädchen	Jungen	Differenz ^A
	M _m (SE)	M _j (SE)	M _j - M _m (SE)	M _m (SE)	M _j (SE)	M _j - M _m (SE)	M _m (SE)	M _j (SE)	M _j - M _m (SE)
² Tschechische Republik	547 (3.7)	552 (3.5)	6 (4.2)	506 (3.6)	531 (3.7)	25 (3.5)	530 (3.9)	544 (4.3)	14 (4.5)
² Österreich	525 (2.8)	527 (3.7)	3 (3.8)	526 (3.2)	544 (4.0)	18 (4.2)	526 (4.6)	550 (4.6)	24 (5.5)
Deutschland	525 (2.8)	525 (3.1)	0 (2.9)	526 (4.0)	543 (3.2)	17 (3.9)	507 (4.2)	533 (4.6)	26 (4.9)
Chile	486 (2.7)	493 (2.8)	7 (3.2)	465 (3.6)	477 (3.3)	11 (4.8)	465 (3.3)	485 (3.2)	19 (3.6)
³ Belgien (Fläm. Gem.)	507 (3.0)	513 (2.7)	6 (2.9)	503 (2.6)	511 (2.5)	8 (3.0)	493 (3.5)	516 (3.0)	23 (3.4)
³ Niederlande	536 (2.1)	538 (2.8)	2 (3.3)	518 (2.4)	535 (3.0)	17 (3.7)	517 (4.4)	534 (2.9)	18 (5.2)
^{2,3} USA	544 (2.4)	550 (2.1)	6 (1.9)	538 (2.4)	550 (2.4)	12 (2.6)	531 (2.6)	547 (2.1)	15 (1.9)
² Spanien	510 (2.9)	516 (3.4)	7 (2.8)	490 (3.0)	503 (3.3)	13 (3.2)	493 (4.7)	505 (3.9)	12 (4.3)
Slowakei	533 (4.3)	535 (4.2)	3 (4.8)	519 (4.5)	535 (4.1)	16 (3.0)	530 (4.4)	540 (3.6)	9 (2.9)
Republik Korea (Südkorea)	570 (2.2)	572 (2.9)	2 (2.6)	591 (4.1)	602 (2.5)	11 (4.2)	596 (2.8)	610 (2.6)	13 (4.0)
³ Italien	534 (2.9)	537 (3.2)	3 (3.1)	504 (3.3)	514 (4.2)	10 (4.6)	518 (4.2)	529 (5.2)	11 (6.1)
Taiwan	535 (3.1)	541 (2.7)	6 (3.2)	564 (2.6)	572 (2.3)	8 (2.6)	546 (3.5)	559 (2.6)	13 (3.6)
¹ Malta	437 (3.0)	440 (3.0)	4 (3.7)	448 (3.4)	458 (3.2)	10 (4.3)	442 (2.5)	452 (4.3)	10 (5.5)
Polen	514 (2.8)	513 (3.0)	-1 (3.0)	489 (4.5)	500 (3.6)	11 (4.6)	491 (3.4)	500 (4.3)	9 (4.2)
Slowenien	525 (3.4)	523 (3.3)	-1 (4.3)	515 (3.7)	531 (4.2)	16 (4.1)	503 (3.4)	509 (3.4)	6 (4.0)
Japan	538 (1.9)	542 (2.6)	4 (2.5)	588 (2.7)	590 (2.2)	3 (3.1)	544 (2.6)	559 (2.5)	15 (3.6)
Portugal	520 (5.2)	521 (4.0)	1 (4.0)	511 (4.7)	523 (4.2)	12 (2.9)	526 (5.7)	536 (5.0)	10 (6.0)
Ungarn	554 (4.0)	549 (3.8)	-5 (3.5)	514 (4.3)	527 (4.2)	13 (4.0)	519 (5.1)	529 (4.7)	10 (4.4)
Irland	514 (4.6)	511 (4.1)	-3 (5.0)	516 (4.5)	518 (3.6)	2 (5.3)	518 (4.2)	522 (4.7)	5 (4.6)
VG EU	525 (0.8)	523 (0.8)	-1 (0.8)	512 (0.8)	524 (0.8)	11 (0.9)	512 (0.9)	523 (0.9)	11 (1.1)
VG OECD	526 (0.7)	525 (0.7)	-1 (0.7)	517 (0.7)	527 (0.7)	10 (0.8)	516 (0.8)	528 (0.8)	12 (0.9)
² Singapur	598 (4.5)	597 (4.7)	-1 (3.1)	596 (3.6)	601 (4.0)	5 (3.0)	536 (3.9)	546 (3.8)	10 (4.9)
² Norwegen	498 (3.5)	493 (3.6)	-5 (3.7)	476 (4.2)	489 (3.8)	13 (4.1)	502 (3.8)	511 (4.1)	9 (5.1)
Schweden	538 (2.8)	530 (3.5)	-9 (3.3)	521 (3.0)	534 (2.9)	12 (3.1)	533 (4.0)	543 (3.7)	10 (4.2)
^{2,3} Dänemark	533 (3.4)	527 (3.4)	-6 (3.8)	523 (3.5)	528 (3.1)	5 (4.4)	522 (4.1)	531 (5.1)	9 (6.9)
² Litauen	524 (3.4)	517 (3.1)	-7 (2.8)	510 (3.7)	518 (3.3)	9 (3.5)	498 (3.8)	503 (3.7)	4 (4.7)
^{1,3} Neuseeland	499 (3.2)	496 (3.1)	-3 (3.9)	493 (3.1)	494 (3.2)	1 (3.2)	494 (4.3)	504 (3.2)	11 (4.0)
Finnland	580 (2.8)	569 (3.6)	-11 (3.2)	564 (3.6)	572 (3.2)	8 (3.8)	562 (3.2)	569 (3.7)	8 (4.0)
Rumänien	507 (6.8)	501 (6.4)	-6 (5.3)	504 (6.3)	512 (5.7)	7 (3.8)	499 (7.0)	504 (6.1)	5 (5.4)
Australien	518 (3.5)	513 (3.8)	-6 (3.7)	512 (3.3)	516 (4.3)	4 (4.2)	516 (4.8)	523 (3.8)	7 (5.0)
^{1,3} England	534 (3.6)	527 (4.1)	-7 (5.2)	532 (3.5)	538 (4.9)	6 (4.8)	520 (4.5)	524 (3.9)	5 (3.7)
² Russische Föderation	561 (3.8)	552 (4.0)	-8 (3.2)	546 (4.1)	551 (4.7)	5 (3.7)	551 (4.9)	554 (4.6)	4 (4.7)
³ Nordirland	523 (3.5)	514 (3.4)	-9 (4.0)	519 (3.5)	522 (3.8)	3 (3.7)	503 (3.8)	512 (4.8)	9 (6.9)
Internationaler Mittelwert	489 (0.6)	481 (0.6)	-8 (0.7)	484 (0.6)	485 (0.7)	2 (0.7)	479 (0.7)	483 (0.7)	3 (0.8)
Türkei	463 (5.2)	457 (4.5)	-6 (3.5)	469 (5.1)	464 (5.0)	-5 (3.7)	456 (5.7)	455 (5.9)	0 (5.5)



■ Statistisch signifikante Unterschiede (p < .05).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülererebene erreichen nicht die internationalen Vorgaben.

A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Im Inhaltsbereich *Biologie* zeigt sich ein sehr heterogenes Bild der Leistungsunterschiede zwischen Jungen und Mädchen. Während in Deutschland Mädchen und Jungen eine identische Punktzahl (525) erreichen, zeigen sich in anderen Staaten (mitunter deutliche) Unterschiede. So erzielen in Chile, der Flämischen Gemeinschaft in Belgien, den USA und Spanien die Jungen signifikant höhere Leistungswerte. In Schweden, Litauen, Finnland, der Russischen Föderation und Nordirland zeigt sich hingegen, dass Mädchen im Inhaltsbereich *Biologie* höhere Kompetenzen als Jungen aufweisen. Insgesamt in 23 Teilnehmerstaaten und Regionen ergibt sich kein signifikanter Unterschied. Auch in den Vergleichsgruppen EU (VG_{EU}) und OECD (VG_{OECD}) ergeben sich keine signifikanten Leistungsunterschiede im Inhaltsbereich *Biologie*. International zeigt sich für den Inhaltsbereich eine signifikante Differenz von 8 Punkten zugunsten der Mädchen.

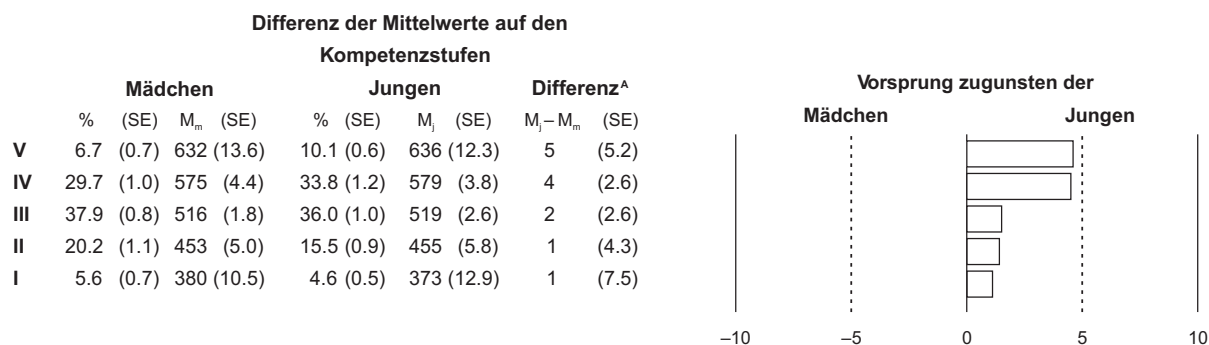
Für den Inhaltsbereich *Physik/Chemie* ergibt sich ein einheitliches Bild der geschlechtsspezifischen Leistungsunterschiede für die einzelnen Teilnehmer. Für Deutschland ist in diesem Inhaltsbereich ein signifikanter Unterschied von 17 Leistungspunkten zugunsten der Jungen zu verzeichnen – die Mädchen erreichen im Mittel 526 Punkte und die Jungen 543 Punkte. Ein ähnliches Ergebnis kann für 22 weitere Teilnehmerstaaten und Regionen festgehalten werden. Dagegen zeigen sich in Japan, Irland, Singapur, Dänemark, Neuseeland, Australien, England, der Russischen Föderation, Nordirland sowie der Türkei keine signifikanten Unterschiede zwischen Jungen und Mädchen. Für die Vergleichsgruppe EU (VG_{EU}) ergibt sich eine Differenz von 11 und für die Vergleichsgruppe OECD (VG_{OECD}) eine Differenz von 10 Punkten. International erreichen die Jungen einen signifikanten Vorsprung von 2 Punkten.

Ähnlich wie für den Inhaltsbereich *Physik/Chemie* sind die Leistungsunterschiede für den Inhaltsbereich *Geographie* zwischen Jungen und Mädchen über die Teilnehmer hinweg relativ homogen. In Deutschland erreichen Mädchen einen Mittelwert von 507 und Jungen einen Mittelwert von 533 Punkten. Ebenso wie in Deutschland ergeben sich auch für 16 weitere Teilnehmer signifikant bessere Leistungen für Jungen als für Mädchen, wobei im Inhaltsbereich *Geographie* die Differenz im internationalen Vergleich für Deutschland am größten ausfällt. Der Leistungsvorsprung der Jungen liegt in der Vergleichsgruppe EU (VG_{EU}) bei 11 Punkten und in der Vergleichsgruppe OECD (VG_{OECD}) bei 12 Punkten. International fällt der Leistungsvorsprung von 3 Punkten signifikant und zugunsten der Jungen aus.

Insgesamt erzielen in Chile, Belgien, den USA und Spanien die Jungen in allen drei Inhaltsbereichen signifikant höhere Leistungswerte. In Irland, Dänemark, Australien und England zeigen sich hingegen in den drei Inhaltsbereichen keine signifikanten geschlechtsspezifischen Leistungsunterschiede.

Geschlechtsspezifische Unterschiede nach kognitiven Anforderungsbereichen. Wie in Mathematik werden nach der TIMSS-Rahmenkonzeption auch für die Naturwissenschaften die drei kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen* differenziert, wobei die Ergebnisse für Jungen und Mädchen in Deutschland im folgenden Abschnitt berichtet werden. Die Ergebnisse aller, außer denen für die Vergleichsgruppen, können im internationalen Ergebnisbericht nachgelesen werden (Martin, Mullis, Foy & Stanco, 2012).

Für den Anforderungsbereich *Reproduzieren* zeigt sich in Deutschland, dass die Jungen mit 531 Punkten einen Vorsprung von 14 Punkten vor den Mädchen haben. Im EU-Durchschnitt haben mit 7 Punkten und im OECD-Durchschnitt mit 8 Punkten die Jungen einen signifikanten Leistungsvorsprung vor den Mädchen.

Abbildung 6.7: Mittlere Leistungen und prozentuale Verteilungen von Jungen und Mädchen auf den Kompetenzstufen der Naturwissenschaften

■ Statistisch signifikante Unterschiede ($p < .05$).

A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Im internationalen Mittel zeigt sich eine Differenz von einem Punkt zugunsten der Mädchen. Dieser Unterschied ist jedoch nicht signifikant.

Im Anforderungsbereich *Anwenden* zeigt sich, dass Mädchen in Deutschland eine mittlere Leistung von 527 Punkten und Jungen von 539 Punkten erreichen, wobei die Differenz von 12 Punkten signifikant ist. Die mittlere Differenz für die Vergleichsgruppen EU (VG_{EU}) und OECD (VG_{OECD}) liegt bei 6 Punkten. Im internationalen Mittel ergibt sich für den Anforderungsbereich *Anwenden* ebenfalls ein Leistungsvorsprung von einem Punkt zugunsten der Mädchen, wobei diese Differenz wiederum nicht signifikant ist.

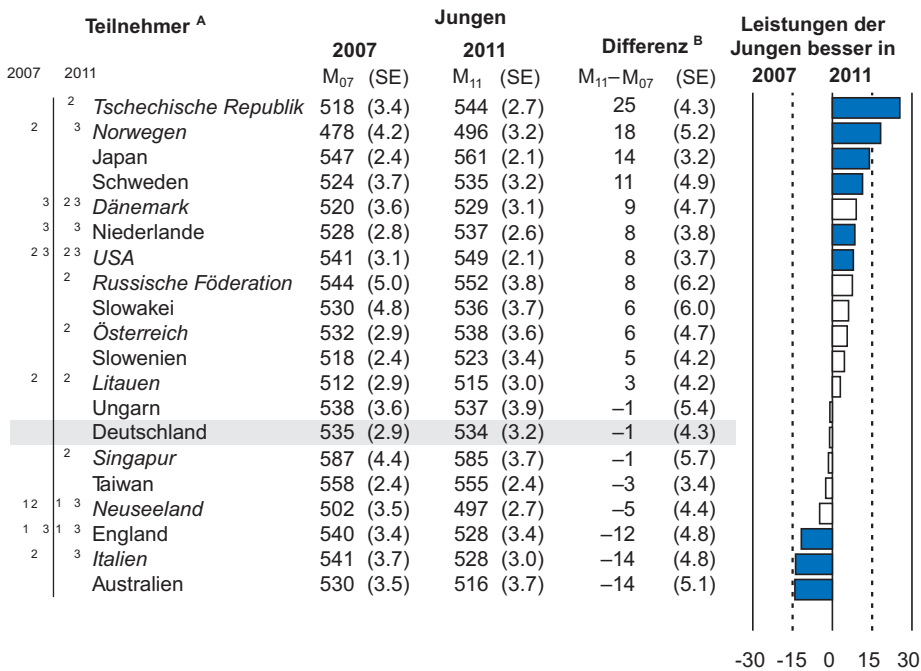
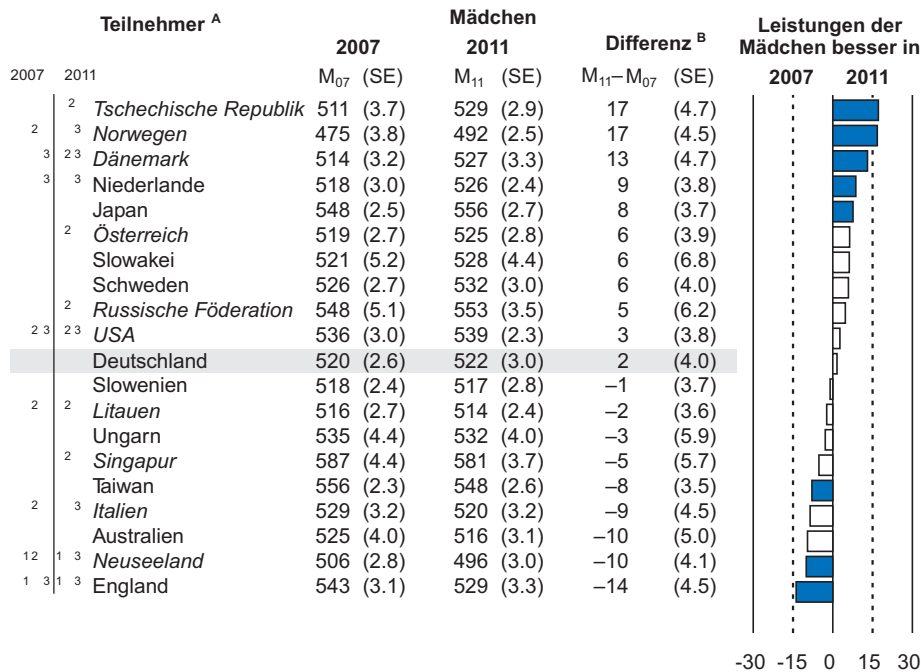
In Deutschland zeigt sich im Anforderungsbereich *Problemlösen* ein Unterschied von 9 Punkten. Für die Vergleichsgruppen EU und OECD ergibt sich jeweils eine nicht signifikante Differenz von einem Punkt zugunsten der Jungen. International wird ein Leistungsvorsprung der Mädchen deutlich. Die Differenz ist mit 6 Punkten signifikant.

Mittlere Leistung auf den Kompetenzstufen. Die mittleren Leistungen sowie die prozentualen Anteile von Jungen und Mädchen auf den Kompetenzstufen der Naturwissenschaften sind in der Abbildung 6.7 analog zur Abbildung 6.3 dargestellt. Die mittleren Leistungen, die Mädchen und Jungen auf den jeweiligen Kompetenzstufen erreichen, unterscheiden sich nicht signifikant voneinander. Vergleicht man allerdings jeweils den Anteil aller Kinder auf den Kompetenzstufen, lässt sich feststellen, dass signifikant mehr Jungen Leistungen auf den Niveaus der Kompetenzstufen IV und V aufweisen.

Vergleich der Kompetenzen in TIMSS 2007 und TIMSS 2011. Im Folgenden sollen noch die Veränderungen der geschlechtsspezifischen Leistungsunterschiede im Vergleich von TIMSS 2007 und TIMSS 2011 untersucht werden. Diese sind in Abbildung 6.8 analog zur Abbildung 6.4 dargestellt.

Im internationalen Vergleich zeigt sich, dass Mädchen im Vergleich von TIMSS 2007 zu TIMSS 2011 in der Tschechischen Republik, Norwegen, Dänemark, den Niederlanden und Japan eine signifikante Veränderung der Leistungsmittelwerte erzielt haben, wobei sie 2011 höhere Leistungen verzeichnen konnten. In Taiwan, Neuseeland und England haben sich hingegen seit 2007 die mittleren Leistungen in den Naturwissenschaften verschlechtert. Für 12 Teilnehmer,

Abbildung 6.8: Vergleich der Leistungsunterschiede in den Naturwissenschaften von Jungen und Mädchen in TIMSS 2007 und TIMSS 2011 (internationaler Vergleich)



■ Statistisch signifikante Unterschiede (p < .05).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

darunter auch Deutschland, ergeben sich keine signifikanten Veränderungen in den Naturwissenschaftsleistungen der Mädchen.

Jungen schneiden im Mittel in der Tschechischen Republik, Norwegen, Japan, Schweden, den Niederlanden und den USA in TIMSS 2011 signifikant besser ab als 2007. Ähnlich wie in 10 weiteren Staaten zeigt sich in Deutschland keine signifikante Veränderung der Leistungsergebnisse der Jungen. In England, Italien und Australien erzielten Jungen 2011 im Mittel eine geringere Leistung als im vorangegangenen Studienzyklus 2007. Des Weiteren kann festgehalten werden, dass es keinem Mitglied der EU oder OECD gelungen ist, die Leistungsunterschiede in den Naturwissenschaften zwischen den Geschlechtern zu verringern.

In der mittleren Naturwissenschaftsleistung haben sich demnach weder für Jungen noch für Mädchen in Deutschland Veränderungen ergeben. Betrachtet man jedoch die Veränderungen zwischen 2011 und 2007 in den Subdomänen, zeigt sich, dass Jungen 2011 signifikant niedrigere Leistungen im Inhaltsbereich *Biologie* und Mädchen signifikant höhere Leistungen im Anforderungsbereich *Anwenden* erzielen. Auch in den Naturwissenschaften zeigt sich wie in Mathematik, dass die Leistungsentwicklungen von Jungen und Mädchen nicht synchron verlaufen. Dabei wäre im nächsten Studienzyklus zu überprüfen, ob sich dies auch im Trend feststellen lässt.

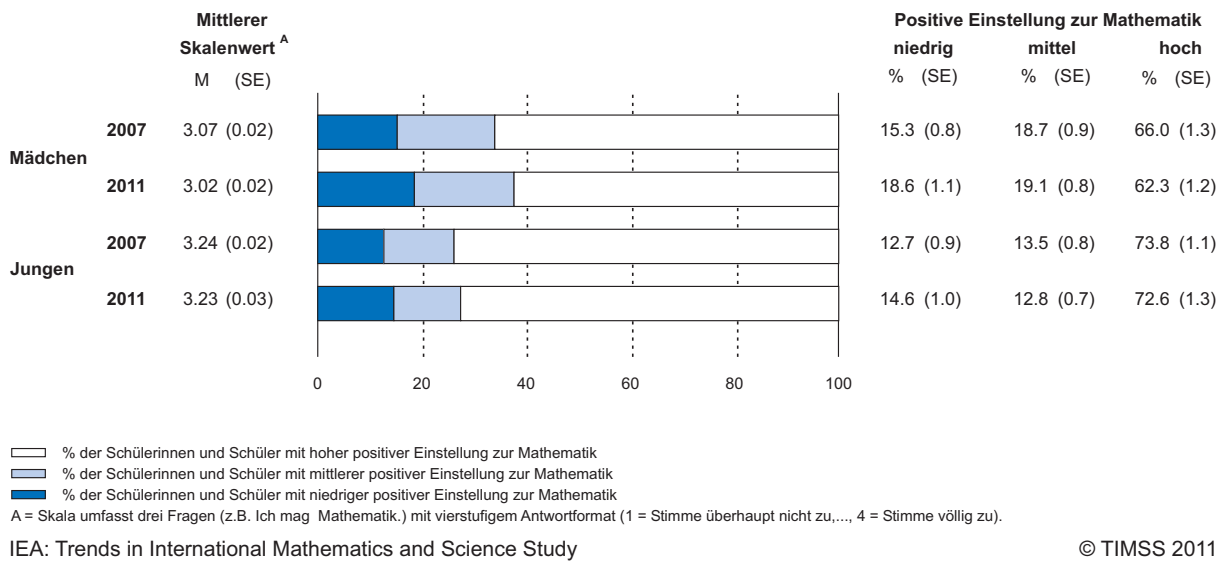
6 Ergebnisse zu geschlechtsspezifischen Unterschieden in Einstellungen und Selbstkonzept

Im Folgenden werden die Ergebnisse zu fachbezogener positiver Einstellung und Selbstkonzept für die deutsche Stichprobe für Jungen und Mädchen dargestellt. Zunächst werden die Ergebnisse zur positiven Einstellung zur Mathematik sowie zum mathematikbezogenen Selbstkonzept berichtet. Anschließend werden Ergebnisse zu Einstellungen zum Fach Sachunterricht und zum sachunterrichtsbezogenen Selbstkonzept beschrieben. Dabei werden jeweils die mittleren Einstellungen und Selbstkonzepte in 2011 und 2007 sowie die Verteilung auf die Kategorien hohe, mittlere, niedrige positive Einstellung sowie hohes, mittleres und niedriges Selbstkonzept dargestellt. Detaillierte Informationen zur Skalenbildung finden sich in den Kapiteln 3 und 4.

In der Abbildung 6.9 sind die mittleren Skalenwerte zur positiven Einstellung zur Mathematik in 2007 und 2011 für Jungen und Mädchen dargestellt. Außerdem werden die Verteilungen der Schülerinnen und Schüler auf die Gruppen von Kindern mit niedrigen, mittleren und hohen positiven Einstellungen dargestellt.

Es zeigt sich, dass Jungen im Mittel signifikant höhere positive Einstellungen als Mädchen haben, wobei die Differenz 2011 0.21 Punkte beträgt. Diese Differenz entspricht einer Effektgröße von Cohens $d = .24$, was als kleiner Effekt bezeichnet werden kann. Dabei kann erwähnt werden, dass sowohl Jungen als auch Mädchen im Durchschnitt eine positive Einstellung zur Mathematik haben.

Vergleicht man die mittleren Skalenwerte der positiven Einstellung zur Mathematik in der Abbildung 6.9, zeigt sich für Mädchen in Deutschland ein leicht, jedoch nicht signifikant geringerer Wert für 2011 (3.02 Punkte) im Vergleich zu 2007 (3.07 Punkte). Auch für die Jungen ergeben sich keine signifikanten Veränderungen. Vergleicht man die drei Gruppen niedrige positive

Abbildung 6.9: Prozentuale Verteilung auf der Skala ‚positive Einstellung zur Mathematik‘ im Vergleich

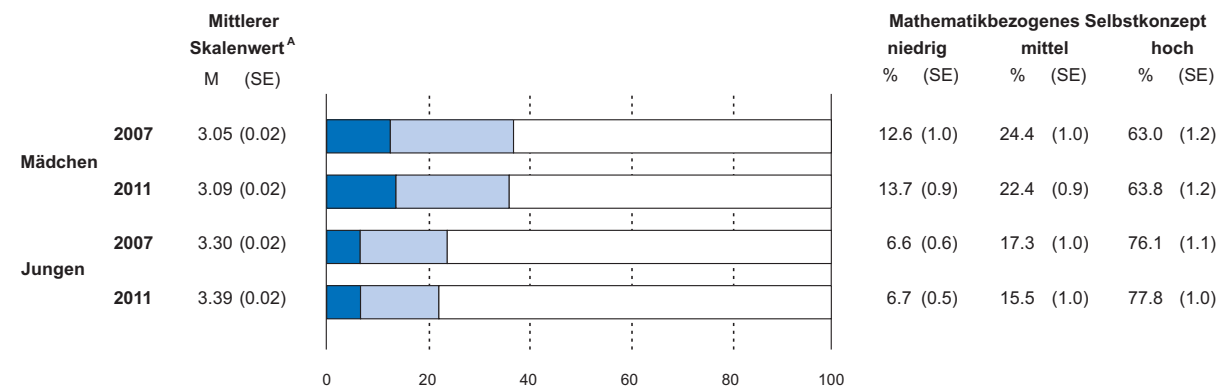
Einstellung, mittlere positive Einstellung und hohe positive Einstellung miteinander, so fällt auf, dass es gegenüber 2007 leichte Verschiebungen für die Mädchen gegeben hat. Der Anteil der Mädchen mit niedriger positiver Einstellung ist um 3.3 Prozentpunkte auf 18.6 Prozent gestiegen. Der Anteil der mittleren Gruppe ist mit 19.1 Prozent für die Mädchen stabil. Der Anteil der Mädchen mit hoher positiver Einstellung fällt hingegen um 3.7 Prozentpunkte ab.

Für die Jungen ergeben sich für die drei Gruppen nur minimale Verschiebungen, wobei auch hier der Anteil mit niedriger positiver Einstellung leicht gestiegen und der Anteil mit hoher positiver Einstellung leicht gesunken ist.

In der Abbildung 6.10, in der die mittleren Skalenwerte sowie die prozentualen Anteile für die Kategorien hohes, mittleres und niedriges Selbstkonzept dargestellt sind, wird ersichtlich, dass in 2011 Jungen ein um 0.3 Punkte höheres Selbstkonzept als Mädchen haben. Dieser Unterschied ist signifikant und es handelt sich nach Cohens Effektgröße mit $d = .40$ um einen mittleren Effekt. Außerdem zeigt sich, dass sich die mittleren Skalenwerte der Mädchen im mathematikbezogenen Selbstkonzept nicht signifikant verändert haben. Das mittlere Selbstkonzept der Jungen ist hingegen um einen Wert von 0.09 Punkten angestiegen. Die Jungen stimmen den Aussagen zum Selbstkonzept also eher zu und artikulieren wie 2007 ein recht positives mathematikbezogenes Selbstkonzept. Vergleicht man die drei Gruppen hohes, mittleres und niedriges Selbstkonzept, so fällt auf, dass es gegenüber 2007 lediglich leichte Verschiebungen gegeben hat, die allesamt aber nicht signifikant sind. Beim Vergleich der Jungen mit den Mädchen wird deutlich, dass der Anteil der Mädchen in der Gruppe der Schülerinnen und Schüler mit niedrigem Selbstkonzept 2011 doppelt so hoch ist wie der der Jungen. So haben 13.7 Prozent der Mädchen ein niedriges Selbstkonzept, doch nur 6.7 Prozent der Jungen. Ein hohes Selbstkonzept geben hingegen 77.8 Prozent der Jungen und nur 63.8 Prozent der Mädchen an.

Insgesamt wird für die Einstellung zur Mathematik und das mathematikbezogene Selbstkonzept ersichtlich, dass Jungen sowohl eine höhere positive Einstellung zur Mathematik haben als auch ein höheres Selbstkonzept verzeichnen können.

Für die Naturwissenschaften wurden im Rahmen von TIMSS die Einstellungen und das Selbstkonzept im Hinblick auf das Fach Sachunterricht er-

Abbildung 6.10: Prozentuale Verteilung auf der Skala ‚mathematikbezogenes Selbstkonzept‘ im Vergleich

□ % der Schülerinnen und Schüler mit hohem mathematikbezogenen Selbstkonzept
 □ % der Schülerinnen und Schüler mit mittlerem mathematikbezogenen Selbstkonzept
 ■ % der Schülerinnen und Schüler mit niedrigem mathematikbezogenen Selbstkonzept

A = Skala umfasst vier Fragen (z.B. Normalerweise bin ich gut in Mathematik.) mit vierstufigem Antwortformat (1 = Stimme überhaupt nicht zu, ..., 4 = Stimme völlig zu).

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

fasst, denn in Deutschland werden naturwissenschaftliche Inhalte in der Grundschule nicht in einem eigenen Fach unterrichtet, so dass nicht davon ausgegangen werden kann, dass die Schülerinnen und Schüler eine klare Vorstellung davon haben, was Naturwissenschaften beziehungsweise naturwissenschaftliche Inhalte sind.

Die Einstellung zum Fach Sachunterricht und das sachunterrichtsbezogene Selbstkonzept werden ebenso wie in Mathematik in TIMSS mit Hilfe eines Schülerfragebogens erfasst, wobei nähere Informationen dem Kapitel 4 entnommen werden können.

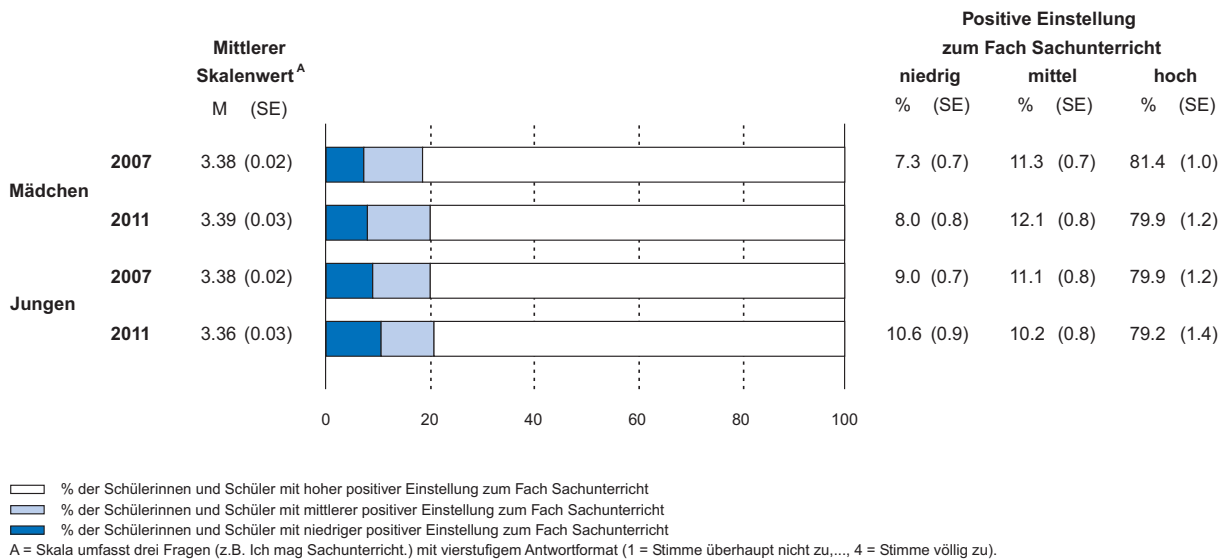
In der Abbildung 6.11 sind sowohl die mittleren Einstellungen zum Sachunterricht von Jungen und Mädchen im Vergleich 2007 zu 2011 dargestellt als auch die prozentualen Verteilungen von Jungen und Mädchen auf die Kategorien niedrige, mittlere und hohe positive Einstellung zum Sachunterricht abgebildet.

Es zeigt sich, dass die Einstellungen der Schülerinnen und Schüler zum Sachunterricht am Ende der Grundschulzeit sehr positiv ausgeprägt sind. Die Mittelwerte von 3.39 Punkten für die Mädchen und 3.36 Punkten für die Jungen liegen recht nah am theoretischen Maximum der Skala (4), wobei dieser Unterschied nicht signifikant ist. Außerdem zeigen sich weder für die Jungen noch für die Mädchen signifikante Veränderungen im Vergleich zwischen TIMSS 2007 und TIMSS 2011.

Dem hohen Mittelwert in 2011 entsprechend groß ist der Anteil der Schülerinnen und Schüler, die eine hohe positive Einstellung aufweisen: Etwa 80 Prozent der Mädchen und 79 Prozent der Jungen erfüllen dieses Kriterium 2011. 12.1 Prozent der Mädchen und 10.2 Prozent der Jungen weisen eine mittlere positive Einstellung auf und nur etwa 8 Prozent der Mädchen und 11 Prozent der Jungen eine niedrige positive Einstellung. Die Anteile der Schülerinnen und Schüler in diesen drei Gruppen unterscheiden sich zwischen TIMSS 2007 und TIMSS 2011 praktisch nicht.

Es zeigt sich, wie schon bei den Einstellungen zum Sachunterricht, dass auch das sachunterrichtsbezogene Selbstkonzept (siehe Abbildung 6.12) der Schülerinnen und Schüler am Ende der Grundschulzeit sehr positiv ausgeprägt ist. In 2011 haben Mädchen im Mittel ein Selbstkonzept von 3.35 Punkten erreicht und Jungen von 3.39 Punkten, wobei dieser Unterschied nicht signifikant

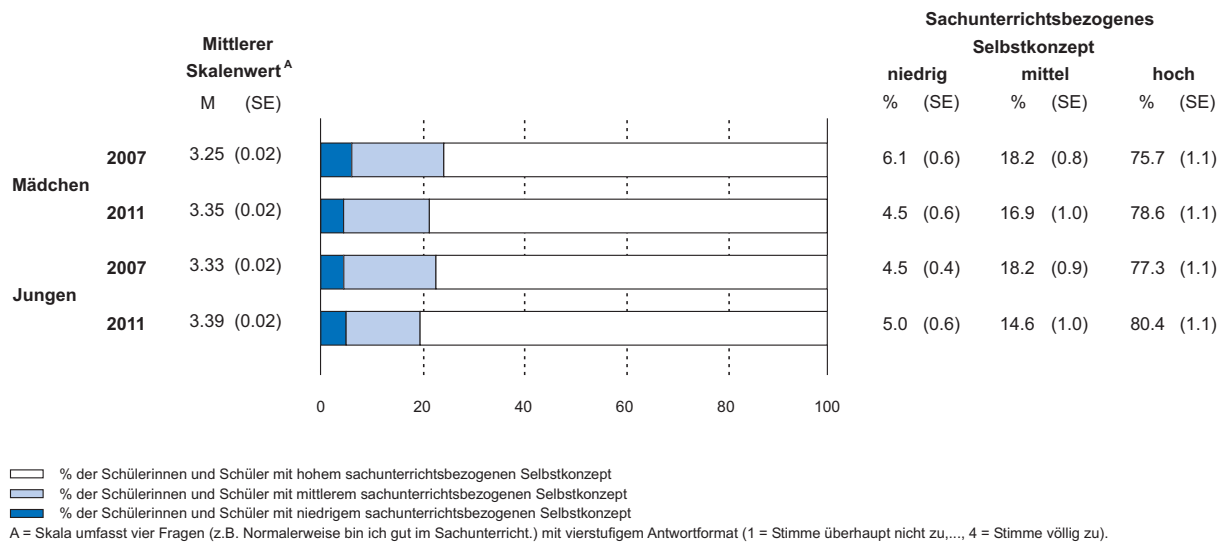
Abbildung 6.11: Prozentuale Verteilung auf der Skala ‚positive Einstellung zu dem Fach Sachunterricht‘ im Vergleich



IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Abbildung 6.12: Prozentuale Verteilung auf der Skala ‚sachunterrichtsbezogenes Selbstkonzept‘ im Vergleich



IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

ist. Im Vergleich zu TIMSS 2007 zeigt sich, dass Mädchen 2011 ein signifikant höheres Selbstkonzept haben (Differenz 0.10 Punkte). Der Mittelwert der Jungen ist mit 3.39 Punkten 2011 im Vergleich zu 2007 signifikant um 0.06 Punkte höher. Dabei können sowohl die Differenz für Jungen als auch für Mädchen als sehr kleine Effekte bezeichnet werden. Dem hohen Mittelwert der Mädchen und Jungen entsprechend groß ist der Anteil der Schülerinnen und Schüler, die ein hohes positives sachunterrichtsbezogenes Selbstkonzept aufweisen: 78.6 Prozent der Mädchen und 80.4 Prozent der Jungen erfüllen dieses Kriterium 2011. Außerdem haben 16.9 Prozent der Mädchen und 14.6 Prozent der Jungen ein mittleres positives Selbstkonzept und nur etwa 5 Prozent der Mädchen und Jungen ein niedriges positives Selbstkonzept. Die Anteile der Jungen in den Gruppen mit hohem und mittlerem positiven Selbstkonzept unterscheiden sich zwischen TIMSS 2007 und TIMSS 2011 signifikant. Dabei sind die prozentualen Anteile in der

Gruppe der Schülerinnen und Schüler bezogen auf das hohe sachunterrichtsbezogene Selbstkonzept gestiegen und in der Gruppe mit mittlerem sachunterrichtsbezogenen Selbstkonzept gesunken.

Insgesamt wird auch für den Sachunterricht deutlich, dass Grundschülerinnen und Grundschüler eine positive Einstellung und ein positives Selbstkonzept haben. Dabei sind die Unterschiede zwischen Jungen und Mädchen nicht so stark ausgeprägt wie in Mathematik.

Im folgenden Kapitelabschnitt wird schließlich noch der Frage nachgegangen, ob Leistungsunterschiede zwischen Mädchen und Jungen in Mathematik und Naturwissenschaften auch dann bestehen, wenn weitere Variablen, die im Zusammenhang mit der Leistungsentwicklung stehen sollten, kontrolliert werden. Dazu werden ausgewählte Prädiktoren von Schulleistungen in einer schrittweisen Regression untersucht. Als abhängige Variable wird dabei die mathematische und naturwissenschaftliche Leistung definiert. Neben dem Geschlecht werden das fachspezifische Selbstkonzept, Noten in Mathematik und im Sachunterricht, das Bildungsniveau der Eltern sowie der sozioökonomische Status als unabhängige Variablen betrachtet. Nicht untersucht wird die positive Einstellung zur Mathematik und zum Sachunterricht, um Multikollinearität mit dem Selbstkonzept zu vermeiden (Bortz & Schuster, 2010).

7 Geschlechtsspezifische Leistungsunterschiede nach Kontrolle weiterer Variablen

Wie oben bereits beschrieben sind die Leistungsunterschiede zwischen Jungen und Mädchen durch eine Vielzahl an Faktoren vermittelt, so dass eine einfache Analyse von einzelnen Merkmalen das „multikausale Beziehungsgeflecht“ (Hellmich & Jahnke-Klein, 2008, S. 114–115) nicht angemessen beschreiben würde. Verschiedene Variablen zeigen jeweils individuell betrachtet Zusammenhänge mit den getesteten mathematischen und naturwissenschaftlichen Kompetenzen, weisen aber auch Zusammenhänge untereinander auf. So können individuell beobachtete Effekte bei umfassenden Analysen in ihren Wirkungen nicht einfach addiert werden. In multivariaten Analysen können hingegen individuelle Effekte zwischen erklärenden Variablen abgeschätzt werden.

Zur Untersuchung des Verhältnisses von Geschlecht, fachspezifischem Selbstkonzept, den jeweiligen Fachnoten, dem Bildungsniveau der Eltern sowie dem sozioökonomischen Status sind in Tabelle 6.1 die Ergebnisse schrittweise erweiterter Regressionsanalysen dargestellt.

Da neben dem Geschlecht auch das fachbezogene Selbstkonzept, die Note sowie der sozioökonomische und kulturelle Status der Familie einen Zusammenhang mit dem Bildungserfolg aufweisen, werden diese in den Analysen berücksichtigt.

Dabei handelt es sich beim Selbstkonzept um die Selbsteinschätzungen der Schülerinnen und Schüler zu ihren jeweiligen Fachleistungen. Die Noten wurden durch die Lehrkräfte in sogenannte ‚Schülerteilnahmelisten‘ eingetragen. Beim höchsten Bildungsabschluss der Eltern werden die Familien in den Blick genommen, in denen mindestens ein Elternteil ein Studium abgeschlossen hat und beim höchsten Berufsstatus werden die Familien betrachtet, in denen mindestens ein Elternteil einen akademischen Beruf oder eine führende Position in einem Unternehmen innehat. Weitere vertiefende Analysen zum sozioökonomischen und kulturellen Status finden sich im Kapitel 7 in diesem Band.

Tabelle 6.1: Regression von Leistungsunterschieden in Mathematik und Naturwissenschaften auf das Geschlecht, das mathematikbezogene bzw. naturwissenschaftsbezogene Selbstkonzept, den höchsten Bildungsabschluss der Eltern und den sozialen Status

	Mathematik		Naturwissenschaft	
	β	(SE)	β	(SE)
Modell I				
Referenzwert	523.5	(2.7) ***	521.7	(3.0) ***
Geschlecht ^A	8.4	(2.7) **	11.9	(2.5) ***
	R ² = .00		R ² = .01	
Modell II				
Referenzwert	398.6	(4.6) ***	373.6	(9.4) ***
Geschlecht	-	- ns	15.5	(2.6) ***
Selbstkonzept ^B	21.5	(3.0) ***	27.9	(2.8) ***
Noten ^C	34.2	(1.4) ***	35.1	(2.2) ***
	R ² = .37		R ² = .26	
Modell III				
Referenzwert	408.1	(4.9) ***	390.2	(9.4) ***
Geschlecht	-	- ns	17.0	(2.5) ***
Selbstkonzept	21.1	(3.0) ***	25.7	(3.7) ***
Noten	30.5	(1.6) ***	30.0	(2.6) ***
Familie mit hohem Bildungsniveau ^D	20.0	(2.7) ***	27.2	(2.9) ***
	R ² = .38		R ² = .28	
Modell IV				
Referenzwert	409.2	(4.9) ***	396.8	(9.8) ***
Geschlecht	-	- ns	16.8	(2.4) ***
Selbstkonzept	20.8	(3.1) ***	26.2	(3.6) ***
Noten	30.3	(1.5) ***	28.2	(2.6) ***
Familie mit hohem Bildungsniveau	12.6	(2.9) ***	15.6	(3.7) ***
Familie mit hohem sozioökonomischen Status ^E	10.0	(2.6) ***	16.4	(3.1) ***
	R ² = .38		R ² = .28	

β = Regressionsgewichte (unstandardisiert)

Signifikanzniveau: ns= nicht signifikant; *= signifikant ($p < .05$); **= signifikant ($p < .01$); ***= signifikant ($p < .001$)

A= Geschlecht (0= Mädchen; 1= Junge)

B= mathematikbezogenes bzw. sachunterrichtsbezogenes Selbstkonzept (0= niedriges und mittleres Selbstkonzept; 1= hohes Selbstkonzept)

C= Mathematik- bzw. Sachunterrichtnote (1= ungenügend/mangelhaft; 2= ausreichend; 3= befriedigend; 4= gut; 5= sehr gut)

D= Bildungsniveau nach höchstem Bildungsabschluss der Eltern (0= kein Elternteil mit (Fach-)Hochschulabschluss; 1= mindestens ein Elternteil mit (Fach-)Hochschulabschluss)

E= Sozialschicht nach Angabe der Eltern (0= Manuell Tätige, Angestellte und kleinere Unternehmer; 1= Akademiker, Techniker und Führungskräfte)

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Tabelle 6.1 ist zu entnehmen, dass sich in dem sukzessiven Aufbau der Modelle zwischen Mathematik und Naturwissenschaften Unterschiede ergeben, die jedoch nicht im Widerspruch zueinander stehen. Vielmehr werden Spezifika der verschiedenen Erhebungen und Fächer deutlich.

Im *Modell I* zeigt sich, dass Jungen in Bezug auf ihre mathematischen (+8 Punkte) und naturwissenschaftlichen (+12 Punkte) Kompetenzen signifikant bessere Leistungsergebnisse erzielen als ihre Mitschülerinnen. Dieses Ergebnis wurde bereits in den Abbildungen 6.1 und 6.5 illustriert.

Werden das mathematikbezogene beziehungsweise sachunterrichtsbezogene Selbstkonzept sowie die jeweiligen Fachnoten zur Vorhersage der Leistungen in die Modellrechnung (*Modell II*) einbezogen, zeigt sich, dass sich der beschriebene Geschlechtereffekt für Mathematik nicht mehr signifikant nachweisen lässt,

jedoch für die Naturwissenschaften erhalten bleibt und sogar um 3.6 Punkte ansteigt. Für beide Domänen erweist sich zudem das Selbstkonzept als signifikanter Leistungsprädiktor. Dies bedeutet, dass Kinder mit hohem Selbstkonzept in Mathematik 22 Punkte und in den Naturwissenschaften 28 Punkte Vorsprung vor Kindern mit niedrigem und mittlerem Selbstkonzept haben. Ein höherer Prädiktor als das fachspezifische Selbstkonzept scheinen die Fachnoten zu sein, wobei dies nicht verwunderlich ist, da diese zu einem großen Teil aus Vorleistungen bestehen. Mit *Modell II* lassen sich außerdem in Mathematik 37 Prozent und in den Naturwissenschaften 26 Prozent der Leistungsvarianz zwischen Kindern erklären.

Bei zusätzlicher Berücksichtigung des Bildungsniveaus der Eltern (*Modell III*) zeigt sich, dass sich für Mathematik weiterhin kein Geschlechtereffekt signifikant nachweisen lässt, dieser jedoch für die Naturwissenschaften um weitere 1.5 Punkte ansteigt. Jungen weisen damit in ihren naturwissenschaftlichen Kompetenzen um 17 Punkte signifikant höhere Leistungsergebnisse auf als ihre Mitschülerinnen. Die Varianzaufklärung steigt marginal in Mathematik auf 38 Prozent und in den Naturwissenschaften auf 28 Prozent an.

Unter Hinzunahme der Information der Zugehörigkeit zur Gruppe der Familien mit hohem sozioökonomischen Status (*Modell IV*) verändern sich die bereits im Modell III betrachteten Werte für den Zusammenhang zwischen Geschlecht und mathematischen wie naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern nur geringfügig.

8 Zusammenfassung

In diesem Kapitel wurden zum einen geschlechtsspezifische Leistungsunterschiede in Mathematik und den Naturwissenschaften sowie positive mathematikbezogene und sachunterrichtsbezogene Einstellungen und Selbstkonzepte von Jungen und Mädchen dargestellt. Außerdem wurden geschlechtsspezifische Leistungsunterschiede unter Kontrolle weiterer Variablen untersucht.

Leistungsunterschiede zwischen Jungen und Mädchen. Wie bereits in TIMSS 2007 lassen sich auch in TIMSS 2011 signifikante geschlechtsspezifische Unterschiede in den mathematischen und naturwissenschaftlichen Leistungen von Viertklässlerinnen und Viertklässlern in Deutschland feststellen. Es zeigt sich, dass Jungen in Deutschland sowohl in Mathematik als auch in den Naturwissenschaften signifikant höhere Leistungen erzielen als Mädchen, wobei die Unterschiede vor allem in den Naturwissenschaften vergleichsweise hoch ausfallen. Im Vergleich mit den teilnehmenden Mitgliedsstaaten der EU und OECD zeigt sich, dass lediglich in der Tschechischen Republik die geschlechtsspezifischen Unterschiede in den Naturwissenschaften um drei Leistungspunkte höher ausfallen als in Deutschland (12). Im europäischen Vergleich zeigt sich, dass für Teilnehmerstaaten wie England, Irland, Litauen, Nordirland, Portugal, Rumänien, Schweden und Ungarn sowohl für die Gesamtskala Mathematik als auch für die Gesamtskala Naturwissenschaften keine signifikanten geschlechtsspezifischen Leistungsunterschiede festzustellen sind. Für Dänemark, Finnland, Norwegen und Slowenien zeigen sich signifikante Leistungsunterschiede, die zugunsten der Jungen ausfallen, lediglich im Bereich Mathematik.

Relative Leistungsstärken zeigen die Jungen in Deutschland in den mathematischen Inhaltsbereichen *Arithmetik* und *Geometrie/Messen*. Auch in den mathematischen Anforderungsbereichen *Reproduzieren* und *Problemlösen* erreichen Jungen in Deutschland höhere Leistungswerte als Mädchen. Für die naturwissenschaftlichen Inhaltsbereiche zeigen sich relative Leistungsstärken der Jungen

in *Physik/Chemie* und *Geographie*. In den Naturwissenschaften erzielen Jungen in Deutschland in allen kognitiven Anforderungsbereichen signifikant bessere Leistungen als Mädchen.

Im Vergleich der beiden Zyklen TIMSS 2007 und TIMSS 2011 scheint sich auf den ersten Blick weder für die Mädchen noch für die Jungen in Deutschland eine Veränderung ergeben zu haben. Bei genauerer Betrachtung der Inhalts- und Anforderungsbereiche zeigt sich jedoch, dass für Jungen im Bereich *Geometrie/Messen* und für Mädchen im Bereich *Umgang mit Daten* Leistungsverbesserungen gefunden werden können. Allerdings haben die Jungen im Inhaltsbereich *Biologie* im Vergleich signifikant niedrigere Leistungen erzielt als 2007. In den Anforderungsbereichen zeigen sich sowohl für die Jungen als auch für die Mädchen in Mathematik im Anforderungsbereich *Reproduzieren* signifikant höhere Kompetenzen. Außerdem können die Mädchen in den Naturwissenschaften im Bereich *Anwenden* Leistungssteigerungen verzeichnen. Im internationalen Vergleich zeigt sich, dass sich in einigen Teilnehmerstaaten zwar signifikante Veränderungen der mittleren Leistungen von Jungen und Mädchen feststellen lassen, keinem Mitgliedsstaat der EU oder OECD es aber gelungen ist, im Vergleich zu TIMSS 2007 Leistungsunterschiede zwischen den Geschlechtern signifikant zu verringern.

Geschlechtsspezifische Unterschiede in Einstellung und Selbstkonzept. Wie bereits 2007 zeigen die Grundschülerinnen und Grundschüler in Deutschland sowohl zur Mathematik als auch zum Fach Sachunterricht ausgesprochen positive Einstellungen. Für das Fach Mathematik wird deutlich, dass Jungen sowohl eine höhere positive Einstellung zur Mathematik haben als auch ein höheres mathematikbezogenes Selbstkonzept aufweisen. Die geschlechtsspezifischen Unterschiede in den positiven Einstellungen zur Mathematik sind aber als eher gering zu bewerten. Allerdings zeigt sich im Vergleich zu 2007, dass der Anteil an Mädchen mit hoher positiver Einstellung um knapp 4 Prozentpunkte gesunken ist. In Bezug auf das mathematische Selbstkonzept lassen sich keine signifikanten Veränderungen zu TIMSS 2007 beobachten. Weiterhin zeigt sich, dass der Anteil an Mädchen in der Gruppe der Schülerinnen und Schüler mit einem niedrigen Selbstkonzept auch 2011 doppelt so hoch ist wie der der Jungen.

Unterschiede in den Einstellungen zum Sachunterricht und dem sachunterrichtsbezogenen Selbstkonzept sind hingegen nicht signifikant. Etwa 80 Prozent der Mädchen und 79 Prozent der Jungen berichten von hohen positiven Einstellungen zum Fach Sachunterricht. Ebenso groß sind die Anteile von Jungen und Mädchen mit hohen positiven sachunterrichtsbezogenen Selbstkonzepten.

Geschlechtsspezifische Leistungsunterschiede nach Kontrolle weiterer Variablen. Insgesamt lässt sich feststellen, dass sowohl Leistungsunterschiede als auch Differenzen in Einstellungen und Selbstkonzepten zwischen Jungen und Mädchen in der Grundschule bestehen. Multivariate Analysen unter Einbezug der Variablen Geschlecht, Selbstkonzept, Note, Bildungsniveau der Eltern sowie sozioökonomischer Status zeigen in Mathematik, dass nach den Noten insbesondere das Selbstkonzept einen entscheidenden Anteil der Leistungsdifferenzen zwischen Jungen und Mädchen erklärt. Für die Naturwissenschaften wird hingegen deutlich, dass Geschlechterunterschiede sich auch unter Hinzunahme des Selbstkonzepts und der Noten als signifikant erweisen und sogar um etwa 4 Punkte ansteigen. Unter Hinzunahme der Information zur Zugehörigkeit zu der Gruppe mit hohem Bildungsniveau der Eltern sowie der Information zum familiären Berufsstatus verändert sich der Zusammenhang zwischen Geschlecht, Selbstkonzept und mathematischen wie naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern nur geringfügig. Diese Befunde verdeutlichen,

dass hier Wechselwirkungen zwischen den fachspezifischen Leistungen und Einstellungen und Selbstkonzepten bestehen, die sich zugunsten der Jungen auswirken. Dieser Befund verweist darauf, dass Programme, die auf einen optimierten Kompetenzerwerb von Mädchen wie Jungen abzielen, auch die fachspezifischen Einstellungen und Selbstkonzepte der Schülerinnen und Schüler berücksichtigen sollten, da die unterschiedliche Entwicklung von Selbstkonzepten von Jungen und Mädchen (s. dazu König, Wagner & Valtin, 2011) insbesondere im weiteren Verlauf der Bildungskarriere an Bedeutung gewinnt und beispielsweise zu unterschiedlichem Kurswahlverhalten in der Oberstufe führen kann (Köller, Daniels, Schnabel & Baumert, 2000). Dabei könnten beispielsweise Angebote, die darauf abzielen, Mädchen in ihren Selbstkonzepten zu stärken, helfen, Geschlechterdisparitäten entgegenzuwirken. Bei Angeboten wie dem Attributionstraining (Ziegler & Finsterwald, 2008) könnten sowohl Lehrkräfte lernen, wie sie günstige Ursachenerklärungen und funktionale Wirklichkeitskonstruktionen aufbauen können, als auch Schülerinnen und Schüler günstige Ursachenzuschreibungen ihrer eigenen Leistungen lernen.

Literatur

- Autorengruppe Bildungsberichterstattung. (2006). *Bildung in Deutschland 2006: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Zugriff am 11.10.2012 unter <http://www.bildungsbericht.de/daten/gesamtbericht.pdf>.
- 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.
- Bonsen, M., Lintorf, K. & Bos, W. (2008). Kompetenzen von Jungen und Mädchen. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 125–140). Münster: Waxmann.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7., vollständig überarbeitete und erweiterte Auflage). Berlin: Springer.
- Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.). (2008). *TIMSS 2007: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K. & Walther, G. (Hrsg.). (2004). *IGLU: Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*. Münster: Waxmann.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (Soziale Welt, Sonderband 2, S. 183–198). Göttingen: Schwartz.
- Deci, E. L. & Ryan, R. M. (2000). The „what“ and „why“ of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11 (4), 227–268.
- Eccles, J. S. (1983). Expectancies, values and academic choice: Origins and changes. In J. Spence (Hrsg.), *Achievement and achievement motivation*. San Francisco: W. H. Freeman.
- Eccles, J. S. (2006). Families, schools, and developing achievement-related motivations and engagement. In J. E. Grusec & P. D. Hastings (Hrsg.), *Handbook of socialization. Theory and research* (S. 665–691). New York: Guilford.
- Eccles, J. S. (2007). Subjective task value and the Eccles et al. model of achievement-related choices. In R. Emmerich & C. S. Dweck (Hrsg.), *Handbook of competence and motivation* (S. 105–121). New York: The Guilford Press.
- EcKes, T. (2008). Geschlechterstereotype: Von Rollen, Identitäten und Vorurteilen. In R. Becker & B. Kortendiek (Hrsg.), *Handbuch Frauen & Geschlechterforschung. Theorie, Methoden, Empirie* (2. Aufl., S. 171–182). Wiesbaden: Verlag für Sozialwissenschaften.

- Ehmke, T. & Siegle, T. (2008). Einfluss elterlicher Mathematikkompetenz und familialer Prozesse auf den Kompetenzerwerb von Kindern in Mathematik. *Psychologie in Erziehung und Unterricht*, 55 (4), 253–264.
- Frey, A., Heinze, A., Mildner, D., Hochweber, J. & Asseburg, R. (2010). Mathematische Kompetenz von PISA 2003 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. 153–176). Münster: Waxmann.
- Geary, D. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and brain sciences*, 19 (2), 229–284.
- Geissler, G. (2011). *Schulgeschichte in Deutschland: Von den Anfängen bis in die Gegenwart*. Frankfurt a. M.: Peter Lang.
- Halpern, D. F. (2004). A cognitive-process taxonomy for sex differences in cognitive abilities. *Current Directions in Psychological Science*, 13 (4), 135–139.
- Halpern, D. F. (2012). *Sex differences in cognitive abilities* (4. Aufl.). New York: Psychology Press.
- Hannover, B. (2010). Sozialpsychologie. In G. Steins (Hrsg.), *Handbuch Psychologie und Geschlechterforschung* (S. 27–42). Wiesbaden: Verlag für Sozialwissenschaften.
- Hansford, B. C. & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52 (1), 123–142.
- Heckhausen, J. & Heckhausen, H. (2010). Motivation und Entwicklung. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 427–486). Berlin: Springer.
- Hellmich, F. & Günther, H. (2011). Entwicklung von Selbstkonzepten bei Kindern im Grundschulalter – ein Überblick. In F. Hellmich (Hrsg.), *Selbstkonzepte im Grundschulalter. Modelle, empirische Ergebnisse, pädagogische Konsequenzen* (S. 17–46). Stuttgart: Kohlhammer.
- Hellmich, F. & Jahnke-Klein, S. (2008). Selbstbezogene Kognitionen und Interessen von Mädchen und Jungen im Mathematikunterricht der Grundschule. In B. Rendtorff & A. Prengel (Hrsg.), *Kinder und ihr Geschlecht* (S. 111–120). Opladen: Budrich.
- Hornberg, S., Valtin, R., Potthoff, B., Schwippert, K. & Schulz-Zander, R. (2007). Lesekompetenzen von Mädchen und Jungen im internationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195–223). Münster: Waxmann.
- Kammermeyer, G. & Martschinke, S. (2003). Schulleistung und Fähigkeitsbild im Anfangsunterricht – Ergebnisse aus dem KILIA-Projekt. *Empirische Pädagogik*, 17 (4), 486–503.
- Köller, O., Daniels, Z., Schnabel, U. K. & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse. *Zeitschrift für Pädagogische Psychologie*, 14 (1), 26–37.
- König, J., Wagner, C. & Valtin, R. (2011). *Jugend – Schule – Zukunft. Psychosoziale Bedingungen der Persönlichkeitsentwicklung. Ergebnisse der Längsschnittstudie AIDA*. Münster: Waxmann.
- Linn, M. C. & Petersen, A. C. (1985). Emergence and characterisation of gender differences in spatial abilities: A meta-analysis. *Child Development*, 56 (6), 1479–1498.
- Lintorf, K. (2012). *Wie vorhersagbar sind Grundschulnoten? Prädiktionskraft individueller und kontextspezifischer Merkmale*. Wiesbaden: Verlag für Sozialwissenschaften.
- Marsh, H. W. & Craven, R. G. (2006). Reciprocal effects of self-concept and achievement: Competing multidimensional and unidimensional perspectives. *Perspectives on Psychological Science*, 1 (2), 133–163.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, 76 (2), 397–416.
- Martin, M., Mullis, I., Foy, P. & Stanco, G. (2012). *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Möller, J. & Trautwein, U. (2009). Selbstkonzept. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 180–203). Heidelberg: Springer.
- Moschner, B., Dickhäuser, O. & Rost, D. H. (2006). *Selbstkonzept* (3. Aufl.). Weinheim: Beltz.
- Mullis, I. V. S., Martin, M. O., Foy, P. & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

- 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. 23–72). Münster: Waxmann.
- Nyborg, H. (1983). Spatial ability in men and women: Review and new theory. *Advances in Behavior Research and Therapy*, 5 (2), 89–140.
- Prenzel, M., Geiser, H., Langeheine, R. & Lobemeier, K. (2003). Das naturwissenschaftliche Verständnis am Ende der Grundschule. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 143–187). Münster: Waxmann.
- Prenzel, M., Geiser, H., Langeheine, R. & Lobemeier, K. (2004). Naturwissenschaftliche Kompetenz am Ende der Grundschulzeit: Vergleiche zwischen einigen Ländern der Bundesrepublik Deutschland. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hrsg.), *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (S. 93–115). Münster: Waxmann.
- Rönnebeck, S., Schöps, K., Prenzel, M., Mildner, D. & Hochweber, J. (2010). Naturwissenschaftliche Kompetenz von PISA 2006 bis 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. 177–198). Münster: Waxmann.
- Schiefele, U. (2009). Motivation. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 151–178). Heidelberg: Springer.
- Schöps, K., Walter, O., Zimmer, K. & Prenzel, M. (2006). Disparitäten zwischen Jungen und Mädchen in der mathematischen Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. H. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 209–224). Münster: Waxmann.
- Stanat, P. & Bergmann, S. (2009). Geschlechtsbezogene Disparitäten in der Bildung. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 513–528). Wiesbaden: Verlag für Sozialwissenschaften.
- Statistisches Bundesamt (Hrsg.). (2012). *Statistisches Jahrbuch für die Bundesrepublik Deutschland 2012: Für die Bundesrepublik Deutschland mit „Internationalen Übersichten“*. Wiesbaden: Statistisches Bundesamt.
- Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotation, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599–604.
- Walther, G., Geiser, H., Langeheine, R. & Lobemeier, K. (2003). Mathematische Kompetenzen am Ende der vierten Jahrgangsstufe. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 189–226). Münster: Waxmann.
- Walther, G., Geiser, H., Langeheine, R. & Lobemeier, K. (2004). Mathematische Kompetenzen am Ende der vierten Jahrgangsstufe in einigen Ländern der Bundesrepublik Deutschland. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hrsg.), *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (S. 117–140). Münster: Waxmann.
- Walther, G., Selter, C., Bensen, M. & Bos, W. (2008). Mathematische Kompetenz im internationalen Vergleich. Testkonzeption und Ergebnisse. In W. Bos, M. Bensen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 49–85). Münster: Waxmann.
- Wigfield, A. & Eccles, J. S. (2002). The development of competence beliefs, expectancies of success and achievement through adolescence. In A. Wigfield & J. S. Eccles (Hrsg.), *Development of Achievement Motivation*. San Diego: Academic Press.
- Wittwer, J., Saß, S. & Prenzel, M. (2008). Naturwissenschaftliche Kompetenz im internationalen Vergleich: Testkonzeption und Ergebnisse. In W. Bos, M. Bensen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 87–124). Münster: Waxmann.
- Ziegler, A. & Finsterwald, M. (2008). Attributionstrainings. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (Bd. 10, S. 416–427). Göttingen: Hogrefe.



Kapitel VII

Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften

Tobias C. Stubbe, Irmela Tarelli und Heike Wendt

1 Einleitung

Ein stabiler Befund der empirischen Bildungsforschung ist die für eine Vielzahl von Staaten nachgewiesene enge Kopplung zwischen dem sozioökonomischen Status von Familien und dem Bildungserfolg ihrer Kinder. Ungleiche Startbedingungen (primäre Herkunftseffekte) verstärken sich dabei im Zeitverlauf meist noch durch herkunftsbedingte Unterschiede in den familiären und institutionellen Bildungsverlaufsentscheidungen (sekundäre Herkunftseffekte; vgl. Stubbe, Bos & Euen, 2012).

In Folge der Bildungsexpansion stand dieses Thema in der Bundesrepublik Deutschland lange Zeit nicht mehr im Fokus des öffentlichen Interesses. Dies hat sich erst mit der Teilnahme an den großen Schulleistungstudien geändert. In dieser Perspektive lassen sich für die Bundesrepublik Deutschland im internationalen Vergleich regelmäßig sehr große soziale Disparitäten nachweisen. Im Zuge des *Programme for International Student Assessment* (PISA) 2000 schnitt Deutschland sogar von allen Teilnehmerstaaten in diesem Aspekt am ungünstigsten ab (Baumert & Schümer, 2001). Entsprechend konstatierten zahlreiche Autorinnen und Autoren, dass die Bildungsexpansion insgesamt nicht zu einem Abbau an sozialen Disparitäten beigetragen hat (vgl. z.B. Becker & Lauterbach, 2004; Blossfeld, 1993; Schimpl-Neimanns, 2000; Solga, 2005; Vester, 2004, 2005).

In den weiteren PISA-Zyklen fiel der Zusammenhang zwischen Herkunft und erreichten Kompetenzen in Deutschland zwar etwas geringer aus, aber weiterhin gibt es nur wenige Staaten, in denen das Ausmaß der sozialen Disparitäten in der Sekundarstufe höher ausfällt (vgl. zum Überblick Ehmke & Jude, 2010). Auch im Rahmen der bisherigen Publikationen zur *Internationalen Grundschul-Lese-Untersuchung* (IGLU) und der *Trends in International Mathematics and Science Study* (TIMSS) wurde für Deutschland im internationalen Vergleich eine

eher überdurchschnittliche Kopplung zwischen dem sozialen Status und dem Bildungserfolg der Schülerinnen und Schüler berichtet (Bonsen, Frey & Bos, 2008; Bos, Schwippert & Stubbe, 2007).

Im Vergleich der Länder der Bundesrepublik Deutschland im Rahmen von IGLU-E 2006 wurden zum Teil deutliche Unterschiede festgestellt. In Berlin und Hamburg waren die sozialen Disparitäten statistisch signifikant höher als im Bundesdurchschnitt und in Bayern signifikant niedriger (Stubbe, Bos & Hornberg, 2008). Im Rahmen des vom *Institut zur Qualitätsentwicklung im Bildungswesen* (IQB) durchgeführten Ländervergleichs 2011 berichten Richter, Kuhl und Pant (2012) im Bereich Mathematik für Deutschland insgesamt einen sozialen Gradienten von 35 Punkten. In diesem Kompetenzbereich weisen Brandenburg und Hessen die nominell niedrigsten Werte (28 bis 31 Punkte) auf, während für die drei Stadtstaaten und Baden-Württemberg soziale Gradienten (Steigerung der Regressionsgeraden auf Individualebene) von 37 bis 43 Punkten berichtet werden. Signifikant unterscheidet sich vom deutschen Mittelwert allerdings nur Bremen.

Ein zentraler Aspekt bei der Analyse sozialer Disparitäten ist die Frage, wie der sozioökonomische Status von Familien valide gemessen werden kann. Theoretisch lassen sich die Erhebungsinstrumente, die zurzeit in der empirischen Bildungsforschung zur Operationalisierung des sozialen Status eingesetzt werden, sehr gut auf der Grundlage der Arbeiten von Pierre Bourdieu zur Unterscheidung von ökonomischem, kulturellem und sozialem Kapital (Bourdieu, 1983) ordnen. Als ökonomisches Kapital beschreibt Bourdieu alles, was direkt in Geld umwandelbar ist. Beim kulturellen Kapital unterscheidet Bourdieu (1) das erworbene Wissen einer Person (inkorporiertes Kulturkapital), (2) Besitztümer wie Bücher oder Musikinstrumente, die nur von Personen adäquat genutzt werden können, die über das notwendige inkorporierte Kulturkapital verfügen (objektiviertes Kulturkapital) und (3) offizielle schulische und akademische Titel (institutionalisiertes Kulturkapital). Soziales Kapital besteht nach Bourdieu aus den Beziehungsnetzwerken der Menschen. Dabei hängt das Ausmaß des sozialen Kapitals, über das eine Person verfügt, zum einen von der Größe dieses Netzes ab und zum anderen von dem Kapital, das die übrigen Personen im Netzwerk besitzen. Coleman (1988) grenzt von dem sozialen Kapital außerhalb der Familie das soziale Kapital innerhalb der Familie ab und betont dessen Wichtigkeit für den Erwerb von schulischen Kompetenzen.

Schließlich sind für die Analyse sozialer Disparitäten noch die Transformationen einer Kapitalform in eine andere von großer Bedeutung. Grundsätzlich kann jede Kapitalform in jede andere transformiert werden, auch wenn damit unter Umständen Transformationskosten (z.B. zu investierende Zeit) verbunden sind. Beispielsweise kann eine Familie ihre Kinder beim Erwerb von kulturellem Kapital (also Bildung) unterstützen, indem vorhandenes ökonomisches (z.B. Investition in Nachhilfeunterricht), kulturelles (z.B. Unterstützung durch die Eltern bei der Arbeit für die Schule) und soziales (z.B. Unterstützung durch befreundete Lehrkräfte) Kapital genutzt wird.

Verschiedene Indikatoren, die sich in den vergangenen Jahren in der empirischen Bildungsforschung bewährt haben, sollen im Folgenden kurz vorgestellt werden. Die Frage nach der Anzahl der im Haushalt vorhandenen *Bücher* kann mithin als „Klassiker“ unter den Instrumenten zur Erfassung von sozialem Status in der internationalen Schulleistungsforschung bezeichnet werden. Die Frage weist sowohl für Schüler- als auch Elternbefragungen gute Messeigenschaften auf und ist zudem einfach und ökonomisch in ihrer Administration. Die Variable

erfasst primär das objektivierte Kulturkapital, liefert damit aber auch einen Hinweis auf das ökonomische Kapital. Indirekt misst die Variable zusätzlich das inkorporierte (und institutionalisierte) Kulturkapital, da ein Zusammenhang besteht zwischen dem Bildungsniveau einer Familie und der Anzahl der Bücher, die diese besitzt (Bildungsnähe des Elternhauses). Auch Angaben zu weiteren *Besitztümern im Haushalt* liefern Informationen zum ökonomischen beziehungsweise kulturellen Kapital von Familien (z.B. Anzahl Autos, Musikinstrumente, Tageszeitung).

Beim *Bildungsniveau der Eltern* handelt es sich um einen direkten Indikator für das kulturelle Kapital in der Schülerfamilie. Zusätzlich bietet diese Variable aber auch einen Hinweis auf das ökonomische Kapital im Haushalt, da ein höherer Bildungsabschluss oftmals auch mit einem höheren Einkommen einhergeht. Im Rahmen von IGLU/TIMSS 2011 wurde – wie auch schon in den vorangegangenen Erhebungen – der höchste Bildungsabschluss von Mutter und Vater mit Hilfe der internationalen Bildungsskala der *United Nations Educational, Scientific and Cultural Organization* (UNESCO) (*International Standard Classification of Education* – ISCED) erhoben (Schroedter, Lechert & Lüttinger, 2006; UNESCO, 2003).

Der *Berufsstatus von Personen* liefert indirekt Informationen über das kulturelle (Bildungsniveau, das für bestimmte Berufe notwendig ist) und ökonomische (Einkommen in den einzelnen Berufen) Kapital. Bedingt durch die sehr hohe Anzahl von unterschiedlichen Berufen muss dieser zunächst offen erfragt werden. Anschließend wird der Berufsstatus mit Hilfe der Nominalskala *International Standard Classification of Occupations* (ISCO) kodiert. Dabei wird jedem Beruf ein vierstelliger Code zugewiesen, der allerdings noch keine Aussagen über den Status eines Berufes erlaubt (Hoffmann, 2003). Diese Codes lassen sich dann in die verschiedenen Indizes des Berufsstatus überführen. Alle zwanzig Jahre werden die ISCO-Codes aktualisiert (zuletzt 2008). Daher wurde im Rahmen der bisherigen IGLU- und TIMSS-Erhebungszyklen ISCO-88 genutzt und im aktuellen Erhebungszyklus ISCO-08. In der empirischen Bildungsforschung werden vor allem zwei Indizes des Berufsstatus genutzt, die auf unterschiedliche theoretische Annahmen zurückgreifen und unterschiedliche statistische Eigenschaften haben:

- Der *International Socio-Economic Index of Occupational Status* (ISEI) gibt an, wie gut ein Beruf geeignet ist, um die Ausbildung einer Person in Einkommen umzuwandeln. Die Skala ist eindimensional und nimmt ganzzahlige Werte zwischen 16 und 90 an (Ganzeboom, de Graaf & Treiman, 1992; Ganzeboom & Treiman, 1996). In der empirischen Bildungsforschung wird häufig der höchste ISEI im Haushalt (HISEI) genutzt.
- Erikson, Goldthorpe und Portocarero (1979) teilen Berufe in distinkte Gruppen ein, die sich im Hinblick auf unterschiedliche Dimensionen durch hohe interne Homogenität und hohe externe Heterogenität auszeichnen. Bei diesen sogenannten EGP-Klassen handelt es sich also um eine Nominalskala.

Schließlich liefert das *Einkommen* von Haushalten direkte Angaben über das ökonomische Kapital. Eine weitere Variable, die aus dem Haushaltseinkommen abgeleitet werden kann, ist die Einkommensarmut von Familien (vgl. Abschnitt 2). Im Rahmen von IGLU/TIMSS 2011 wurden in Deutschland national alle in diesem Abschnitt vorgestellten Indikatoren eingesetzt. International kamen allerdings deutlich weniger Variablen für die Ermittlung des sozioökonomischen Status von Familien zum Einsatz. Aus diesem Grund werden in Abschnitt 3 zunächst internationale Vergleiche unter Berücksichtigung der Anzahl der Bücher im Haushalt

berichtet, bevor dann im vierten Abschnitt vertiefend soziale Disparitäten im deutschen Bildungssystem dargestellt werden.

2 Einkommensarmut

Spätestens mit den Armuts- und Reichtumsberichten der Bundesregierung (Bundesministerium für Arbeit und Soziales, 2001, 2005, 2008)¹ ist das Thema der Einkommensarmut auch in der Bundesrepublik Deutschland in den Fokus der Öffentlichkeit gerückt. Insbesondere die Frage der Kinderarmut ist dabei von besonderer Bedeutung, zumal aktuelle Daten zeigen, dass selbst auf dem Höhepunkt des letzten Konjunkturaufschwungs im März 2007 etwa jedes fünfte bis sechste Kind in Deutschland von Armut betroffen war (Butterwegge, 2010).

Trotz der offensichtlichen (bildungs-)politischen Relevanz wurde die Einkommensarmut von Schülerfamilien in den meisten großen Schulleistungsstudien der vergangenen Jahre nicht berücksichtigt. Erstmals im Jahr 2007 wurde im Rahmen von IGLU Belgien und TIMSS 2007 in Deutschland das Armutsrisiko – nach der Definition der Europäischen Union (Bardone & Guio, 2005) – in empirischen Bildungsstudien erhoben (Bos, Stubbe & Buddeberg, 2010a, 2010b; Stubbe & Bos, 2008).

Als armutsgefährdet gelten Haushalte, die über weniger als 60 Prozent des medianen Nettoäquivalenzeinkommens in dem jeweiligen Staat verfügen. Zur Berechnung muss zunächst die Armutsgefährdungsgrenze eines Haushalts ermittelt werden, die als Produkt aus dem medianen Nettoäquivalenzeinkommen und der gewichteten Haushaltsgröße (Haushaltsvorstand: 1.0; jedes weitere Haushaltsmitglied, das mindestens 14 Jahre alt ist: 0.5; jedes Kind, das maximal 13 Jahre alt ist: 0.3) definiert ist. Liegt dieser Wert oberhalb des jeweiligen Haushaltsnettoeinkommens, wird der Haushalt als armutsgefährdet bezeichnet.

Empirisch besteht selbstverständlich ein enger Zusammenhang zwischen Armut und anderen Indikatoren des sozialen Status. Der Einsatz dieser Variable verspricht allerdings trotzdem einen zusätzlichen Erkenntnisgewinn, da in den letzten Jahren in Deutschland auch vermehrt Familien mit vergleichsweise hohem kulturellen Kapital von Armut bedroht sind. So wird in einer Studie zum Niedriglohnsektor festgestellt: „Im deutschen Niedriglohnsektor sind vor allem Qualifizierte beschäftigt, die angesichts des Überangebots an Arbeitskräften auch in den einfachen Jobs die gering Qualifizierten verdrängt haben“ (Bosch & Kalina, 2007, S. 97). Hinzu kommen die Alleinerziehenden und Mehrkindfamilien, die – über alle Bildungsniveaus hinweg – überproportional häufig von Armut betroffen sind (Butterwegge, 2000).

Dies lässt den Schluss zu, dass insbesondere für Familien mit einem relativ hohen kulturellen und sozialen Kapital, die von Armut betroffen sind, damit zu rechnen ist, dass die Armut negative Auswirkungen auf den Bildungserfolg der Kinder hat, während in bildungsfernen Haushalten ein deutlich geringerer Effekt von Armut zu finden ist.

¹ Der vierte Armutsbericht liegt den anderen Bundesministerien seit dem 17.09.2012 als Entwurf zur Abstimmung vor (<http://www.handelsblatt.com/politik/deutschland/armuts-und-reichtumsbericht-roessler-ministerium-lehnt-von-der-leyens-bericht-ab/7157010.html>).

3 Der Zusammenhang zwischen sozialer Herkunft und Leistungen in Mathematik und Naturwissenschaften im internationalen Vergleich

International stehen – wie bereits dargestellt – nur wenige Indikatoren für die soziale Herkunft zur Verfügung. Da im Rahmen von TIMSS 2007 ausschließlich die Anzahl der Bücher im Haushalt (fünfstufige Skala: 0–10 Bücher, 11–25 Bücher, 26–100 Bücher, 101–200 Bücher und mehr als 200 Bücher) für internationale Vergleiche genutzt wurde, soll dies an dieser Stelle entsprechend gehandhabt werden, um Vergleiche zwischen den Studienzyklen zu ermöglichen. In diesem Abschnitt werden die Teilnehmerstaaten verglichen, die der *Europäischen Union* (EU) und/oder der *Organisation for Economic Co-operation and Development* (OECD) angehören sowie die Staaten, die auf den Gesamtskalen Mathematik und Naturwissenschaften signifikant besser abschneiden als Deutschland.

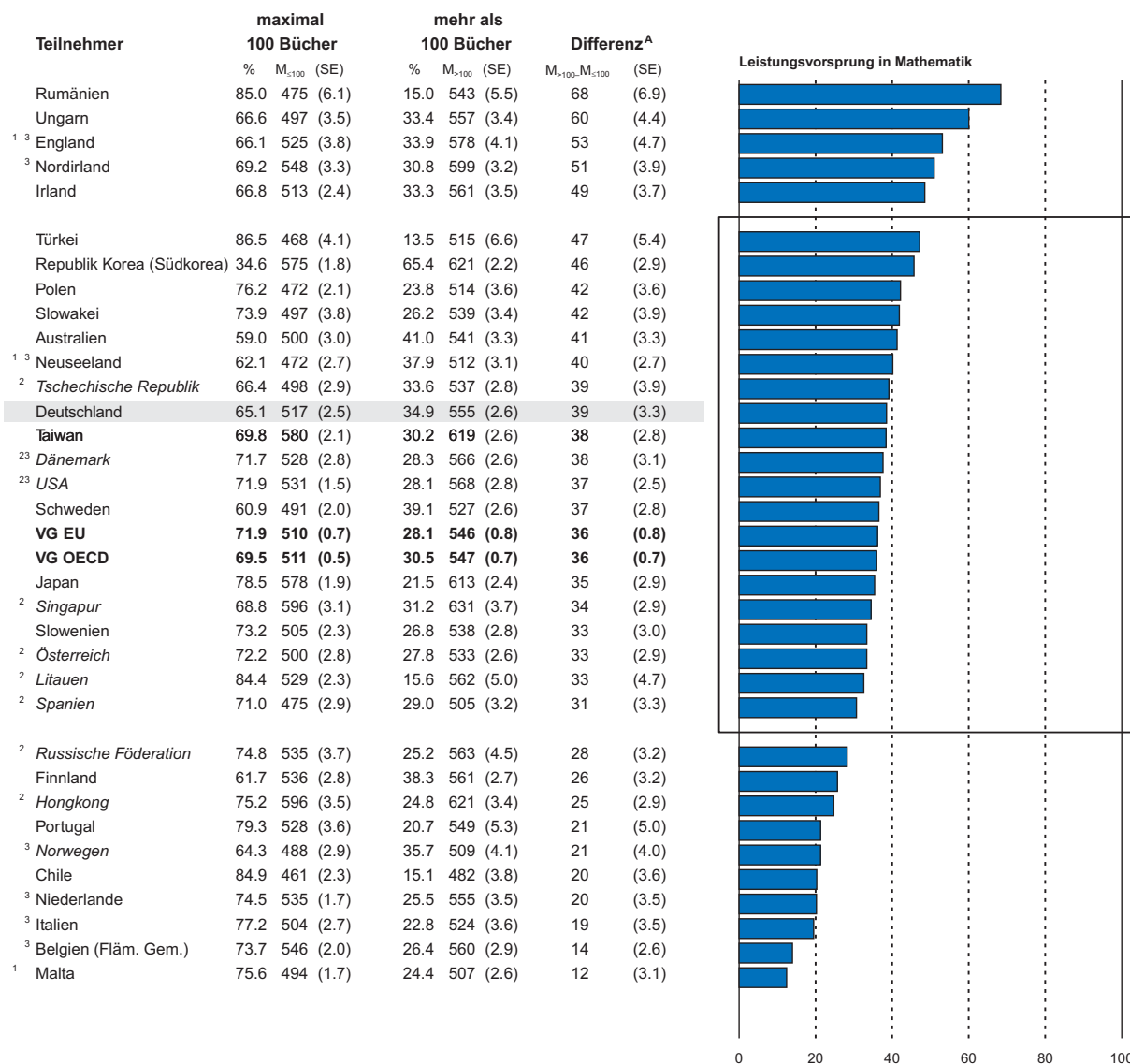
Für die folgenden Analysen werden die Schülerinnen und Schüler, bei denen es zu Hause mehr als 100 Bücher gibt, mit denen verglichen, bei denen es maximal 100 Bücher gibt. Abbildung 7.1 zeigt für die beiden Gruppen die prozentualen Anteile in den einzelnen Staaten sowie die Leistungsmittelwerte für Mathematik. Zudem wird die Differenz zwischen den Leistungsmittelwerten der beiden Gruppen graphisch dargestellt.

Besonders wenige Familien mit mehr als 100 Büchern gibt es in der Türkei (13.5%) sowie in Rumänien (15.0%), Chile (15.1%) und Litauen (15.6%). Der mit Abstand höchste Anteil von Familien mit mehr als 100 Büchern findet sich mit 65.4 Prozent in der Republik Korea (Südkorea), gefolgt von Australien (41.0%), Schweden (39.1%) und Finnland (38.3%). Auch in Deutschland ist der Anteil von Familien mit vielen Büchern überdurchschnittlich hoch (34.9%).

In allen Teilnehmerstaaten findet sich bei den Schülerinnen und Schülern mit mehr als 100 Büchern ein signifikanter Leistungsvorsprung in Mathematik vor denen mit maximal 100 Büchern. In Deutschland liegt dieser mit 39 Punkten bei knapp einem Lernjahr (vgl. Bos et al., 2007). Im internationalen Vergleich liegt Deutschland damit in einer großen Gruppe im Mittelfeld, zu der auch die Vergleichswerte der EU- und der OECD-Staaten gehören. Signifikant größer als in Deutschland sind die sozialen Disparitäten in Rumänien, Ungarn, England, Nordirland und Irland. Signifikant geringer als in Deutschland sind die sozialen Disparitäten in der Russischen Föderation und in Hongkong. Signifikant und besonders auffällig niedriger als in Deutschland sind die Leistungsdifferenzen mit weniger als 25 Punkten in Malta, der Flämischen Gemeinschaft in Belgien, Italien, den Niederlanden, Chile, Norwegen und Portugal.

In Entsprechung zu Abbildung 7.1 zeigt die Abbildung 7.2 die Ergebnisse für den Kompetenzbereich Naturwissenschaften. Hinsichtlich der Anteile von Familien mit mehr als beziehungsweise maximal 100 Büchern gibt es keine Unterschiede zur vorherigen Abbildung und auch hier zeigt sich in allen Staaten ein signifikanter Leistungsvorsprung der Schülerinnen und Schüler mit mehr als 100 Büchern vor Viertklässlerinnen und Viertklässlern mit maximal 100 Büchern.

Abbildung 7.1: Leistungsvorsprung in der Mathematikkompetenz von Kindern aus Familien mit mehr als 100 Büchern vor denen mit maximal 100 Büchern



□ Kein statistisch signifikanter Unterschied zum Differenzwert von Deutschland (p > .05).

■ Statistisch signifikante Unterschiede (p < .05).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

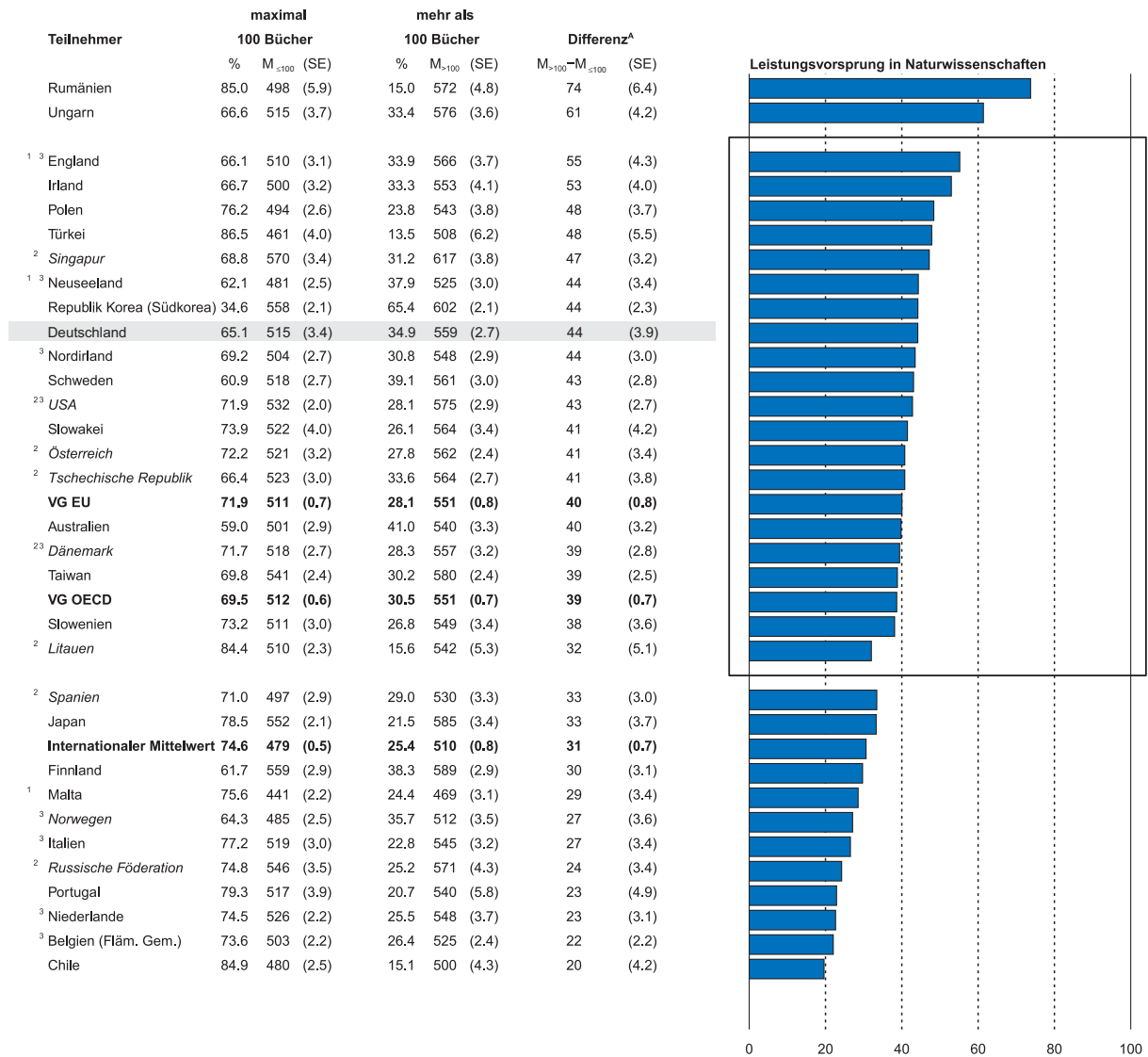
1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A= Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Abbildung 7.2: Leistungsvorsprung in der Naturwissenschaftskompetenz von Kindern aus Familien mit mehr als 100 Büchern vor denen mit maximal 100 Büchern



□ Kein statistisch signifikanter Unterschied zum Differenzwert von Deutschland ($p > .05$).

■ Statistisch signifikante Unterschiede ($p < .05$).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

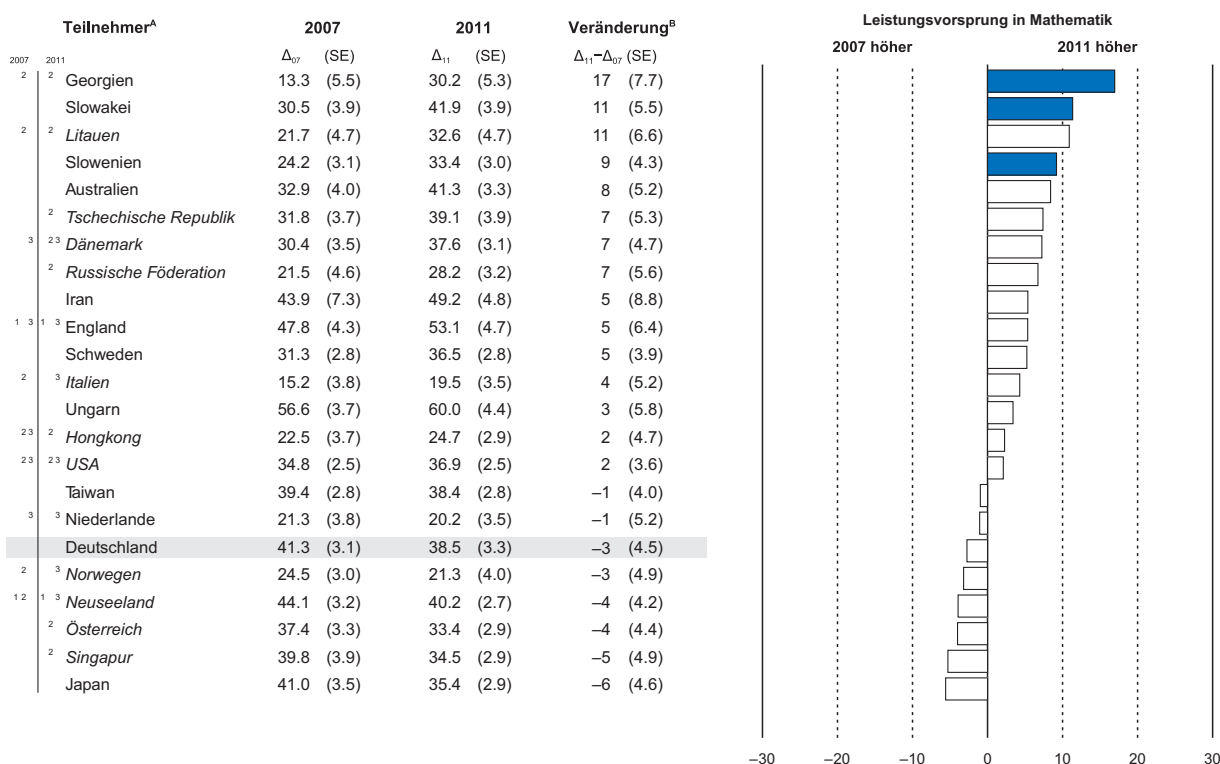
3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

In Deutschland liegt die Differenz zwischen den Leistungsmittelwerten der beiden Gruppen bei 44 Punkten und damit minimal oberhalb des entsprechenden Werts für Mathematik. Nur in zwei Staaten finden sich signifikant größere soziale Disparitäten: 74 Punkte in Rumänien und 61 Punkte in Ungarn. Der Gruppe, in der die Unterschiede im jeweiligen Vergleich zu Deutschland signifikant geringer ausfallen, gehört knapp ein Drittel der hier berichteten Teilnehmerstaaten an. Weniger als 25 Punkte beträgt die Differenz in Chile, der Flämischen Gemeinschaft in Belgien, den Niederlanden, Portugal und der Russischen Föderation.

Da die entsprechenden Daten auch für TIMSS 2007 vorliegen, zeigen die Abbildungen 7.3 und 7.4 den Vergleich der Differenz zwischen den Leistungsmittelwerten der beiden Gruppen für beide Erhebungen. Zudem ist die Veränderung dieser Differenzen von TIMSS 2007 zu TIMSS 2011 dargestellt. In den Tabellen sind diejenigen Staaten aufgeführt, die sowohl an TIMSS 2007 als auch an TIMSS 2011 teilgenommen haben.

Abbildung 7.3: Unterschiede im Leistungsvorsprung in der Mathematikkompetenz von Kindern aus Familien mit mehr als 100 Büchern vor denen mit maximal 100 Büchern im Vergleich von TIMSS 2007 und TIMSS 2011



■ Statistisch signifikante Unterschiede ($p < .05$).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

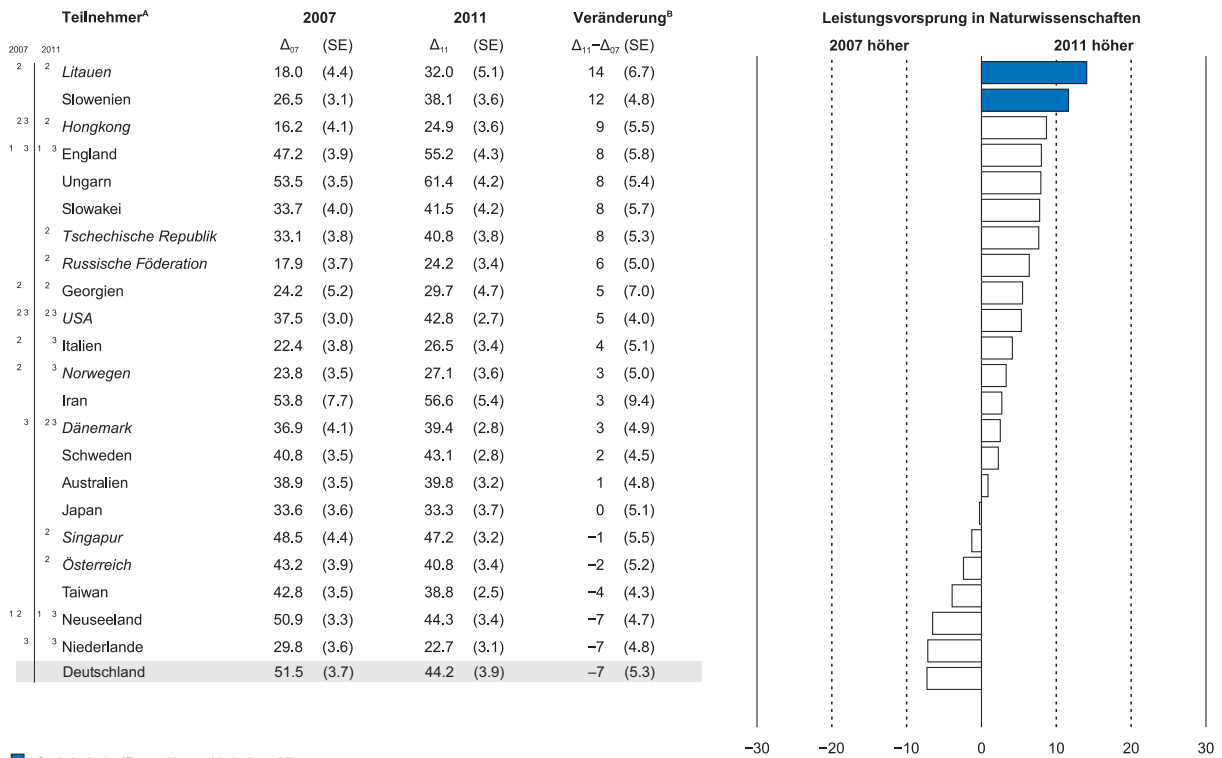
3= Die Teilnahmequoten auf Schul- und/oder Schülerenebene erreichen nicht die internationalen Vorgaben.

A= Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

B= Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Δ = Differenz in den Leistungsmittelwerten.

Abbildung 7.4: Unterschiede im Leistungsvorsprung in der Naturwissenschaftskompetenz von Kindern aus Familien mit mehr als 100 Büchern vor denen mit maximal 100 Büchern im Vergleich von TIMSS 2007 und TIMSS 2011



■ Statistisch signifikante Unterschiede ($p < .05$).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerschicht erreichen nicht die internationalen Vorgaben.

A = Die Ergebnisse von Armenien, Jemen, Kasachstan, Katar, Kuwait, Marokko und Tunesien werden auf Grund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007 und 2011 hier nicht berichtet.

B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Δ = Differenz in den Leistungsmittelwerten.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Für Deutschland zeigen sich beim Vergleich von TIMSS 2007 und 2011 in beiden Domänen keine signifikanten Unterschiede im Ausmaß der sozialen Disparitäten in den Schülerleistungen. Nominell liegt die Differenz zwischen Kindern mit mehr als beziehungsweise maximal 100 Büchern bei TIMSS 2011 in Mathematik 3 Punkte und in den Naturwissenschaften 7 Punkte unter den Werten aus TIMSS 2007. Auch in den meisten anderen Teilnehmerstaaten gibt es in den vier Jahren keine signifikanten Veränderungen. In Georgien, der Slowakei und Slowenien liegen die Differenzen in Mathematik 2011 signifikant über den Werten aus 2007. In den Naturwissenschaften trifft dies auf Litauen und Slowenien zu.

4 Der Zusammenhang zwischen sozialer Herkunft und Mathematik- beziehungsweise Naturwissenschaftskompetenz in Deutschland

International wurden in den TIMSS-Fragebögen nur sehr wenige Indikatoren der sozialen Herkunft eingesetzt (vgl. Abschnitt 3), so dass die Fragebögen in Deutschland um weitere Fragen zu diesem Thema ergänzt wurden. In diesem Abschnitt werden daher ausschließlich nationale Befunde berichtet.

Wie bereits dargestellt, sind die unterschiedlichen Variablen zum Berufsstatus von Personen besonders geeignete Indikatoren für den sozialen Status von Familien. Daher wurden die Eltern in Anlehnung an TIMSS 2007 in offener Form nach ihren Berufen gefragt, um anschließend EGP-Klassen und den ISEI berechnen zu können.

Tabelle 7.1 zeigt die mittlere Mathematik- und Naturwissenschaftskompetenz für die einzelnen EGP-Klassen für die TIMS-Studien, an denen sich Deutschland mit der Grundschule beteiligt hat. Die Veränderungen zwischen den Messzeitpunkten in den einzelnen EGP-Klassen sind mit 1 bis 10 Punkten (sehr) gering und in keinem Fall signifikant. In Mathematik betrug die Differenz zwischen der oberen Dienstklasse und den (Fach-)Arbeitern 2007 43 Punkte und in der aktuellen Studie 33 Punkte. Im Vergleich der oberen Dienstklasse mit den un- und angelernten Arbeitern liegen die entsprechenden Werte bei 60 Punkten (2007) beziehungsweise 46 Punkten (2011). Somit kann festgestellt werden, dass die sozialen Disparitäten in der aktuellen Studie nominell geringer ausfallen als vor vier Jahren. Allerdings sind auch diese Differenzen zwischen den beiden Erhebungszeitpunkten nicht signifikant und lassen auch noch keine Aussagen über Entwicklungsverläufe zu.

Sehr ähnlich fallen die Ergebnisse für den Bereich Naturwissenschaften aus. So liegt die Differenz zwischen der oberen Dienstklasse und den (Fach-)Arbeitern bei 39 Punkten (2007: 52 Punkte) und zwischen der oberen Dienstklasse und den un- und angelernten Arbeitern bei 55 Punkten (2007: 70 Punkte).

In beiden Bereichen führen leichte Verbesserungen in den EGP-Klassen V, VI und VII sowie unveränderte Werte und leichte Verschlechterungen in den EGP-Klassen I, II, III und IV (jeweils nicht signifikant) 2011 nominell zu geringeren Leistungsunterschieden zwischen den EGP-Klassen als 2007.

Der Zusammenhang zwischen den beiden Kompetenzdomänen und einer weiteren Variable für den Berufsstatus – dem ISEI – lässt sich am besten mit Regressionsanalysen darstellen. 2011 liegt der soziale Gradient (Steigung der Regressionsgeraden auf Individualebene) des höchsten ISEI im Haushalt (HISEI) in Mathematik bei 1.07 und somit signifikant unter dem entsprechenden Wert aus TIMSS 2007 (1.41). Auch in den Naturwissenschaften ist der soziale Gradient 2011 signifikant kleiner als vier Jahre zuvor (2011: 1.27; 2007: 1.63). Die Befunde bestätigen also die Tendenz, die bereits bei der Analyse der EGP-Klassen zu erkennen war. Für TIMSS 2011 erklären die Modelle 13.0 Prozent der Varianz der Mathematikkompetenz (2007: 12.7%) und 14.7 Prozent der Varianz der Naturwissenschaftskompetenz (2007: 12.7%).

Ein weiterer Indikator für den sozialen Status, der ausschließlich für Deutschland vorliegt, ist die Armutsgefährdung der Schülerfamilien. Nach der offiziellen Definition der EU gelten 25.4 Prozent der Familien, deren Kinder sich an IGLU/TIMSS 2011 beteiligt haben, als armutsgefährdet. Wie Abbildung 7.5

Tabelle 7.1: Mittlere Mathematik- und Naturwissenschaftskompetenz je EGP-Klasse in Deutschland bei TIMSS 2007 und TIMSS 2011

Sozioökonomische Stellung der Familie (EGP-Klasse) ^A	Mathematik							
	2007				2011			
	M	(SE)	SD	(SE)	M	(SE)	SD	(SE)
Obere Dienstklasse (I)	564	(2.7)	56	(1.9)	558	(3.1)	59	(2.3)
Untere Dienstklasse (II)	548	(2.6)	59	(1.8)	548	(3.0)	58	(2.6)
Routinedienstleistungen (III)	526	(7.6)	66	(6.9)	526	(4.9)	58	(3.9)
Selbstständige (IV)	534	(4.1)	63	(3.7)	524	(4.8)	54	(4.4)
(Fach-)Arbeiter (V, VI)	521	(3.0)	65	(2.1)	524	(2.9)	58	(2.7)
Un- und angelernte Arbeiter (VII)	504	(3.9)	67	(3.3)	512	(3.1)	57	(2.1)
Gesamt	525	(2.3)	68	(1.2)	528	(2.2)	62	(1.4)

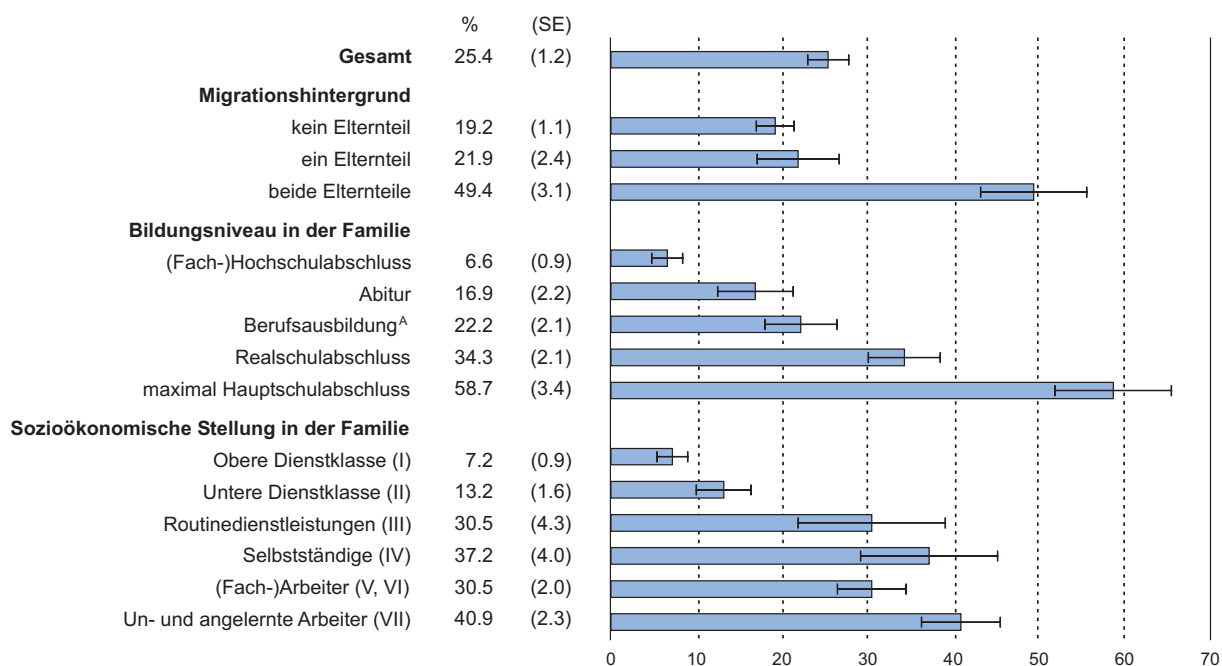
Sozioökonomische Stellung der Familie (EGP-Klasse) ^A	Naturwissenschaften							
	2007				2011			
	M	(SE)	SD	(SE)	M	(SE)	SD	(SE)
Obere Dienstklasse (I)	572	(3.1)	64	(2.4)	563	(3.5)	64	(2.5)
Untere Dienstklasse (II)	554	(2.9)	68	(2.8)	552	(3.4)	64	(2.6)
Routinedienstleistungen (III)	536	(6.8)	70	(6.9)	528	(6.1)	65	(3.8)
Selbstständige (IV)	533	(5.2)	72	(3.8)	528	(5.8)	62	(4.1)
(Fach-)Arbeiter (V, VI)	521	(3.3)	76	(2.8)	524	(3.1)	64	(2.7)
Un- und angelernte Arbeiter (VII)	503	(4.2)	76	(3.3)	508	(3.6)	63	(2.2)
Gesamt	528	(2.4)	79	(1.5)	528	(2.9)	70	(1.3)

A = Die EGP- Klassifikation erfolgte für den Studienzyklus 2007 auf Basis von ISCO-88-Kodes (ILO, 1990) und für den Studienzyklus 2011 auf Basis von ISCO-08-Kodes (ILO, 2012).

zu entnehmen ist, variiert diese Quote allerdings stark zwischen unterschiedlichen gesellschaftlichen Gruppen.

So ist knapp die Hälfte der Familien, in denen beide Eltern im Ausland geboren wurden, armutsgefährdet, während dies nur etwa jede fünfte Familie betrifft, bei denen mindestens ein Elternteil in Deutschland geboren wurde. Ein noch deutlicherer Zusammenhang zeigt sich zwischen dem höchsten Bildungsniveau der Eltern und der Armutsgefährdungsquote. Verfügen die Eltern maximal über einen Hauptschulabschluss, beträgt die Armutsgefährdungsquote 58.7 Prozent. Hingegen sind etwa ein Drittel der Familien armutsgefährdet, in denen mindestens ein Elternteil über einen Realschulabschluss verfügt, während dies etwa ein Sechstel der Familien betrifft, in denen mindestens ein Elternteil das Abitur gemacht hat. Besitzt mindestens ein Elternteil einen Hochschulabschluss, liegt die Armutsgefährdungsquote bei nur 6.6 Prozent. Ähnliche Ergebnisse zeigen sich für die EGP-Klassen: Hier variieren die Quoten zwischen 7.2 Prozent in der oberen Dienstklasse und 40.9 Prozent bei den un- und angelernten Arbeitern.

Wie der Tabelle 7.2 entnommen werden kann, liegt die Mathematikkompetenz von Kindern aus nicht armutsgefährdeten Elternhäusern mit 544 Punkten 37 Punkte über der Mathematikkompetenz von armutsgefährdeten

Abbildung 7.5: Armutsgefährdungsquoten nach Migrationshintergrund, höchstem Bildungsabschluss und EGP-Klasse (in Prozent)

A = Berufsausbildung (duales System, Berufsschulsystem) bzw. Fachhochschulreife.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Schülerinnen und Schülern. Für den Bereich Naturwissenschaften fällt diese Differenz mit 43 Punkten geringfügig größer aus. Im Vergleich zu TIMSS 2007 ergeben sich keine signifikanten Veränderungen: Die Differenz für die Mathematikkompetenz betrug vier Jahre zuvor 44 Punkte und die Differenz für die Naturwissenschaftskompetenz 53 Punkte (vgl. Bos et al., 2010a).

Bedingt durch die geringe Größe der Teilstichproben – insbesondere für armutsgefährdete Kinder aus bildungsnahen Elternhäusern – erlauben die vorliegenden Daten kaum generalisierbare Aussagen zu Unterschieden zwischen den einzelnen EGP-Klassen. Um die Hypothese prüfen zu können, ob in bildungsnahen Elternhäusern ein engerer Zusammenhang zwischen Armutsgefährdung und Kompetenz besteht als in bildungsfernen, zeigt die Tabelle 7.3 daher die Ergebnisse von Regressionsanalysen zur Erklärung der Mathematik- beziehungsweise Naturwissenschaftskompetenz durch den HISEI und den Interaktionseffekt aus HISEI und Armutsgefährdung. Der Zusammenhang mit dem HISEI fällt auch in diesen Modellen positiv aus (s.o.). Der negative Interaktionseffekt aus HISEI und Armutsgefährdung macht deutlich, dass bei armutsgefährdeten Familien der Zusammenhang zwischen dem Berufsstatus und der Mathematik- beziehungsweise Naturwissenschaftsleistung der Kinder erwartungsgemäß geringer ausfällt als bei nicht armutsgefährdeten Familien.

Tabelle 7.2: Mittlere Mathematik- und Naturwissenschaftskompetenz nach Armutsgefährdung und EGP-Klasse

Sozioökonomische Stellung der Familie (EGP- Klassen)	Mathematik									
	nicht armutsgefährdet				armutsgefährdet				Differenz ^A	
	M _n	(SE)	SD	(SE)	M _a	(SE)	SD	(SE)	M _n -M _a	(SE)
Obere Dienstklasse (I)	560	(3.2)	57	(2.5)	532	(12.1)	67	(9.1)	28	(11.7)
Untere Dienstklasse (II)	554	(3.5)	56	(2.5)	513	(7.3)	56	(6.5)	42	(8.4)
Routinedienstleistungen (III)	534	(7.7)	62	(5.8)	505	(8.6)	46	(5.9)	29	(12.5)
Selbstständige (IV)	530	(4.9)	51	(3.8)	514	(9.1)	59	(8.9)	16	(9.6)
(Fach-)Arbeiter (V, VI)	532	(3.2)	56	(3.8)	514	(5.4)	58	(3.7)	18	(5.9)
Un- und angelernte Arbeiter (VII)	524	(3.2)	56	(3.1)	500	(4.5)	54	(3.7)	24	(5.1)
Gesamt	544	(2.1)	58	(1.5)	507	(3.0)	59	(2.3)	36	(3.0)

Sozioökonomische Stellung der Familie (EGP- Klassen)	Naturwissenschaften									
	nicht armutsgefährdet				armutsgefährdet				Differenz ^A	
	M _n	(SE)	SD	(SE)	M _a	(SE)	SD	(SE)	M _n -M _a	(SE)
Obere Dienstklasse (I)	566	(3.3)	62	(2.9)	528	(12.1)	72	(12.5)	38	(11.1)
Untere Dienstklasse (II)	558	(3.6)	62	(2.2)	515	(8.6)	66	(6.7)	43	(9.4)
Routinedienstleistungen (III)	535	(8.2)	69	(6.2)	507	(9.7)	53	(7.9)	28	(11.4)
Selbstständige (IV)	534	(6.7)	58	(4.7)	514	(8.9)	71	(7.8)	20	(10.8)
(Fach-)Arbeiter (V, VI)	533	(3.7)	64	(2.7)	511	(5.6)	62	(5.5)	22	(6.9)
Un- und angelernte Arbeiter (VII)	522	(4.2)	61	(2.7)	495	(4.5)	60	(2.8)	27	(5.9)
Gesamt	547	(2.4)	65	(1.3)	504	(3.6)	65	(2.9)	43	(3.5)

A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Tabelle 7.3: Regressionsanalysen zum Interaktionseffekt zwischen HISEI und Armutsgefährdung bei der Erklärung von Mathematik- und Naturwissenschaftskompetenz (TIMSS 2007 und 2011)

	Mathematik		Naturwissenschaften	
	2007	2011	2007	2011
Konstante	474.77	489.36	469.43	482.26
Berufsstatus (HISEI ^A)	1.31	0.98	1.50	1.16
Interaktion: Berufsstatus mit Armutsgefährdung ^B	-0.37	-0.40	-0.42	-0.43
R ²	0.14	0.14	0.14	0.15

A = Höchster ISEI (*International Socio-Economic Index of Occupational Status*) im Haushalt.

B = Interaktion von HISEI mit Armutsgefährdung.

Alle berichteten Koeffizienten sind signifikant ($p < .05$).

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

5 Zusammenfassung

Ebenso wie bei TIMSS 2007 finden sich auch in der aktuellen Studie in allen Teilnehmerstaaten signifikante Unterschiede in den mathematischen und naturwissenschaftlichen Kompetenzen zwischen den Schülerinnen und Schülern der oberen beziehungsweise der unteren sozialen Lagen. Die Notwendigkeit, soziale Disparitäten bei der Beschreibung von Bildungssystemen zu berücksichtigen, wird somit auch durch die aktuelle Untersuchung bekräftigt.

Allerdings muss auch betont werden, dass sich das Ausmaß der sozialen Disparitäten zwischen den Teilnehmerstaaten zum Teil erheblich unterscheidet. Dies kann entweder bedeuten, dass die Unterschiede zwischen den oberen und den unteren sozialen Lagen in einigen Staaten ausgeprägter sind als in anderen, oder dass es einigen Staaten besser gelingt als anderen, die primären Herkunftseffekte der Schülerinnen und Schüler im Laufe der ersten vier Schuljahre zu verringern.

Geringe soziale Disparitäten in den Schülerleistungen in Mathematik und Naturwissenschaft finden sich in der Flämischen Gemeinschaft in Belgien, Chile, Finnland, Italien, Malta, den Niederlanden, Norwegen, Portugal und der Russischen Föderation. Die mit Abstand größten Unterschiede zwischen den sozialen Lagen weisen die osteuropäischen Staaten Rumänien und Ungarn auf, in Mathematik sind aber auch in England, Nordirland und Irland die Differenzen sehr ausgeprägt. Deutschland liegt in diesem Punkt im Mittelfeld mit leicht überdurchschnittlichen Werten. Zwischen 2007 und 2011 hat es in Deutschland – wie auch in den meisten anderen Staaten – keine signifikanten Veränderungen gegeben.

National erlauben die Daten aus TIMSS 2011 vertiefende Analysen, da hierzulande mehr Indikatoren der sozialen Lage eingesetzt wurden. Es bestätigen sich die internationalen Befunde dahingehend, dass ausgeprägte soziale Disparitäten im deutschen Bildungssystem vorhanden sind, wobei die Leistungsunterschiede zwischen Kindern, deren Eltern den oberen Dienstklassen angehören, und Kindern, deren Eltern den Gruppen der Arbeiter angehören, etwa knapp ein Lernjahr betragen. Selbiges gilt, wenn Viertklässlerinnen und Viertklässler aus armutsgefährdeten Elternhäusern mit Schülerinnen und Schülern verglichen werden, die nicht armutsgefährdet sind. Beim Vergleich mit den Ergebnissen aus TIMSS 2007 zeigen sich hinsichtlich der Unterschiede zwischen den EGP-Klassen zwar keine signifikanten Veränderungen, Regressionsanalysen mit dem HISEI zeigen aber, dass das Ausmaß der sozialen Disparitäten 2011 etwas geringer ausfällt als noch vier Jahre zuvor.

Literatur

- Bardone, L. & Guio, A.-C. (2005). *In-work poverty. New commonly agreed indicators at the EU level*. 5. Zugriff am 18.10.2012 unter http://ec.europa.eu/employment_social/social_inclusion/docs/statistics5-2005_en.pdf.
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–407). Opladen: Leske + Budrich.

- Becker, R. & Lauterbach, W. (2004). Dauerhafte Bildungsungleichheiten – Ursachen, Mechanismen, Prozesse und Wirkungen. In R. Becker & W. Lauterbach (Hrsg.), *Bildung als Privileg?* (S. 9–40). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Blossfeld, H.-P. (1993). Changes in educational opportunities in the Federal Republic of Germany. A longitudinal study of cohorts born between 1916 and 1965. In Y. Shavit & H.-P. Blossfeld (Hrsg.), *Persistent inequality. Changing educational attainment in thirteen countries* (S. 51–74). Boulder: Westview Press.
- Bonsen, M., Frey, K. A. & Bos, W. (2008). Soziale Herkunft. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 141–156). Münster: Waxmann.
- Bos, W., Schwippert, K. & Stubbe, T. C. (2007). Die Koppelung von sozialer Herkunft und Schülerleistung im internationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 225–247). Münster: Waxmann.
- Bos, W., Stubbe, T. C. & Buddeberg, M. (2010a). Einkommensarmut und schulische Kompetenzen. In J. Fischer & R. Merten (Hrsg.), *Armut und soziale Ausgrenzung von Kindern und Jugendlichen – Problembestimmungen und Interventionsansätze*. (Grundlagen der sozialen Arbeit, Band 26, S. 58–72). Baltmannsweiler: Schneider Verlag Hohengehren.
- Bos, W., Stubbe, T. C. & Buddeberg, M. (2010b). Gibt es eine armutsbedingte Benachteiligung? Die Operationalisierung verschiedener Indikatoren der sozialen Herkunft in der empirischen Bildungsforschung. In D. H. Rost (Hrsg.), *Intelligenz, Hochbegabung, Vorschulerziehung, Bildungsbenachteiligung* (S. 165–208). Münster: Waxmann.
- Bosch, G. & Kalina, T. (2007). Niedriglöhne in Deutschland – Zahlen, Fakten, Ursachen. In G. Bosch & C. Weinkopf (Hrsg.), *Arbeiten für wenig Geld. Niedriglohnbeschäftigung in Deutschland* (S. 20–105). Frankfurt a. M.: Campus.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (Soziale Welt, Sonderband 2, S. 183–198). Göttingen: Schwartz.
- Bundesministerium für Arbeit und Soziales. (2001). *Lebenslagen in Deutschland. Daten und Fakten. Materialband zum ersten Armuts- und Reichtumsbericht der Bundesregierung*. Zugriff am 18.10.2012 unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/a268-lebenslagen-in-deutschland-armutsbericht-1-daten-fakten.pdf?__blob=publicationFile.
- Bundesministerium für Arbeit und Soziales. (2005). *Lebenslagen in Deutschland. Der 2. Armuts- und Reichtumsbericht der Bundesregierung*. Zugriff am 18.10.2012 unter http://www.sozialpolitik-aktuell.de/tl_files/sozialpolitik-aktuell/_Kontrovers/NeueArmutDeutschland/Lebenslagen%20in%20Deutschland_EndBericht.pdf.
- Bundesministerium für Arbeit und Soziales. (2008). *Lebenslagen in Deutschland. Der 3. Armuts- und Reichtumsbericht der Bundesregierung*. Zugriff am 18.10.2012 unter http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen-DinA4/forschungsprojekt-a333-dritter-armuts-und-reichtumsbericht.pdf;jsessionid=EDDB008BB946B813E7D1BB8BB9237860?__blob=publicationFile.
- Butterwegge, C. (2010). Kinderarmut und Bildung. In G. Quenzel & K. Hurrelmann (Hrsg.), *Bildungsverlierer. Neue Ungleichheiten* (S. 537–555). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Butterwegge, C. (Hrsg.). (2000). *Kinderarmut in Deutschland. Ursachen, Erscheinungsformen und Gegenmaßnahmen*. Frankfurt/Main: Campus.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94 (1), 95–120.
- Ehmke, T. & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. 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. 231–254). Münster: Waxmann.
- Erikson, R., Goldthorpe, H. J. & Portocarero, L. (1979). Intergenerational class mobility in three western european societies: England, France and Sweden. *British Journal of Sociology*, 30 (4), 415–441.

- Ganzeboom, H. B. G. & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 international standard classification of occupations. *Social Science Research*, 25 (3), 201–239.
- Ganzeboom, H. B. G., Graaf, P. M. de & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21 (1), 1–56.
- Hoffmann, E. (2003). International statistical comparisons of occupational and social structures. Problems, possibilities and the role of ISCO-88. In J. H. Hoffmeyer-Zlotnik & C. Wolf (Hrsg.), *Advances in cross-national comparison. A European working book for demographic and socio-economic variables* (S. 137–158). New York: Plenum Press.
- ILO. (1990). *International Standard Classification of Occupations. ISCO-88*. Genève: International Labour Office.
- ILO. (2012). *International Standard Classification of Occupations. ISCO-08*. Genève: International Labour Office.
- 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. 191–208). Münster: Waxmann.
- Schimpl-Neimanns, B. (2000). Soziale Herkunft und Bildungsbeteiligung. Empirische Analysen zu herkunftsspezifischen Bildungsungleichheiten zwischen 1950 und 1989. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 52 (4), 636–669.
- Schroedter, J. H., Lechert, Y. & Lüttinger, P. (2006). *Die Umsetzung der Bildungsskala ISCED -1997 für die Volkszählung 1970, die Mikrozensus-Zusatzerhebung 1971 und die Mikrozensus 1976-2004. ZUMA-Methodenbericht 2006/08*. Zugriff am 06.2006 unter http://www.gesis.org/Publikationen/Berichte/ZUMA_Methodenberichte/documents/pdfs/2006/06_08_Schroedter.pdf.
- Solga, H. (2005). Meritokratie – die moderne Legitimation ungleicher Bildungschancen. In P. A. Berger & H. Kahlert (Hrsg.), *Institutionalisierte Ungleichheiten. Wie das Bildungswesen Chancen blockiert* (S. 19–38). Weinheim: Juventa.
- Stubbe, T. C. & Bos, W. (2008). Die Koppelung von sozialer Herkunft und Lesekompetenz im internationalen Vergleich und in der Deutschsprachigen Gemeinschaft. In W. Bos, S. Sereni & T. C. Stubbe (Hrsg.), *IGLU Belgien. Lese- und Orthografiekompetenzen von Grundschulkindern in der Deutschsprachigen Gemeinschaft* (S. 111–125). Münster: Waxmann.
- Stubbe, T. C., Bos, W. & Euen, B. (2012). Der Übergang von der Primar- in die Sekundarstufe. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Stubbe, T. C., Bos, W. & Hornberg, S. (2008). Soziale und kulturelle Disparitäten der Schülerleistungen in den Ländern der Bundesrepublik Deutschland. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU-E 2006. Die Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (S. 103–109). Münster: Waxmann.
- UNESCO. (2003). International Standard Classification of Education, ISCED 1997. In J. H. Hoffmeyer-Zlotnik & C. Wolf (Hrsg.), *Advances in cross-national comparison. A European working book for demographic and socio-economic variables*. New York: Plenum Press.
- Vester, M. (2004). Die Illusion der Bildungsexpansion. Bildungsöffnungen und soziale Segregation in der Bundesrepublik Deutschland. In S. Engler & B. Kraus (Hrsg.), *Das kulturelle Kapital und die Macht der Klassenstrukturen. Sozialstrukturelle Verschiebungen und Wandlungsprozesse des Habitus* (S. 13–54). Weinheim: Juventa.
- Vester, M. (2005). Die selektive Bildungsexpansion. Die ständische Regulierung der Bildungschancen in Deutschland. In P. A. Berger & H. Kahlert (Hrsg.), *Institutionalisierte Ungleichheiten. Wie das Bildungswesen Chancen blockiert* (S. 39–70). Weinheim: Juventa.



Kapitel VIII

Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund

Irmela Tarelli, Knut Schwippert und Tobias C. Stubbe

1 Einführung

Migration ist kein einheitlicher sozialer Sachverhalt, sondern hat viele Facetten, die bedeutend für die Bildungsintegration sind. Modifizierte Einbürgerungsrechte und sich verändernde Migrationsbewegungen stellen besondere Herausforderungen an das Bildungssystem (Konsortium Bildungsberichterstattung, 2006; Stanat, 2008; Stanat, Rauch & Segeritz, 2010a). Neben der Multikulturalität und der damit einhergehenden Bereicherung des Schulalltags stellt die Integration und Förderung von Kindern und Jugendlichen, die noch nicht lange in Deutschland leben, spezielle Anforderungen an die Gestaltung von Schule und Unterricht. Insbesondere Schülerinnen und Schüler mit defizitären landessprachlichen Kenntnissen stellen eine besondere Schülergruppe dar, die bei der Gestaltung von Unterricht berücksichtigt werden sollte.

Die statistische Erfassung von Zuwanderung hat sich in den vergangenen Jahren konzeptionell gewandelt. Wurde in der amtlichen Statistik lange nach der Staatsangehörigkeit und damit nach der im Pass dokumentierten Nationalität unterschieden, haben auch die internationalen Schulleistungsstudien zu einer Veränderung von einem Ausländer- zu einem Migrationskonzept beigetragen. Ausgangspunkt war, dass sich die Orientierung an der Staatsangehörigkeit zur Identifizierung von Förderbedarfen in der Schul- und Unterrichtsforschung als nicht mehr angemessen herausgestellt hat (Haag, Böhme & Stanat, 2012). Erstmals wurden im Rahmen von TIMSS 1995 Schülerinnen und Schüler mit Migrationshintergrund nach ihrem familiären Sprachgebrauch und ihrem Geburtsland differenziert (Schwippert & Schnabel, 2000). In nachfolgenden Studien rückte zunächst das Geburtsland der Eltern oder auch das Geburtsland der befragten Schülerinnen und Schüler ins Zentrum der Betrachtung. Anhand dieser Indikatoren konnte der Migrationshintergrund weit differenzierter erhoben werden als über die im Pass dokumentierte Nationalität (Walter, 2008).

In der letzten Dekade konnte im Rahmen verschiedener Untersuchungen zur Lage von Kindern und Jugendlichen aus Familien mit Migrationsgeschichte der familiäre Sprachgebrauch als ein zentraler Aspekt für die Erklärung des Bildungserfolges von Schülerinnen und Schülern aufgezeigt werden (Stanat et al., 2010a; Stanat, Schwippert & Gröhlich, 2010b). Die Befundlage erscheint hierbei eindeutig: Kinder mit Migrationserfahrung, die mit ihren Eltern zu Hause Deutsch sprechen, weisen nicht nur in sprachbezogenen Untersuchungen bessere Leistungen auf, sondern auch in Studien, die andere für die Schule zentrale Kompetenzen in den Blick nehmen (Stanat et al., 2010a).

In diesem Kapitel werden Analysen zu den mathematischen und naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern der vierten Jahrgangsstufe aus Familien mit Migrationsgeschichte vorgestellt. Im ersten Teil des Kapitels werden zunächst im Rahmen eines internationalen Vergleichs anhand des familiären Sprachgebrauchs die Kompetenzen von Kindern mit und ohne Migrationshintergrund verglichen. Für den internationalen Vergleich wird als Indikator für einen Migrationshintergrund auf Informationen zum Sprachgebrauch in der Familie zurückgegriffen. Inwiefern eine entsprechende Operationalisierung die Bevölkerungssituationen aller Teilnehmerstaaten angemessen repräsentiert, ist im Einzelfall anhand der TIMSS-Enzyklopädie, in der die Bildungssysteme der Teilnehmerstaaten beschrieben sind, zu überprüfen (Mullis et al., 2012). Für den internationalen Vergleich werden diejenigen TIMSS-Teilnehmerstaaten berücksichtigt, die Mitglieder der *Europäischen Union* (EU) sind, die der *Organisation for Economic Co-operation and Development* (OECD) angehören oder deren Leistungswerte auf den Gesamtskalen Mathematik oder Naturwissenschaften signifikant über dem für Deutschland berichteten Mittelwert liegen.

Im zweiten Teil des Kapitels werden auf Deutschland fokussierte Analysen zu Merkmalen von Schülerinnen und Schülern mit Migrationshintergrund und deren Familien berichtet. Dies geschieht unter Bezugnahme auf das Geburtsland der Eltern, wobei folgende Gruppen von Schülerinnen und Schülern unterschieden werden:

- die Eltern sind in Deutschland geboren,
- ein Elternteil ist in Deutschland und ein Elternteil im Ausland geboren,
- die Eltern sind im Ausland geboren.

Die Befunde werden um eine Betrachtung von Unterschieden auf nationaler Ebene zwischen TIMSS 2007 und 2011 ergänzt.

Die Nutzung unterschiedlicher Indikatoren für die Identifizierung von Schülerinnen und Schülern mit Migrationshintergrund ist im Studiendesign begründet. Aus forschungspragmatischen Gründen hat die internationale Studienleitung die Erfassung des Migrationshintergrunds im Rahmen von TIMSS 2011 auf die Frage reduziert, inwieweit die Testsprache der zu Hause gesprochenen Sprache entspricht. Im Vergleich zur Erhebung von TIMSS 2007 stellt dies einen Konzeptwechsel dar, zumal die bereits bestehenden Antwortkategorien modifiziert wurden, was für die Beschreibung von Veränderungen problematisch ist. Mit Blick auf eine internationale Indikatorisierung von Schülerinnen und Schülern mit Migrationshintergrund erscheint die Erfassung des Geburtslandes als Minimalinformation akzeptabel (Martin, Mullis & Kennedy, 2003, 2007). Um vertiefende Analysen und Vergleiche mit dem Teildatensatz für Deutschland zu ermöglichen, wurde der in Deutschland eingesetzte Schülerfragebogen in Anlehnung an den vorangegangenen Studienzyklus um Fragen zum Geburtsland der Eltern ergänzt.

2 Mathematische und naturwissenschaftliche Kompetenzen der Schülerinnen und Schüler nach familiärem Sprachgebrauch im internationalen Vergleich

Als wesentlicher Indikator für den Integrationserfolg kann die Beherrschung der Sprache des Einwanderungslandes betrachtet werden (Esser, 2006). Migrantinnen und Migranten, deren Heimatsprache sich nicht von der Amts- oder Verkehrssprache im Einwanderungsland unterscheidet, finden sich im Einwanderungsland leichter zurecht; ihre Integration verläuft schneller und erfolgreicher, als dies bei Einwanderern der Fall ist, die die Amts- oder Verkehrssprache erst erlernen müssen.

Auch am Bildungserfolg und an schulisch vermittelten Kompetenzen lässt sich demnach eine erfolgreiche Integration ausmachen (Walter & Taskinen, 2007). Da in Schulen in Deutschland in der Regel Deutsch auch die Unterrichtssprache ist, sind Kenntnisse der deutschen Sprache dabei zentral (Gogolin, 1994; Gogolin, Neumann & Roth, 2003).

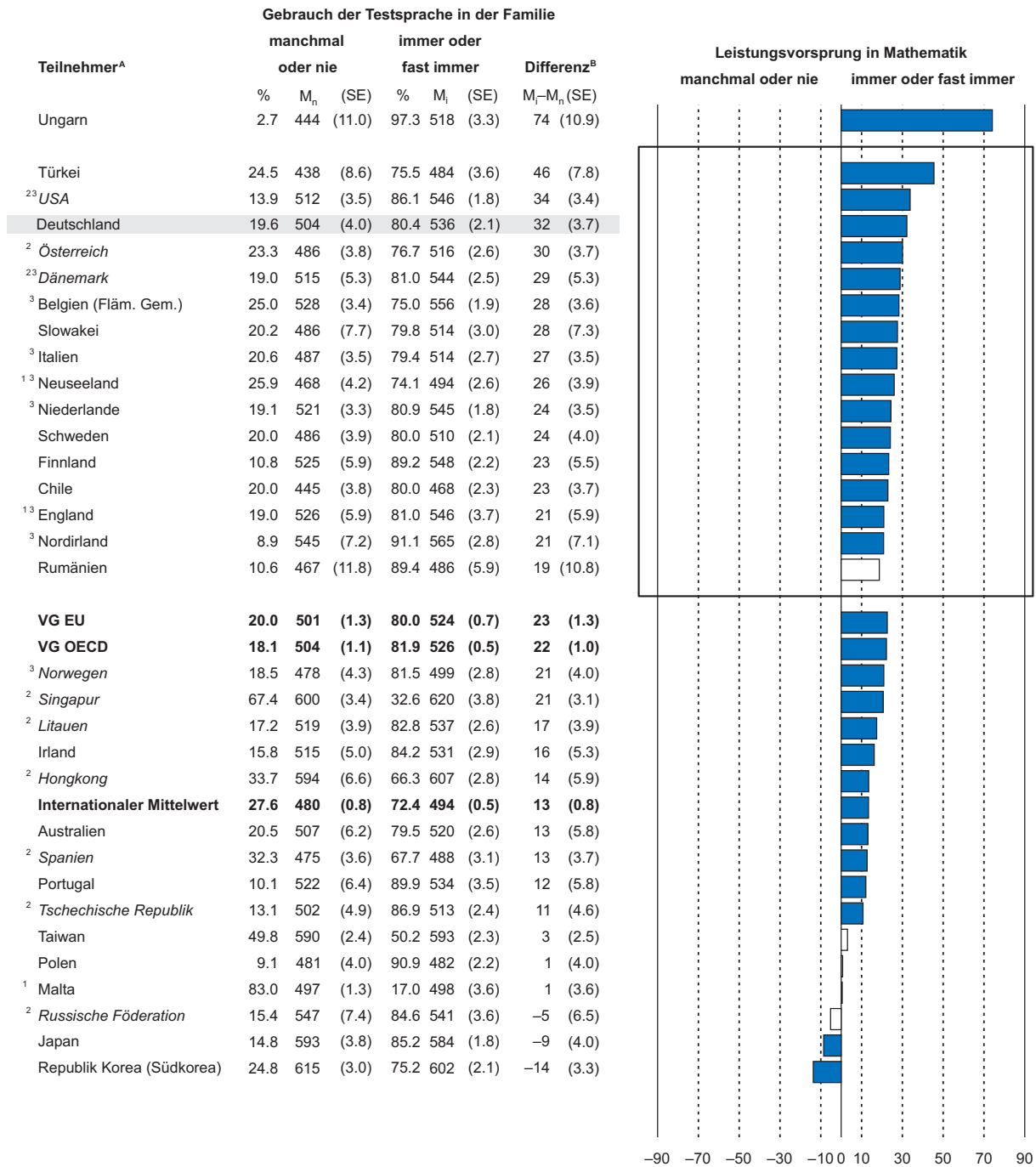
Im tabellarischen Teil der Abbildung 8.1 sind die Verteilungen des familiären Sprachgebrauchs und die mathematischen Leistungen der jeweiligen Gruppen im internationalen Vergleich dargestellt. Der graphische Teil der Abbildung veranschaulicht den Leistungsvorsprung in Mathematik für diejenigen Kinder, die angeben zu Hause immer oder fast immer die Testsprache zu sprechen, vor jenen, die angeben zu Hause nur manchmal oder nie die Testsprache zu sprechen.

Im Schülerfragebogen gaben 19.6 Prozent der Viertklässlerinnen und Viertklässler in Deutschland an, zu Hause nur manchmal (18.9%) oder nie (0.8%) Deutsch zu sprechen (vgl. Schwippert, Wendt & Tarelli, 2012). Für die Analysen in diesem Kapitel wurden diese Kategorien zusammengefasst. Wie in Abbildung 8.1 ersichtlich, ergibt sich bezüglich der Leistungsmittelwerte im Bereich Mathematik zwischen den Schülerinnen und Schülern, die angeben zu Hause immer oder fast immer Deutsch zu sprechen und denjenigen, die angeben zu Hause nur manchmal oder nie Deutsch zu sprechen, eine Differenz von 32 Punkten zugunsten der Kinder, die angeben zu Hause immer oder fast immer Deutsch zu sprechen. Diese Differenz entspricht dem Leistungsvorsprung beziehungsweise -rückstand nach rund einem Schuljahr und verweist damit auf durchaus nicht unerhebliche Disparitäten zwischen den beiden Schülergruppen.

Vergleichbare Unterschiede beziehungsweise Unterschiede, die sich nicht signifikant vom Differenzwert für Deutschland unterscheiden, finden sich in 15 der hier betrachteten 32 Teilnehmerstaaten, zum Beispiel in Österreich (30 Punkte), Dänemark (29), in der Flämischen Gemeinschaft in Belgien (28), in der Slowakei (28) und Italien (27). Die für Deutschland festgestellte Differenz liegt signifikant über den mittleren Differenzen der Teilnehmerstaaten, die Mitglieder der EU (VG_{EU} 23 Punkte) und der OECD (VG_{OECD} 22 Punkte) sind, und auch über der mittleren Differenz der TIMSS-Teilnehmerstaaten (Internationaler Mittelwert 13 Punkte).

Signifikant größer als in Deutschland ist dieser Leistungsunterschied nur in Ungarn (74 Punkte). Dieser Befund deckt sich auch mit dem Befund für den Bereich Naturwissenschaften (vgl. Abbildung 8.2) und ist ähnlich wie der entsprechende Befund zum Leseverständnis (vgl. Schwippert et al., 2012). Keine signifikanten Unterschiede zwischen Schülerinnen und Schülern, die immer die Testsprache zu Hause sprechen und Kindern, auf die dies nicht zutrifft, lassen sich für Taiwan, Polen, Japan und die Russische Föderation verzeichnen. In der

Abbildung 8.1: Mathematische Kompetenzen der Schülerinnen und Schüler nach zu Hause gesprochener Sprache im internationalen Vergleich



□ Kein statistisch signifikanter Unterschied zum Differenzwert von Deutschland (p > .05).

■ Statistisch signifikante Unterschiede (p < .05).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

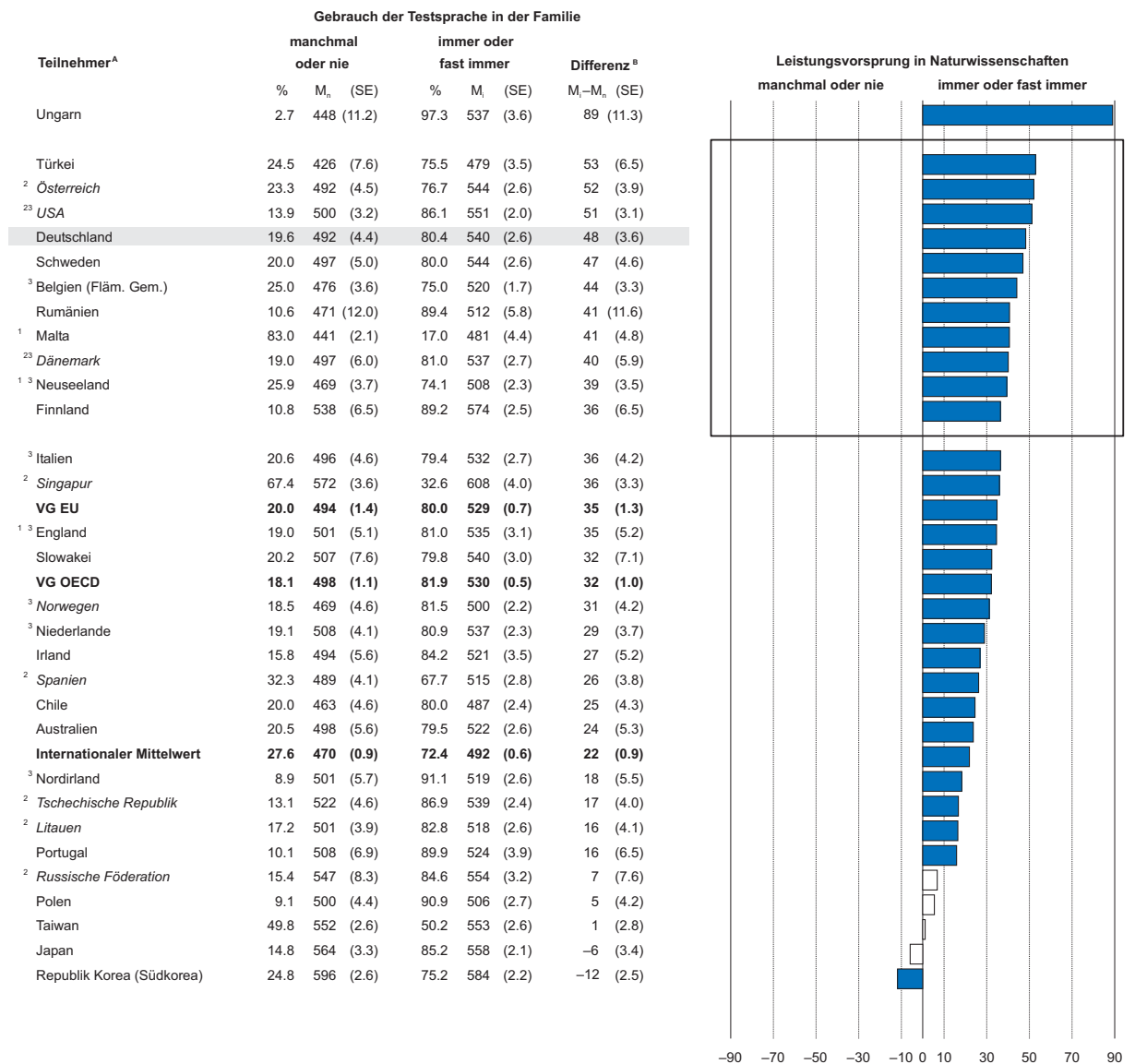
2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Da in Slowenien die Testsprache nicht erhoben wurde, können hier für diesen Teilnehmer keine Ergebnisse berichtet werden.

B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Abbildung 8.2: Naturwissenschaftliche Kompetenzen der Schülerinnen und Schüler nach zu Hause gesprochener Sprache im internationalen Vergleich



□ Kein statistisch signifikanter Unterschied zum Differenzwert von Deutschland ($p > .05$).

■ Statistisch signifikante Unterschiede ($p < .05$).

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1= Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2= Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3= Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Da in Slowenien die Testsprache nicht erhoben wurde, können hier für diesen Teilnehmer keine Ergebnisse berichtet werden.

B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Tabelle 8.1: Verteilung der Kinder mit und ohne Migrationshintergrund auf unterschiedliche Kompetenzstufen der Mathematik (Angaben in Prozent)

Teilnehmer ^A	Die Testsprache wird manchmal oder nie zu Hause gesprochen					Die Testsprache wird immer oder fast immer zu Hause gesprochen				
	Kompetenzstufe					Kompetenzstufe				
	I	II	III	IV	V	I	II	III	IV	V
	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)
² Singapur	1.1(0.3)	5.7(0.7)	17.7(1.2)	36.2(1.5)	39.3(2.2)	0.9(0.2)	3.1(0.6)	11.9(1.4)	33.0(2.0)	51.1(2.4)
Republik Korea (Südkorea)	0.3(0.3)	1.6(0.6)	13.3(1.9)	40.6(1.9)	44.2(1.7)	0.4(0.2)	3.5(0.4)	17.9(0.9)	40.7(1.4)	37.6(1.6)
² Hongkong	1.9(1.4)	4.6(1.3)	16.4(1.6)	42.0(2.0)	35.0(2.6)	0.2(0.1)	2.1(0.5)	14.7(1.2)	43.4(1.4)	39.6(1.9)
Taiwan	1.0(0.3)	5.1(0.7)	20.5(1.4)	41.1(1.3)	32.3(1.5)	0.9(0.4)	6.3(0.7)	18.6(1.2)	39.1(1.4)	35.1(1.4)
Japan	1.0(0.5)	4.5(0.9)	19.5(2.3)	41.5(2.5)	33.5(2.5)	0.9(0.2)	6.0(0.6)	23.3(0.9)	40.7(1.1)	29.1(1.1)
³ Nordirland	6.4(2.1)	14.2(2.2)	29.0(4.1)	31.8(3.0)	18.6(2.9)	3.6(0.5)	10.4(1.1)	25.9(1.2)	35.3(1.0)	24.7(1.3)
³ Belgien (Fläm. Gem.)	1.4(0.7)	16.4(1.8)	46.6(2.1)	30.5(1.9)	5.0(0.9)	0.4(0.2)	8.0(0.7)	36.5(1.2)	43.2(1.0)	11.9(0.8)
Finnland	4.6(1.5)	19.7(3.2)	37.0(3.2)	31.6(3.1)	7.1(2.2)	1.8(0.4)	12.5(0.9)	34.9(1.2)	38.5(1.4)	12.4(0.8)
¹ England	8.8(1.5)	18.8(2.9)	30.0(2.5)	29.5(3.0)	12.9(2.3)	6.1(0.7)	15.0(1.3)	27.8(1.3)	31.8(1.7)	19.3(1.5)
² Russische Föderation	2.5(0.8)	14.7(2.9)	34.1(2.7)	32.7(2.8)	15.9(2.8)	3.2(0.6)	15.1(1.1)	35.2(1.2)	34.1(1.1)	12.3(1.3)
² USA	7.4(1.0)	24.0(1.6)	37.9(1.7)	23.8(1.6)	6.9(1.2)	3.2(0.4)	13.7(0.6)	33.5(1.0)	35.8(0.9)	13.9(0.8)
³ Niederlande	2.0(0.9)	18.0(2.6)	49.3(2.3)	28.9(2.3)	1.8(0.6)	0.3(0.1)	9.1(0.8)	42.6(1.4)	42.6(1.3)	5.3(0.8)
² Dänemark	6.7(1.7)	19.7(1.7)	40.4(2.5)	28.3(2.5)	4.9(1.2)	2.2(0.4)	13.0(0.9)	37.1(1.4)	36.7(1.6)	11.1(1.2)
² Litauen	5.9(1.7)	20.4(2.3)	39.3(2.2)	27.7(2.3)	6.8(1.0)	4.0(0.7)	15.9(1.1)	35.4(1.4)	33.7(1.3)	11.1(0.9)
Portugal	4.4(1.7)	19.1(3.6)	41.8(5.0)	28.8(3.8)	5.9(2.9)	3.1(0.6)	15.8(1.4)	40.1(1.3)	32.2(1.7)	8.8(1.2)
Deutschland	4.2(1.7)	28.4(2.7)	44.3(3.0)	21.0(1.8)	2.2(0.7)	2.2(0.6)	12.5(0.9)	42.5(1.4)	36.2(1.2)	6.4(0.6)
Irland	8.7(1.7)	18.3(2.5)	38.6(3.2)	27.5(2.3)	6.8(1.7)	5.3(0.6)	17.1(1.2)	35.2(1.6)	32.4(1.5)	10.0(1.0)
VG OECD	11.1(0.4)	23.7(0.5)	35.4(0.6)	22.5(0.5)	7.3(0.3)	6.4(0.1)	18.2(0.2)	36.1(0.3)	29.4(0.3)	9.9(0.2)
VG EU	10.2(0.5)	24.1(0.6)	38.3(0.7)	22.6(0.6)	4.9(0.3)	6.2(0.2)	18.2(0.3)	37.5(0.3)	29.8(0.3)	8.3(0.2)
Australien	12.3(2.1)	21.8(2.8)	35.0(2.5)	21.5(1.6)	9.4(1.9)	8.6(0.8)	19.6(1.0)	35.2(1.1)	26.6(1.1)	10.1(1.0)
Ungarn	32.1(5.3)	26.4(4.8)	27.3(5.5)	12.5(3.4)	1.8(1.0)	9.7(0.9)	18.9(1.0)	33.4(1.0)	28.2(1.0)	9.8(0.8)
² Tschechische Republik	7.1(1.5)	23.6(2.9)	45.0(3.9)	21.8(3.1)	2.5(1.0)	6.6(0.8)	21.0(1.1)	41.6(1.5)	26.4(1.5)	4.5(0.6)
² Österreich	8.4(1.8)	34.7(1.9)	41.6(2.3)	14.4(1.9)	0.9(0.4)	3.5(0.6)	21.7(1.8)	44.9(1.4)	27.0(1.5)	2.8(0.4)
³ Italien	11.1(1.5)	31.6(2.1)	38.9(2.0)	16.5(1.6)	1.9(0.7)	6.0(0.7)	22.0(1.0)	41.0(1.4)	25.6(1.3)	5.4(0.7)
Slowakei	14.8(3.1)	24.9(2.3)	37.9(2.4)	19.1(2.6)	3.2(1.1)	7.7(0.7)	20.8(1.0)	38.7(1.1)	27.0(1.3)	5.8(0.8)
Schweden	11.6(1.6)	30.8(2.0)	39.7(1.8)	16.0(1.8)	2.0(0.7)	5.3(0.6)	22.9(1.3)	45.1(1.7)	23.5(1.3)	3.2(0.5)
¹ Malta	11.3(0.7)	24.8(0.9)	39.0(1.2)	21.3(0.9)	3.6(0.3)	10.7(1.5)	25.4(2.9)	37.9(2.6)	22.0(1.8)	4.0(1.0)
³ Norwegen	13.6(2.4)	33.8(3.5)	37.5(3.4)	13.8(2.4)	1.2(0.6)	7.8(1.0)	27.1(1.5)	42.7(1.8)	19.8(1.4)	2.6(0.5)
Internationaler Mittelwert	20.6(0.3)	23.4(0.3)	29.7(0.4)	19.2(0.3)	7.1(0.2)	18.0(0.2)	20.1(0.2)	29.7(0.2)	23.1(0.2)	9.2(0.1)
¹ Neuseeland	21.2(2.0)	30.8(1.9)	30.2(1.9)	14.5(1.3)	3.3(0.6)	12.5(0.8)	25.3(1.1)	37.2(1.0)	21.2(1.1)	3.8(0.6)
² Spanien	14.6(1.8)	33.7(1.8)	37.4(2.3)	13.3(1.3)	1.0(0.4)	11.1(1.3)	29.8(1.6)	40.7(1.5)	16.9(1.1)	1.5(0.3)
Rumänien	26.9(4.7)	24.1(4.6)	26.2(3.9)	17.2(3.5)	5.6(1.6)	20.1(1.9)	21.4(1.0)	29.5(1.4)	22.1(1.4)	6.8(0.6)
Polen	12.6(1.9)	34.3(2.8)	36.3(2.8)	14.1(2.5)	2.7(1.1)	13.4(1.0)	30.4(1.0)	39.3(1.0)	15.2(1.0)	1.7(0.3)
Türkei	34.0(3.1)	26.8(1.7)	23.9(1.9)	12.8(1.4)	2.6(0.6)	17.3(1.1)	26.4(1.2)	32.2(1.0)	19.3(1.0)	4.8(0.6)
Chile	28.3(2.3)	35.9(2.3)	25.6(1.8)	9.0(1.3)	1.3(0.4)	19.9(1.2)	32.8(1.0)	32.3(1.1)	13.0(0.6)	2.0(0.3)

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

A = Da in Slowenien die Testsprache nicht erhoben wurde, können hier für diesen Teilnehmer keine Ergebnisse berichtet werden.

Republik Korea (Südkorea) ergeben sich Leistungsvorsprünge in Mathematik zugunsten der Schülerinnen und Schüler, die zu Hause nur manchmal oder nie die Testsprache sprechen. Zur Interpretation dieser Befunde müssen im internationalen Vergleich sicherlich Unterschiede in der Zusammensetzung der Bevölkerungen und in den Einwanderungsgeschichten sowie die jeweils unterschiedlich ausgeprägten sprachlichen Barrieren für Migrantinnen und Migranten in verschiedenen Staaten ebenso berücksichtigt werden wie staatspezifische Integrationsprogramme.

Ebenso wie für Mathematik zeigt sich bei den naturwissenschaftlichen Leistungen in fast allen Teilnehmerstaaten ein Leistungsvorsprung der Schülerinnen und Schüler, die zu Hause immer oder fast immer die Testsprache sprechen, vor Schülerinnen und Schülern, auf die dies nicht zutrifft (vgl. Abbildung 8.2). Die Differenzen fallen jedoch in den meisten Staaten insgesamt höher aus als für Mathematik. Dies wird auch im Vergleich der mittleren internationalen Differenz deutlich (Mathematik: 13 Punkte; Naturwissenschaften: 22 Punkte). In Deutschland beträgt die Differenz 48 Punkte. Signifikant über der für Deutschland ermittelten Differenz liegt erneut lediglich Ungarn mit einer Differenz von 89 Punkten zwischen den naturwissenschaftlichen Leistungen der beiden Schülergruppen. Im Vergleich der hier betrachteten 31 Teilnehmer unterscheidet sich Deutschland hinsichtlich der Differenz nicht signifikant von sieben anderen teilnehmenden EU-Staaten sowie von den USA, der Türkei und Neuseeland, liegt aber signifikant über dem Mittelwert der Vergleichsgruppe EU (35 Punkte) und über dem Mittelwert der Vergleichsgruppe OECD (32 Punkte). Der internationale Mittelwert für diese Differenz beträgt 22 Punkte und liegt damit ebenfalls signifikant unter dem Wert für Deutschland. Im internationalen Vergleich zeigen sich für die Russische Föderation, Polen, Taiwan und Japan keine signifikanten Differenzen, während sich für die Republik Korea (Südkorea) ein Leistungsvorsprung zugunsten der Kinder ergibt, die nur manchmal oder nie die Testsprache sprechen.

Tabelle 8.1 zeigt die Anteile von Schülerinnen und Schülern auf den fünf Kompetenzstufen für die Gesamtskala Mathematik (siehe Kapitel 3) für die Gruppe der Kinder, die angibt zu Hause immer oder fast immer die Testsprache zu sprechen im Vergleich zu den Schülerinnen und Schülern, die angeben zu Hause nie oder nur manchmal die Testsprache zu sprechen.

Für Deutschland ergibt sich, dass 33 Prozent der Schülerinnen und Schüler, die manchmal oder nie zu Hause die Testsprache sprechen, nicht die Kompetenzstufe II erreichen, während dieser Anteil bei Kindern, die zu Hause immer oder fast immer die Testsprache sprechen, weniger als halb so groß ist (15%). Entsprechend dazu sind in Deutschland die Anteile auf den Kompetenzstufen III und IV für Kinder, die zu Hause immer oder fast immer Deutsch sprechen (79%), höher als für Kinder, bei denen manchmal oder nie die Testsprache zu Hause gesprochen wird (65%). Hingegen zeigt sich, dass lediglich 2 Prozent der Schülerinnen und Schüler, die manchmal oder nie die Testsprache in ihren Familien sprechen, die Kompetenzstufe V erreichen, während es bei den Kindern, die immer oder fast immer zu Hause Deutsch sprechen, 6 Prozent sind.

Für die Vergleichsgruppen EU (VG_{EU}) und OECD (VG_{OECD}) zeigt sich, dass sich die Leistungen von 34 beziehungsweise 35 Prozent der Kinder, die die Testsprache manchmal oder nie zu Hause sprechen, den Kompetenzstufen I und II zuordnen lassen, während dies auf 24 beziehungsweise 25 Prozent der Schülerinnen und Schüler zutrifft, die die Testsprache immer oder fast immer zu Hause sprechen. Im internationalen Durchschnitt erreichen 44 Prozent der Kinder, die nie oder fast nie die Testsprache zu Hause sprechen, nicht das Niveau der Kompetenzstufe III, während es für die andere Schülergruppe 38 Prozent sind. Für die Vergleichsgruppe OECD beträgt der Anteil der Kinder, bei denen die Testsprache manchmal oder nie zu Hause gesprochen wird, auf den Kompetenzstufen III und IV 58 Prozent, für die Vergleichsgruppe EU 61 Prozent und im internationalen Vergleich 49 Prozent. Wird die Testsprache immer oder fast immer in den Familien gesprochen, beträgt der Anteil von Schülerinnen und

Schülern auf den Kompetenzstufen III und IV für die Vergleichsgruppe OECD hingegen 66 Prozent, für die Vergleichsgruppe EU 67 Prozent und im internationalen Mittel 53 Prozent.

Die Verteilung der Kinder mit unterschiedlichem familiären Gebrauch der Testsprache auf die verschiedenen Kompetenzstufen für die Gesamtskala Naturwissenschaften (siehe Kapitel 4) sieht in Deutschland folgendermaßen aus (vgl. Tabelle 8.2): 41 Prozent der Schülerinnen und Schüler, die manchmal oder nie zu Hause die Testsprache sprechen, erreichen nicht das Niveau der Kompetenzstufe III, während dies auf 16 Prozent aller Schülerinnen und Schüler zutrifft, bei denen zu Hause immer oder fast immer die Testsprache gesprochen wird.

Die Kompetenzstufen III und IV erreichen 57 Prozent aller Viertklässlerinnen und Viertklässler in Deutschland, die manchmal oder nie Deutsch zu Hause sprechen, und 75 Prozent der Schülerinnen und Schüler der vierten Jahrgangsstufe, die immer oder fast immer Deutsch sprechen.

Die Kompetenzstufe V erreichen lediglich 3 Prozent der Schülerinnen und Schüler, die manchmal oder nie die Testsprache in ihren Familien sprechen. Der Anteil der Kinder, die Kompetenzstufe V erreichen, ist dreimal so groß bei Kindern, die zu Hause immer oder fast immer Deutsch sprechen (9%).

Im internationalen Vergleich zeigt sich, dass der Anteil der Schülerinnen und Schüler, die die Testsprache manchmal oder nie zu Hause sprechen und die auf dem Niveau der Kompetenzstufen I und II liegen, für die Vergleichsgruppe OECD (VG_{OECD}) 37 Prozent, für die Vergleichsgruppe EU (VG_{EU}) 38 Prozent und im internationalen Mittel 46 Prozent beträgt. Wird hingegen die Testsprache immer oder fast immer zu Hause gesprochen, beträgt der entsprechende Anteil für die VG_{OECD} ebenso wie für die VG_{EU} lediglich 22 Prozent und im internationalen Vergleich 36 Prozent. Der Anteil auf den Kompetenzstufen III und IV beträgt bei Kindern, bei denen die Testsprache manchmal oder nie zu Hause gesprochen wird, für die Vergleichsgruppen OECD und EU 58 Prozent und im internationalen Vergleich 49 Prozent.

Wird die Testsprache immer oder fast immer in den Familien gesprochen, sind 69 Prozent der Viertklässlerinnen und Viertklässler aus der Vergleichsgruppe OECD, ebenso 69 Prozent aus der Vergleichsgruppe EU und 56 Prozent im internationalen Mittel auf den Kompetenzstufen III und IV zu verorten.

Insgesamt wird deutlich, dass fast jedes dritte Kind in Deutschland, bei dem zu Hause manchmal oder nie Deutsch gesprochen wird, in Mathematik nicht die Kompetenzstufe III erreicht. Wird hingegen zu Hause die Testsprache gesprochen, erzielen lediglich ein Sechstel der Kinder keine ausreichenden Leistungen.

Auch in den Naturwissenschaften ist der Anteil der Kinder mit Migrationshintergrund auf den unteren beiden Kompetenzstufen mehr als doppelt so hoch im Vergleich zu den Kindern ohne Migrationshintergrund (vgl. Tabelle 8.2).

Tabelle 8.2: Verteilung der Kinder mit und ohne Migrationshintergrund auf unterschiedliche Kompetenzstufen der Naturwissenschaften (Angaben in Prozent)

Teilnehmer ^A	Die Testsprache wird manchmal oder nie zu Hause gesprochen					Die Testsprache wird immer oder fast immer zu Hause gesprochen				
	Kompetenzstufe					Kompetenzstufe				
	I % (SE)	II % (SE)	III % (SE)	IV % (SE)	V % (SE)	I % (SE)	II % (SE)	III % (SE)	IV % (SE)	V % (SE)
Republik Korea (Südkorea)	0.3(0.2)	2.6(0.7)	18.9(1.7)	46.6(2.5)	31.7(2.1)	0.7(0.2)	4.8(0.4)	23.8(1.0)	43.2(1.2)	27.5(1.5)
² <i>Singapur</i>	3.6(0.5)	9.4(0.9)	24.1(1.5)	35.0(1.4)	27.9(1.7)	1.9(0.4)	4.3(0.8)	14.8(1.3)	34.4(1.7)	44.7(2.3)
Finnland	5.0(1.3)	15.2(2.8)	31.3(3.3)	38.1(3.7)	10.4(3.9)	0.8(0.3)	5.6(0.7)	26.9(1.5)	45.2(1.4)	21.5(1.2)
Japan	0.9(0.6)	7.5(1.4)	30.5(2.2)	45.4(2.2)	15.7(1.9)	1.4(0.2)	8.5(0.7)	33.2(1.5)	42.9(1.2)	14.1(1.2)
² <i>Russische Föderation</i>	2.3(1.2)	15.8(2.9)	32.9(3.4)	33.5(3.2)	15.5(2.7)	2.0(0.4)	11.6(0.9)	33.7(1.4)	36.8(1.3)	15.9(1.4)
Taiwan	2.5(0.4)	11.2(1.1)	33.8(1.2)	38.4(1.5)	14.1(1.2)	3.5(0.6)	11.9(1.1)	29.8(1.3)	38.3(1.5)	16.5(1.1)
^{2,3} <i>USA</i>	10.3(1.0)	26.2(1.5)	37.8(1.9)	20.1(1.8)	5.5(1.0)	3.4(0.4)	12.1(0.6)	31.2(0.7)	36.9(0.9)	16.4(0.8)
² <i>Tschechische Republik</i>	4.1(1.3)	20.5(3.0)	39.0(3.2)	30.0(2.8)	6.4(1.7)	3.3(0.7)	14.6(1.0)	36.3(1.4)	35.1(1.3)	10.7(1.0)
² <i>Hongkong</i>	7.5(2.8)	14.9(1.6)	36.2(2.1)	32.8(2.1)	8.7(1.4)	2.7(0.5)	12.9(1.1)	37.2(1.7)	37.5(2.0)	9.6(1.2)
Ungarn	31.0(5.5)	24.3(4.6)	28.5(5.0)	14.4(3.6)	1.8(1.1)	6.6(0.9)	14.4(1.1)	31.7(1.2)	33.7(1.6)	13.6(0.9)
Schweden	12.1(1.8)	25.8(2.0)	35.2(2.4)	22.4(2.5)	4.5(1.0)	2.7(0.4)	13.5(0.9)	35.5(1.3)	36.6(1.2)	11.7(1.1)
Slowakei	11.2(2.1)	19.4(2.9)	36.8(2.8)	26.7(2.4)	5.8(1.4)	4.0(0.7)	14.1(1.1)	34.5(1.4)	36.5(1.0)	10.9(1.2)
² <i>Österreich</i>	10.5(1.7)	29.5(2.3)	38.4(2.4)	19.2(2.0)	2.4(0.8)	2.0(0.4)	13.0(1.4)	36.9(1.3)	38.0(1.8)	10.2(1.0)
³ <i>Niederlande</i>	3.2(1.3)	24.0(2.6)	50.0(2.9)	21.6(2.2)	1.3(0.6)	0.7(0.2)	10.0(1.4)	48.0(1.2)	37.9(1.7)	3.3(0.5)
^{1,3} <i>England</i>	11.2(1.4)	23.4(2.5)	36.5(2.7)	24.3(2.7)	4.5(1.4)	5.7(0.8)	16.3(0.9)	32.4(1.6)	32.9(1.5)	12.7(1.0)
^{2,3} <i>Dänemark</i>	11.8(2.3)	24.9(2.1)	38.2(3.6)	20.9(2.6)	4.2(1.5)	3.0(0.5)	14.7(1.2)	38.9(1.2)	34.3(1.3)	9.0(0.9)
Deutschland	8.7(1.6)	31.9(3.0)	39.6(2.8)	17.3(1.6)	2.6(0.7)	2.8(0.5)	13.3(1.1)	37.7(1.5)	37.2(1.6)	9.0(0.8)
³ <i>Italien</i>	9.4(1.9)	28.0(2.1)	39.7(2.9)	20.4(2.4)	2.6(0.9)	4.0(0.8)	16.3(1.3)	38.9(1.6)	31.6(1.5)	9.3(0.8)
VG OECD	12.0(0.4)	24.9(0.5)	35.8(0.6)	22.0(0.5)	5.3(0.3)	5.2(0.1)	16.6(0.2)	36.9(0.3)	31.9(0.3)	9.4(0.2)
Portugal	6.3(2.4)	24.5(4.0)	41.4(2.9)	23.8(4.0)	3.9(1.7)	5.3(1.1)	18.6(1.4)	39.7(1.2)	28.9(1.4)	7.6(1.1)
VG EU	12.1(0.5)	25.9(0.6)	37.3(0.7)	20.8(0.6)	4.0(0.3)	5.4(0.2)	16.5(0.3)	37.4(0.3)	31.9(0.3)	8.8(0.2)
³ <i>Nordirland</i>	10.4(3.0)	23.9(3.7)	39.3(5.0)	22.3(3.5)	4.0(1.7)	5.6(0.8)	18.9(1.1)	40.9(1.4)	29.4(1.6)	5.2(0.7)
Irland	12.9(1.8)	25.6(2.6)	35.7(3.2)	21.0(2.7)	4.7(1.2)	6.6(0.9)	19.4(1.5)	36.7(1.6)	29.6(1.6)	7.7(1.0)
Australien	12.2(2.1)	24.4(2.6)	37.1(2.8)	21.1(1.7)	5.2(1.3)	7.4(0.9)	18.3(1.0)	36.2(1.3)	30.0(1.2)	8.1(0.8)
² <i>Litauen</i>	6.0(1.6)	28.1(2.4)	43.7(2.5)	19.8(2.4)	2.5(0.7)	5.2(0.7)	19.9(1.1)	42.0(1.3)	28.2(1.5)	4.7(0.6)
³ <i>Belgien (Fläm. Gem.)</i>	9.6(1.3)	39.5(2.8)	42.0(2.3)	8.6(1.3)	0.3(0.2)	1.5(0.4)	18.5(1.2)	50.9(1.4)	27.1(1.2)	2.0(0.5)
Rumänien	25.8(4.3)	22.8(3.0)	25.8(4.8)	18.4(3.5)	7.2(1.9)	14.0(1.7)	17.4(1.4)	29.3(1.6)	27.3(1.9)	11.9(1.0)
² <i>Spanien</i>	11.5(1.7)	29.1(2.0)	39.6(2.1)	17.1(2.0)	2.7(0.9)	6.2(0.9)	21.6(1.3)	40.0(1.7)	27.0(1.5)	5.2(0.7)
Polen	10.0(2.1)	26.1(2.5)	38.3(2.8)	20.0(2.0)	5.8(2.0)	9.3(0.8)	23.4(1.0)	38.2(1.2)	23.8(1.2)	5.4(0.6)
^{1,3} <i>Neuseeland</i>	21.4(2.0)	29.6(1.6)	31.0(2.6)	15.4(1.6)	2.6(0.5)	10.5(0.7)	21.5(0.9)	35.7(1.1)	26.2(1.1)	6.2(0.6)
³ <i>Norwegen</i>	15.6(2.3)	36.4(2.5)	36.7(2.3)	10.9(2.2)	0.5(0.3)	5.8(0.7)	26.3(1.6)	47.2(1.6)	19.3(1.4)	1.4(0.3)
Internationaler Mittelwert	22.1(0.3)	23.7(0.3)	30.7(0.4)	18.6(0.3)	4.9(0.2)	17.5(0.2)	18.5(0.2)	31.1(0.2)	24.9(0.2)	8.0(0.1)
Chile	21.8(2.5)	33.5(2.8)	30.6(2.1)	12.6(1.4)	1.5(0.4)	12.8(0.9)	29.6(1.2)	36.9(1.1)	18.0(0.9)	2.8(0.4)
Türkei	38.4(2.9)	27.2(2.0)	22.6(1.7)	9.8(1.1)	2.0(0.4)	17.8(1.1)	28.1(1.0)	33.2(1.1)	17.1(1.1)	3.9(0.5)
¹ <i>Malta</i>	31.4(1.3)	30.6(1.3)	26.2(0.9)	10.2(0.9)	1.5(0.3)	18.5(1.8)	25.4(2.2)	32.4(3.0)	18.3(2.3)	5.4(1.2)

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.

2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

3 = Die Teilnahmequoten auf Schul- und/oder Schülersebene erreichen nicht die internationalen Vorgaben.

A = Da in Slowenien die Testsprache nicht erhoben wurde, können hier für diesen Teilnehmer keine Ergebnisse berichtet werden.

3 Deskriptive Befunde zur Lage von Schülerinnen und Schülern mit Migrationshintergrund in Deutschland

Im Folgenden werden Ergebnisse von Analysen beschrieben, die mit dem deutschen Teildatensatz durchgeführt wurden. Zunächst soll es darum gehen, die Situation von Familien mit Migrationshintergrund anhand einiger Strukturmerkmale genauer zu beleuchten. Daran anschließend werden die im ersten Teil des Kapitels berichteten Befunde zu den Leistungsunterschieden vertiefend für Deutschland untersucht und anhand der Daten von TIMSS 2007 und 2011 miteinander verglichen.

Für die nachfolgend vorgestellten Analysen wird zur Beschreibung der Migrationsgeschichte der Familien auf die Angabe zu den Geburtsländern beider Elternteile zurückgegriffen, da hiermit zum einen ein Bezug zu anderen Untersuchungen möglich ist und zum anderen die drei unterschiedenen Gruppen gut zu interpretieren sind und auch durch hinreichend umfangreiche Teilstichproben repräsentiert werden. Im Rahmen der multivariaten Analysen, die dieses Kapitel abschließen, wird die Relation von Geburtsland und dem Sprachgebrauch in der Familie nochmals aufgegriffen.

Strukturelle Merkmale

Betrachtet man die demographische Entwicklung in Deutschland, so wird deutlich, dass sich auch die Lebensbedingungen von Grundschulkindern in Folge eines allgemeinen gesellschaftlichen Strukturwandels verändert haben (vgl. Autorengruppe Bildungsbericht, 2008). In Tabelle 8.3 werden zentrale strukturelle Charakteristika von Familien von Kindern am Ende der vierten Jahrgangsstufe in Deutschland dargestellt und nach dem Migrationsstatus unterschieden. Als zentrale Charakteristika werden die Familienstruktur (Alleinerziehend oder Zwei-Elternteil-Familie), die Anzahl der Kinder unter 14 Jahren im Haushalt, die Erwerbstätigkeit der Eltern (Vollzeit- oder Teilzeittätigkeit) sowie die Dauer des Kindergartenbesuchs betrachtet.

Der für Deutschland ermittelte Anteil von Schülerinnen und Schülern mit Migrationshintergrund (mindestens ein Elternteil im Ausland geboren) liegt bei rund 28 Prozent; der Anteil der Schülerinnen und Schüler aus mehrsprachigen Familien liegt mit knapp 19 Prozent darunter. Der Anteil an Familien, in denen Deutsch nicht in den familiären Alltag integriert ist, liegt unter 1 Prozent.

In der Gesamtstichprobe zeigt sich insgesamt und unabhängig von der Migrationsgeschichte der Familien, dass rund 80 Prozent aller Viertklässlerinnen und Viertklässler in Familien mit zwei Bezugspersonen aufwachsen. Nur für die besonders kleine Gruppe der Kinder, die angeben zu Hause nie Deutsch zu sprechen, ist dieser Anteil mit 72.1 Prozent am geringsten. Rund 10 Prozent aller Eltern erziehen ihr Kind allein, wobei aus der Tabelle zudem deutlich wird, dass es sich hierbei hauptsächlich um alleinerziehende Mütter handelt. Der Anteil von Kindern aus Familien mit einer Bezugsperson ist dabei für die Gruppe der Schülerinnen und Schüler, die nie Deutsch zu Hause sprechen (24.3%) sowie für die Gruppe der Kinder, von denen ein Elternteil im Ausland geboren wurde (14.7%), am höchsten.

Aus Tabelle 8.3 wird zudem deutlich, dass der Großteil der Schülerinnen und Schüler in Deutschland in Familien aufwächst, in denen ein bis zwei Kinder unter 14 Jahre alt sind. Großfamilien mit mehr als vier Kindern unter 14 Jahren sind vergleichsweise selten. Vergleicht man die Familiengröße nach dem Mi-

Tabelle 8.3: Merkmale von Familien. Angaben zur Verteilung von Viertklässlerinnen und Viertklässlern, Familienstruktur, Kinderanzahl und zum Erwerbsstatus von Mutter und Vater sowie zum Kindergartenbesuch nach Migrationsstatus und familiärem Sprachgebrauch (Angaben in Prozent)

	Familientyp					Anzahl der Kinder im Haushalt				Erwerbs-tätigkeit ^A		Kindergarten-besuch ^B
	Viert- klässler	zwei Bezugspersonen	eine Bezugsperson		Keine Angabe	< 14 Jahre				Vater	Mutter	
			weiblich	männlich		1	2	3	4+			
Migrationshintergrund												
beide Elternteile	16.1	89.4	8.9	1.0	0.7	30.8	42.2	17.7	9.4	81.5	62.3	67.2
ein Elternteil	11.6	84.8	14.7	0.3	0.2	33.7	45.9	14.3	6.0	83.6	69.9	68.2
kein Elternteil	72.3	89.6	9.1	0.8	0.5	37.0	45.3	13.5	4.2	90.2	77.3	78.0
Familiärer Sprachgebrauch ist Deutsch												
nie	0.8	72.1	24.3	0.0	3.6	39.8	41.2	15.8	3.2	86.3	68.2	52.9
manchmal	18.9	88.7	9.4	0.7	1.2	33.9	43.5	15.5	7.1	82.0	65.9	68.0
immer oder fast immer	80.4	87.3	10.5	0.9	1.3	36.5	44.8	14.0	4.7	89.0	76.3	77.0

A = Voll- oder Teilzeit

B = mindestens 3 Jahre

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

grationsstatus, ist für Familien mit Migrationsgeschichte ein höherer Anteil an Familien mit drei beziehungsweise vier und mehr Kindern festzustellen.

In Bezug auf die Erwerbstätigkeit der Bezugspersonen ist zunächst unabhängig von der Migrationsgeschichte der Familien in der Stichprobe festzustellen, dass fast zwei Drittel aller Viertklässlerinnen und Viertklässler in Deutschland in Familien aufwachsen in denen beide Elternteile zumindest einer Teilzeitbeschäftigung nachgehen. Hier zeigt sich, dass der Anteil an Vätern, die mindestens einer Teilzeittätigkeit nachgehen, mit etwa 80 bis 90 Prozent deutlich höher ist als der Anteil der teil- oder vollzeitbeschäftigten Mütter (etwa 60% bis 80%). Im Vergleich der Familien mit und ohne Migrationsgeschichte zeigt sich, dass die Anteile an erwerbstätigen Müttern und Vätern in Familien mit Migrationshintergrund geringer sind als in Familien ohne Migrationshintergrund. Inwieweit diese Unterschiede gewollten Familienmodellen entsprechen oder Folge unterschiedlicher Erwerbslagen sind, lässt sich hier nicht klären. Vergleicht man die bis hier skizzierte Situation der Familien mit Befunden aus IGLU 2001 (Schwippert, Bos & Lankes, 2003), so zeigt sich, dass die Unterschiede zwischen den Familien mit und ohne Migrationshintergrund in Bezug auf die elterliche Betreuungssituation, die Kinderzahl und den Erwerbsstatus kleiner geworden sind.

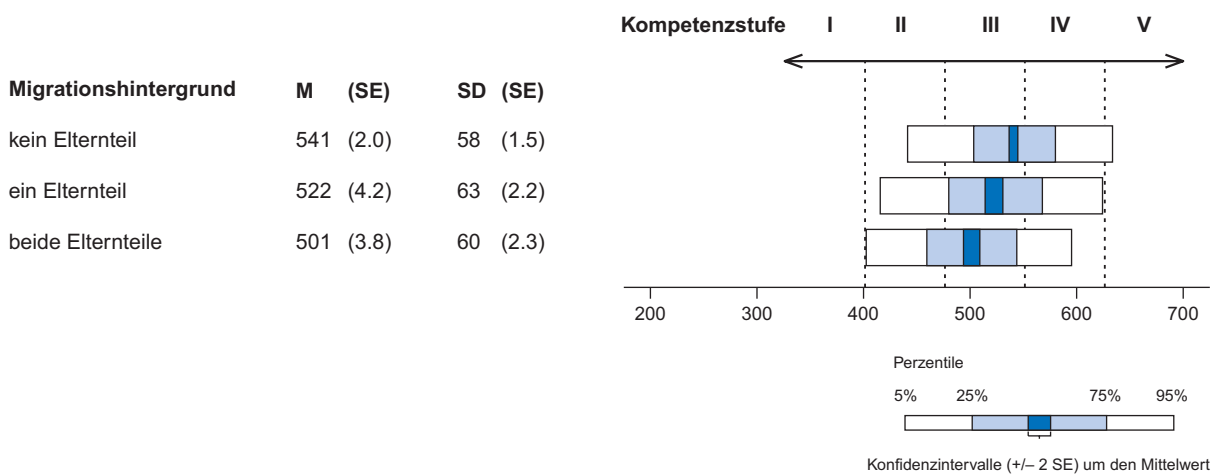
Um Disparitäten im Sprachgebrauch von Kindern zu begegnen, wird oft auf die Bedeutsamkeit einer frühen sprachlichen Förderung verwiesen (Reich & Roth, 2007; Valtin, Bos, Hornberg & Schwippert, 2007). Gerade für Kinder aus Familien, in denen selten oder nie Deutsch gesprochen wird, werden hier positive Effekte erwartet. Der Tabelle 8.3 ist zu entnehmen, dass der Anteil von Kindern, die drei oder mehr Jahre einen Kindergarten besuchen, bei den Familien ohne Migrationsgeschichte mit rund 80 Prozent im Vergleich zu den anderen Familien um rund 10 Prozentpunkte deutlich höher ausfällt. Wenn sich eine frühe Sprachförderung auch empirisch als wirksam bewährt, könnten in Familien, in denen nur manchmal oder nie Deutsch gesprochen wird, durch den einen längeren Zeitraum umfassenden Besuch eines Kindergartens entsprechende Defizite kompensiert werden.

Betrachtet man den Migrationsstatus einerseits anhand des Geburtslandes der Eltern und andererseits anhand des familiären Sprachgebrauchs, ergeben sich zwar Unterschiede, es zeigen sich jedoch dieselben Tendenzen, so dass beide Ansätze zur Unterscheidung von Familien mit und ohne Migrationsgeschichte in einem ersten Zugriff gerechtfertigt erscheinen. Allerdings wäre für Schülerinnen und Schüler im Rahmen von vertiefenden Analysen im Hinblick auf deren schulische Entwicklung die Bedeutsamkeit verschiedener Indikatorisierungen zu untersuchen. Zudem wären Wirkungszusammenhänge zu prüfen. Zu beachten ist darüber hinaus, dass der besonders niedrige Anteil von Kindern, die angeben zu Hause nie Deutsch zu sprechen (unter 1%), keine verallgemeinerbaren Aussagen über diese Personengruppe zulässt.

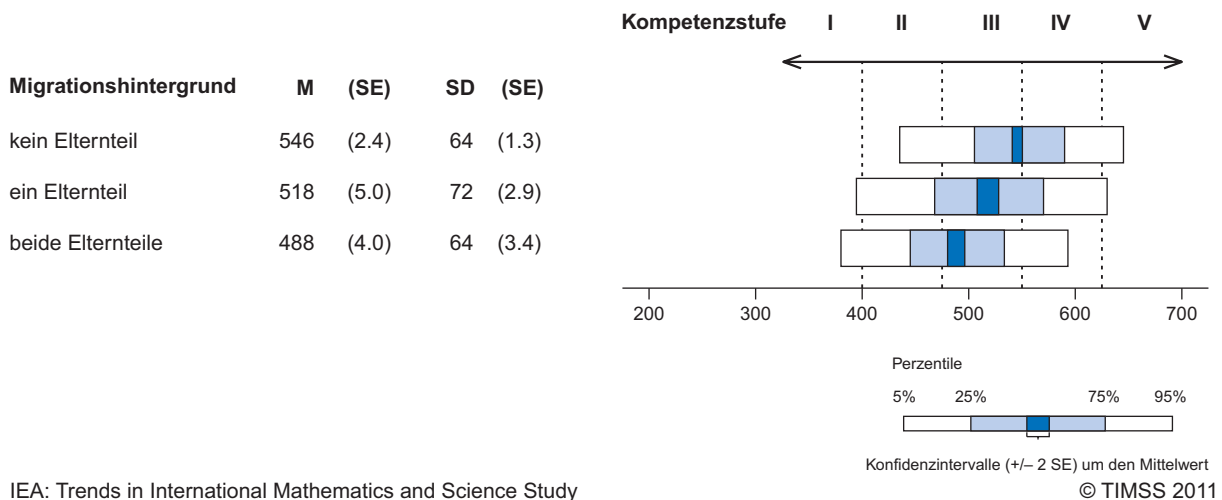
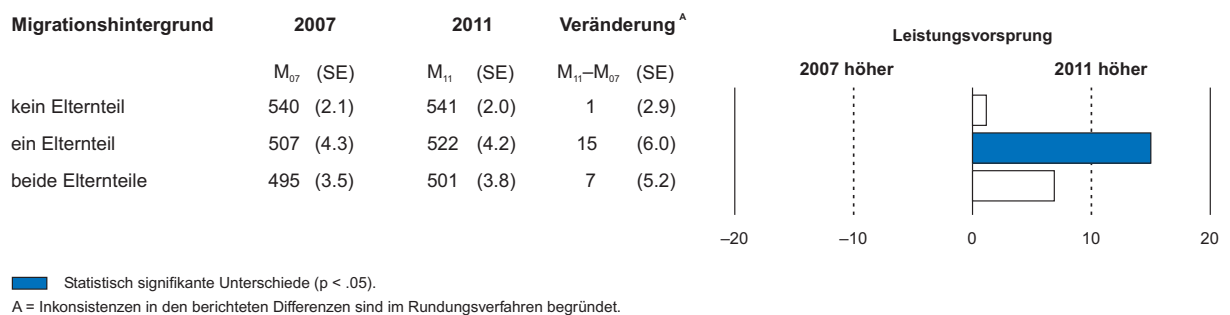
4 Kompetenzunterschiede in Deutschland

In Abbildung 8.3 sind die mathematischen und in Abbildung 8.4 die naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern aus Familien mit unterschiedlicher Migrationsgeschichte für TIMSS 2011 dargestellt. Dabei wird zwischen drei Gruppen von Kindern unterschieden: (1) Kinder, deren Eltern in Deutschland geboren wurden, (2) Kinder, bei denen ein Elternteil im Ausland geboren wurde und (3) Kinder, deren beide Elternteile im Ausland geboren wurden.

Abbildung 8.3: Mathematische Kompetenzen von Schülerinnen und Schülern mit unterschiedlichem Migrationshintergrund in Deutschland im Vergleich



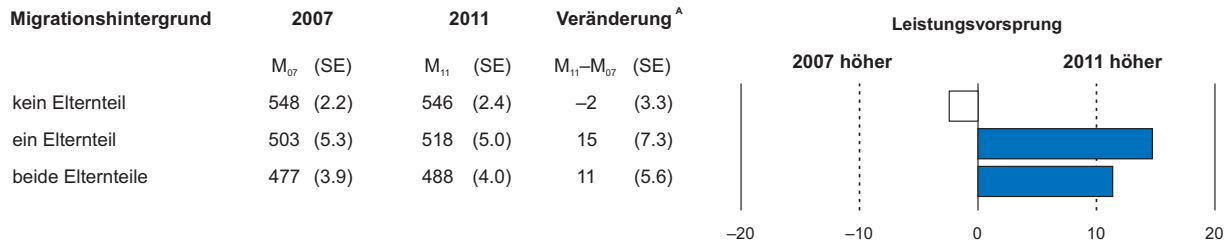
Sowohl für Mathematik als auch für die Naturwissenschaften lässt sich festhalten, dass die Schülerinnen und Schüler, deren Eltern in Deutschland geboren wurden, die höchsten Leistungswerte erreichen. Signifikant niedriger liegen die erreichten Kompetenzen der Kinder, die angeben, dass ein Elternteil oder dass beide Elternteile im Ausland geboren wurden. Vergleicht man die Kompetenzen von Schülerinnen und Schülern, die angeben, dass ein Elternteil im Ausland geboren wurde, mit denen der Schülerinnen und Schüler, die angeben, dass beide Elternteile im Ausland geboren wurden, so zeigen sich für letztere wiederum weitere signifikante Rückstände sowohl im Bereich mathematischer als auch im Bereich naturwissenschaftlicher Kompetenzen.

Abbildung 8.4: Naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit unterschiedlichem Migrationshintergrund in Deutschland im Vergleich

Abbildung 8.5: Unterschiede im Leistungsvorsprung in mathematischen Kompetenzen von Kindern mit unterschiedlichem Migrationshintergrund (nach Geburtsland der Eltern) in Deutschland – TIMSS 2007 und 2011 im Vergleich


Stellt man die Befunde für die mathematischen Kompetenzen von Kindern mit unterschiedlichem Migrationshintergrund einander gegenüber, die in 2007 und in 2011 im Rahmen von TIMSS ermittelt wurden (vgl. Abbildung 8.5), so zeigen sich 2011 signifikant höhere Kompetenzen als 2007 für die Schülerinnen und Schüler, die angeben, dass ein Elternteil im Ausland geboren wurde (2007: 507 Punkte; 2011: 522 Punkte). Für die Schülerinnen und Schüler, bei denen beide Elternteile im Ausland oder beide Elternteile in Deutschland geboren wurden, zeigen sich hingegen keine signifikanten Unterschiede zwischen den beiden Erhebungszyklen.

Betrachtet man die in den Naturwissenschaften erreichten Leistungen jeweils im Vergleich der Jahre 2007 und 2011 (Abbildung 8.6), so zeigen sich für die Kinder, deren Eltern im Ausland geboren wurden, und für Kinder, die angeben, dass ein Elternteil im Ausland geboren wurde, für 2011 jeweils signifikant höhere Kompetenzen als 2007: Kinder aus Familien mit einem im Ausland geborenen Elternteil erzielen 2011 in den Naturwissenschaften – ebenso wie in Mathematik – einen um 15 Punkte höheren Wert als 2007. Bei den Viertklässlerinnen und Viertklässlern, bei denen die Eltern im Ausland geboren wurden, beträgt der Unterschied zwischen den 2007 und den 2011 erreichten mittleren Leistungen 11 Punkte auf der Skala der naturwissenschaftlichen Kompetenzen.

Abbildung 8.6: Unterschiede im Leistungsvorsprung in naturwissenschaftlichen Kompetenzen von Kindern mit unterschiedlichem Migrationshintergrund (nach Geburtsland der Eltern) in Deutschland – TIMSS 2007 und 2011 im Vergleich



■ Statistisch signifikante Unterschiede ($p < .05$).

A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

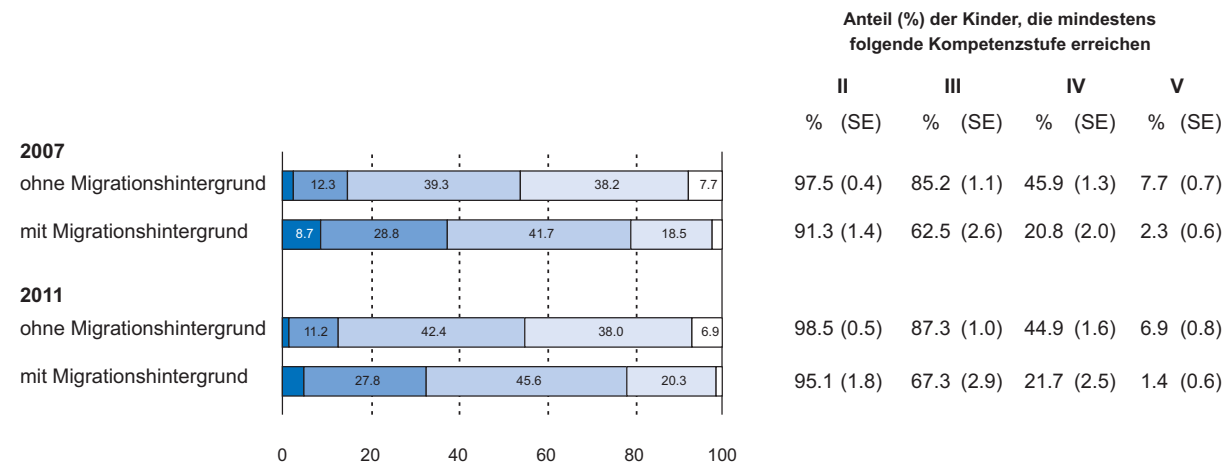
IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Diese Unterschiede können als beachtlich angesehen werden, zumal sich für die Gruppe der Kinder, deren Eltern in Deutschland geboren wurden, keine signifikanten Veränderungen ergeben.

In Abbildung 8.7 ist für 2011 und 2007 die Verteilung der Schülerinnen und Schüler, deren Eltern in Deutschland beziehungsweise im Ausland geboren wurden, auf den fünf Kompetenzstufen für Mathematik gegenübergestellt. Dabei sind die prozentualen Anteile der Kinder, die mindestens die jeweiligen Kompetenzstufen erreichen, rechts in der Tabelle angegeben.

Abbildung 8.7: Anteile von Kindern mit und ohne Migrationshintergrund in Deutschland auf den unterschiedlichen Kompetenzstufen der Mathematik – TIMSS 2007 und 2011 im Vergleich (Angaben in Prozent)



- % der Schülerinnen und Schüler, die genau Kompetenzstufe V erreichen
- % der Schülerinnen und Schüler, die genau Kompetenzstufe IV erreichen
- % der Schülerinnen und Schüler, die genau Kompetenzstufe III erreichen
- % der Schülerinnen und Schüler, die genau Kompetenzstufe II erreichen
- % der Schülerinnen und Schüler, die genau Kompetenzstufe I erreichen

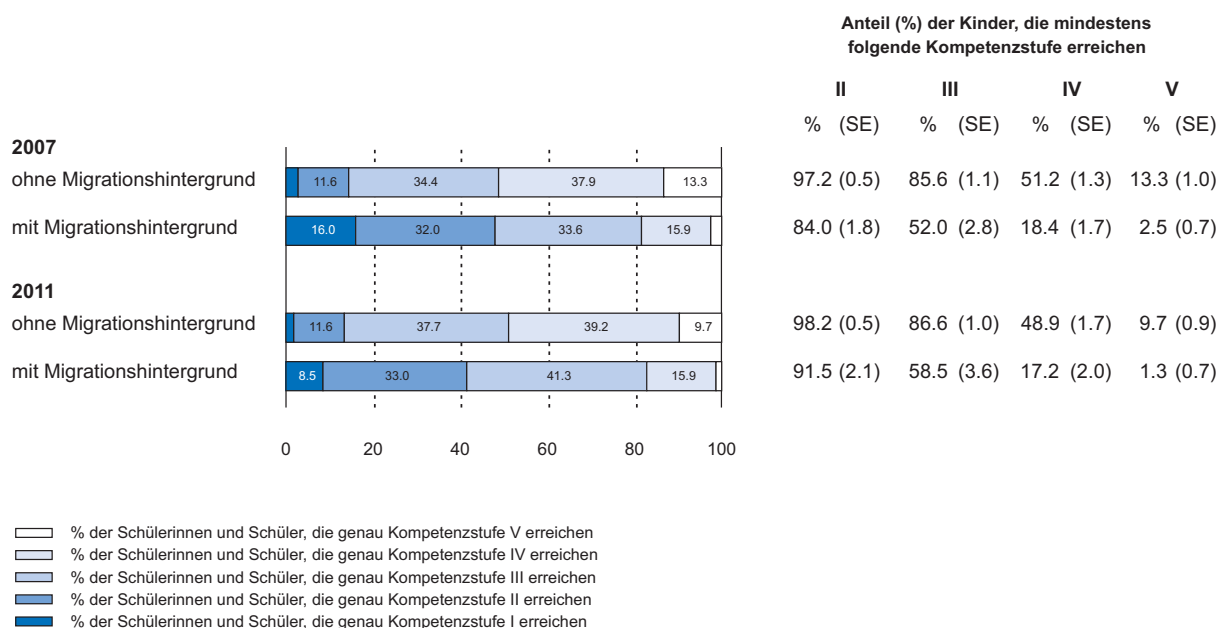
IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Im Vergleich der beiden Gruppen zeigt sich, dass in der Gruppe der Schülerinnen und Schüler, deren Eltern im Ausland geboren wurden, 2007 mit rund 38 Prozent und 2011 mit 33 Prozent – also mit rund einem Drittel – der Kinder, die in Mathematik lediglich über rudimentäres schulisches Anfangswissen oder über elementares Faktenwissen verfügen (Kompetenzstufen I und II), einen substantiellen Anteil ausmachen. Die Anteile der Kinder auf den Kompetenzstufen I und II, deren Eltern in Deutschland geboren wurden, liegen 2007 mit 15 Prozent und 2011 mit 13 Prozent darunter.

Die Anteile der Schülerinnen und Schüler, die auf durchschnittlichem mathematischen Niveau liegen (Kompetenzstufe III), unterscheiden sich weder zwischen den beiden Studienzyklen noch zwischen den betrachteten Gruppen substantiell; die entsprechenden Anteile liegen zwischen 39 und 46 Prozent. Bei Betrachtung der kumulierten Anteile wird darüber hinaus deutlich, dass 2007 91 Prozent aller Schülerinnen und Schüler mit Migrationshintergrund mindestens Kompetenzstufe II und 63 Prozent mindestens Kompetenzstufe III erreicht haben, wobei diese Anteile 2011 im Vergleich zu 2007 leicht angestiegen sind (mindestens Kompetenzstufe II: 95%; mindestens Kompetenzstufe III: 67%). Deutliche Unterschiede sind hingegen auf den höheren Kompetenzstufen (IV und V) zu beobachten. Hier sind die entsprechenden Anteile für die Kinder, deren Eltern in Deutschland geboren wurden, mit 46 und 45 Prozent (2007 und 2011) mehr als doppelt so hoch wie für die Kinder, deren Eltern im Ausland geboren wurden (2007: 21%; 2011: 22%). Im Vergleich der Erhebungen von 2007 und 2011 zeigen sich somit in Mathematik keine herausragenden Unterschiede in der Verteilung der Kinder mit unterschiedlichem Migrationshintergrund.

Abbildung 8.8: Anteile von Kindern mit und ohne Migrationshintergrund in Deutschland auf den unterschiedlichen Kompetenzstufen der Naturwissenschaften – TIMSS 2007 und 2011 im Vergleich (Angaben in Prozent)



In den Naturwissenschaften sind die Anteile von Schülerinnen und Schülern, die lediglich über rudimentäres Anfangswissen (Kompetenzstufe I) oder über elementares Faktenwissen (Kompetenzstufe II) verfügen, in der Gruppe der Kinder, deren Eltern im Ausland geboren wurden, 2007 mit 48 Prozent und 2011 mit 42 Prozent mehr als doppelt so hoch wie in der Gruppe der Kinder, deren Eltern in Deutschland geboren wurden (vgl. Abbildung 8.8). Bei dieser Gruppe liegen die entsprechenden Anteile 2007 bei 14 Prozent und 2011 bei 13 Prozent. Hierbei erscheint besonders bemerkenswert, dass der Anteil von Schülerinnen und Schülern, die lediglich die Kompetenzstufe I (rudimentäres Anfangswissen) erreichen, in der Gruppe der Kinder, deren Eltern im Ausland geboren wurden, 2007 bei 16 Prozent lag und 2011 einen deutlich niedrigeren Prozentsatz von 9 Prozent aufweist, was auch bei Betrachtung der kumulierten Anteile deutlich wird. Im Jahr 2007 lagen 84 Prozent der Schülerinnen und Schüler mit Migrationshintergrund mindestens auf dem Niveau der Kompetenzstufe II. 2011 ist dieser Anteil auf 92 Prozent angestiegen. Außerdem haben 2011 59 Prozent der Schülerinnen und Schüler mit Migrationshintergrund mindestens das Niveau der Kompetenzstufe III erreicht, was im Vergleich zu 2007 eine Erhöhung um 7 Prozentpunkte darstellt. Im Zusammenhang mit der zuvor gezeigten positiven Veränderung der Leistungen dieser Gruppe insgesamt (vgl. Abbildung 8.7) kann es als erfreulicher Befund gewertet werden, dass sich diese Veränderung insbesondere für die leistungsschwächsten Schülerinnen und Schüler ergibt. Inwieweit diese Veränderung als Trend bewertet werden kann, lässt sich allerdings erst mit dem dritten Messzeitpunkt bewerten, der im Kontext von TIMSS 2015 möglich sein wird.

Für die beiden höchsten Kompetenzstufen zeigt sich, dass die Anteile von Kindern auf den Kompetenzstufen IV (Erklären von Alltagsphänomenen) und V (beginnendes naturwissenschaftliches Denken) aus Familien, in denen die Eltern in Deutschland geboren wurden, mit 51 Prozent (2007) beziehungsweise 49 Prozent (2011) mehr als doppelt so hoch sind wie die Anteile der Kinder, deren Eltern beide im Ausland geboren wurden (2007: 18% und 2011: 17%).

5 Zusammenhänge zwischen soziokulturellen Bedingungen und Kompetenzen

Wie oben beschrieben sind die familiären Bedingungen von Schülerinnen und Schülern durch Heterogenität gekennzeichnet – eine einfache Zuordnung zu sozialen Gruppen würde die vielen unterschiedlichen Lebensverhältnisse nicht angemessen beschreiben. Verschiedene Merkmale zeigen jeweils individuell betrachtet Zusammenhänge mit den getesteten mathematischen und naturwissenschaftlichen Kompetenzen, weisen aber auch Zusammenhänge untereinander auf. So können individuell beobachtete Effekte bei umfassenden Analysen in ihren Wirkungen nicht einfach addiert werden. In multivariaten Analysen können individuelle, aber auch Interaktionseffekte zwischen erklärenden Variablen abgeschätzt werden. Zur Untersuchung des Verhältnisses von Migrationsstatus (erfasst anhand des Geburtslandes der Eltern), familiärem Sprachgebrauch, sozialem Hintergrund sowie höchstem Bildungsabschluss und höchstem Berufsstatus der Eltern sind in Tabelle 8.4 die Ergebnisse schrittweise erweiterter Regressionsanalysen für die Erhebungszeitpunkte 2007 und 2011 im Vergleich dargestellt.

Tabelle 8.4: Regressionsmodell zur Erklärung von Unterschieden in mathematischen und naturwissenschaftlichen Kompetenzen zwischen Kindern mit und ohne Migrationshintergrund (Angaben in Skalenpunkten der Gesamtskalen Mathematik und Naturwissenschaften) – Vergleich von TIMSS 2007 und 2011

	Mathematik				Naturwissenschaften			
	2007		2011		2007		2011	
	β	(SE)	β	(SE)	β	(SE)	β	(SE)
Modell I								
Referenzwert	539.6	(2.3) ***	540.7	(2.2) ***	548.0	(2.4) ***	545.6	(2.5) ***
Migrationshintergrund: ein Elternteil ^A	-32.1	(3.6) ***	-18.3	(4.1) ***	-44.7	(5.0) ***	-27.6	(4.8) ***
Migrationshintergrund: beide Elternteile ^B	-45.0	(4.9) ***	-39.2	(3.8) ***	-71.2	(5.4) ***	-57.4	(3.6) ***
	R ² = .07		R ² = .06		R ² = .12		R ² = .09	
Modell II								
Referenzwert	539.8	(2.3) ***	541.5	(2.2) ***	548.6	(2.4) ***	546.8	(2.5) ***
Migrationshintergrund: ein Elternteil	-31.7	(3.8) ***	-12.6	(4.5) **	-42.2	(5.1) ***	-18.4	(4.8) ***
Migrationshintergrund: beide Elternteile	-44.1	(5.2) ***	-29.8	(4.4) ***	-66.9	(5.9) ***	-41.6	(4.3) ***
Familiensprache ist nicht Deutsch ^C	-	- ns	-14.5	(4.0) ***	-	- ns	-23.5	(3.9) ***
	R ² = .07		R ² = .06		R ² = .13		R ² = .11	
Modell III								
Referenzwert	527.5	(2.6) ***	528.3	(2.4) ***	534.1	(2.5) ***	532.5	(3.0) ***
Migrationshintergrund: ein Elternteil	-27.1	(3.7) ***	-10.5	(4.3) *	-36.3	(4.9) ***	-16.1	(4.6) ***
Migrationshintergrund: beide Elternteile	-36.8	(5.2) ***	-21.9	(4.2) ***	-58.2	(5.9) ***	-32.7	(4.3) ***
Familiensprache ist nicht Deutsch	-	- ns	-13.7	(3.8) ***	-	- ns	-22.9	(3.7) ***
bedeutsamer Buchbesitz in der Familie ^D	35.2	(3.0) ***	33.1	(3.0) ***	41.5	(3.3) ***	36.0	(3.5) ***
	R ² = .12		R ² = .13		R ² = .18		R ² = .17	
Modell IV								
Referenzwert	524.0	(2.7) ***	524.0	(2.5) ***	530.3	(2.5) ***	528.4	(2.9) ***
Migrationshintergrund: ein Elternteil	-27.1	(4.0) ***	-	- ns	-35.4	(5.4) ***	-12.4	(4.9) *
Migrationshintergrund: beide Elternteile	-27.3	(5.7) ***	-16.1	(4.4) ***	-47.1	(6.6) ***	-27.6	(5.1) ***
Familiensprache ist nicht Deutsch	-	- ns	-12.5	(4.2) **	-9.2	(3.9) *	-20.2	(4.2) ***
bedeutsamer Buchbesitz in der Familie	23.1	(2.7) ***	25.6	(3.1) ***	29.2	(3.1) ***	27.5	(3.4) ***
Familie mit hohem Bildungsniveau ^E	32.7	(3.0) ***	29.6	(3.0) ***	32.6	(3.1) ***	32.8	(2.9) ***
	R ² = .16		R ² = .17		R ² = .19		R ² = .20	
Modell V								
Referenzwert	523.7	(2.6) ***	523.7	(2.5) ***	530.0	(2.4) ***	528.1	(3.0) ***
Migrationshintergrund: ein Elternteil	-24.8	(4.6) ***	-	- ns	-31.6	(5.3) ***	-11.7	(5.0) *
Migrationshintergrund: beide Elternteile	-25.2	(5.7) ***	-16.7	(4.5) ***	-44.9	(6.9) ***	-27.4	(5.2) ***
Familiensprache ist nicht Deutsch	-	- ns	-11.0	(4.4) *	-	- ns	-19.2	(4.4) ***
bedeutsamer Buchbesitz in der Familie	21.0	(2.6) ***	23.5	(3.3) ***	26.8	(3.0) ***	24.8	(3.4) ***
Familie mit hohem Bildungsniveau	25.9	(3.0) ***	23.1	(4.0) ***	23.7	(3.3) ***	22.0	(3.7) ***
Familie mit hohem sozioökonomischen Status ^F	15.1	(3.5) ***	11.1	(4.2) **	18.4	(3.4) ***	16.4	(4.3) ***
	R ² = .16		R ² = .17		R ² = .19		R ² = .20	

β = Regressionsgewichte (unstandardisiert)

Signifikanzniveau: ns= nicht signifikant; *= signifikant ($p < .05$); **= signifikant ($p < .01$); ***= signifikant ($p < .001$)

A= Migrationshintergrund nach Geburtsland der Eltern (0= beide Elternteile in Deutschland geboren; 1= ein Elternteil im Ausland geboren)

B= Migrationshintergrund nach Geburtsland der Eltern (0= beide Elternteile in Deutschland geboren; 1= beide Elternteile im Ausland geboren)

C= Familiensprache nach Angabe der Schülerinnen und Schüler (0= Deutsch: immer oder fast immer; 1= Deutsch: manchmal oder nie)

D= Heimischer Buchbesitz nach Angabe der Eltern (0= maximal 100 Bücher; 1= mehr als 100 Bücher)

E= Bildungsniveau nach höchstem Bildungsabschluss der Eltern (0= kein Elternteil mit mindestens Fachhochschulabschluss; 1= mindestens ein Elternteil mit mindestens Fachhochschulabschluss)

F= Berufsstatus: Höchster ISEI (*International Socio-Economic Index of Occupational Status*) im Haushalt (0 = Werte unter 65 Punkte; 1 = Werte über 65 Punkte [z.B. Akademiker])

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Da neben der Migrationsgeschichte der Familien und dem familiären Sprachgebrauch in verschiedenen Untersuchungen auch der sozioökonomische und kulturelle Status der Familien einen Zusammenhang mit dem Bildungserfolg aufweist, ist dieser hier anhand des heimischen Buchbesitzes als ein robuster Indikator zum soziokulturellen Status der Familien berücksichtigt. Beim höchsten Bildungsabschluss der Eltern werden die Familien in den Blick genommen, in denen mindestens ein Elternteil einen Fachhochschulabschluss oder höheren Abschluss erreicht hat und beim höchsten Berufsstatus werden die Familien betrachtet, in denen mindestens ein Elternteil einen akademischen Beruf ausübt oder eine führende Position in einem Unternehmen innehat. Weitere vertiefende

Analysen zu sozioökonomischem und kulturellem Status finden sich in Kapitel 7 in diesem Band.

Tabelle 8.4 ist zu entnehmen, dass sich in dem sukzessiven Aufbau der Modelle sowohl zwischen Mathematik und Naturwissenschaften als auch zwischen den beiden Erhebungszyklen von 2007 und 2011 Unterschiede ergeben, die jedoch nicht im Widerspruch zueinander stehen. Vielmehr werden Spezifika der verschiedenen Erhebungen und Fächer deutlich.

Im Modell I zeigt sich, dass im Vergleich zu Schülerinnen und Schülern, deren Eltern in Deutschland geboren wurden, die Kinder mit einem im Ausland geborenen Elternteil 2007 in den mathematischen Kompetenzen einen Leistungsrückstand von 32 Punkten aufweisen. Für Schülerinnen und Schüler, deren Eltern im Ausland geboren wurden, beträgt der Leistungsrückstand gegenüber den Kindern ohne Migrationshintergrund sogar 45 Punkte. Diese Differenz liegt 2011 für Kinder mit einem im Ausland geborenen Elternteil mit 18 Punkten unter dem Wert von 2007. Die Differenz in den mathematischen Kompetenzen zwischen Kindern, deren Eltern im Ausland geboren wurden, und Kindern ohne Migrationshintergrund, liegt mit 39 Punkten zwar nominell unter dem Wert von 2007 – signifikant ist der Unterschied zwischen 2007 und 2011 jedoch nicht.

Wird der familiäre Sprachgebrauch in die Modellrechnungen (Modell II) aufgenommen, so zeigen sich zwar für 2007 keine signifikanten Effekte, jedoch für 2011: Schülerinnen und Schüler, die nie zu Hause Deutsch sprechen, weisen eine um 15 Punkte niedrigere mathematische Kompetenz auf als diejenigen, die immer Deutsch zu Hause sprechen. Unter Berücksichtigung des Sprachindikators verringern sich gleichzeitig die Effekte des Geburtslands der Eltern. Diese Kovariation des Migrationshintergrunds, gemessen am Geburtsort der Eltern und am familiären Sprachgebrauch, bleibt auch unter Kontrolle weiterer Variablen des sozialen Hintergrunds erhalten. Dieser Befund ist auffällig und kann möglicherweise auf etwas veränderte Charakteristika von Schülerfamilien und verändertem Sprachverhalten in Familien mit Migrationsgeschichte in Deutschland zurückgeführt werden. Er sollte Gegenstand weiterer vertiefender Analysen sein.

Bei zusätzlicher Berücksichtigung des kulturellen Hintergrunds der Familien, indikatorisiert durch den heimischen Buchbesitz (Modell III), schwächen sich die für den Migrationshintergrund beschriebenen Effekte für die Kinder mit zwei im Ausland geborenen Elternteilen ab. Dies deutet darauf hin, dass sich die Familien mit Migrationshintergrund bezüglich ihres sozialen Hintergrunds unterscheiden und somit in sich ebenfalls als heterogene Gruppe zu charakterisieren sind. Gleiches zeigt sich bei zusätzlicher Berücksichtigung des Bildungsstatus der Familien (Modell IV). Wird zusätzlich kontrolliert, ob mindestens ein Elternteil einen Fachhochschul- oder höheren Abschluss erreicht hat, verringert sich der Leistungsrückstand von Kindern aus Familien mit Migrationsgeschichte weiter.

Unter Hinzunahme der Information bezüglich der Zugehörigkeit zur Gruppe mit hohem familiären Berufsstatus (2007: 25%; 2011: 30% bei gleicher Unterscheidung) verändern sich weder 2007 noch 2011 die bereits im Modell IV betrachteten Werte für den Zusammenhang zwischen familiären Merkmalen und der mathematischen Kompetenz von Schülerinnen und Schülern.

Für die Naturwissenschaften lässt sich im Vergleich zu den Modellen für die Mathematik im Wesentlichen nur ein Unterschied ausmachen. Die Differenzen in den naturwissenschaftlichen Kompetenzen sind sowohl für die Schülerinnen und Schüler mit einem als auch für die mit zwei im Ausland geborenen Elternteilen stärker ausgeprägt als für die mathematischen Kompetenzen. Mit 45 beziehungsweise 71 Punkten (2007) und 28 beziehungsweise 57 Punkten (2011) sind sie als

substantiell zu bezeichnen. Als markantestes gemeinsames Merkmal der Modelle für Mathematik und Naturwissenschaften bleibt der 2007 nicht auffällige aber 2011 signifikant ausfallende Effekt der Familiensprache festzuhalten, der auch unter Berücksichtigung weiterer familiärer Merkmale erhalten bleibt.

6 Schlussfolgerungen

Schülerinnen und Schüler, die zu Hause nicht die Testsprache sprechen, erzielen sowohl in Mathematik als auch in den Naturwissenschaften in fast allen Teilnehmerstaaten schlechtere Leistungen als ihre Mitschülerinnen und Mitschüler, die zu Hause die Testsprache sprechen. Dies ist auch in Deutschland nicht anders – hier sind die Disparitäten zwischen den Kompetenzen von Kindern mit und ohne Migrationshintergrund im Vergleich von TIMSS 2007 und 2011 gleich.

In der Zusammenschau der Ergebnisse für mathematische und naturwissenschaftliche Kompetenzen gibt es bezüglich der Verteilung der Schülerinnen und Schüler auf die unterschiedlichen Kompetenzstufen kaum überraschende Befunde. Für Deutschland zeigt sich, dass der Anteil der Schülerinnen und Schüler, die nicht die Kompetenzstufe III erreichen, für Kinder aus Familien mit Migrationsgeschichte größer ausfällt als für Kinder ohne Migrationshintergrund. Hingegen ergibt sich für die höheren Kompetenzstufen ein umgekehrtes Verhältnis.

Im Vergleich zu TIMSS 2007 zeigen sich jedoch positive Veränderungen: In Mathematik erzielen Schülerinnen und Schüler mit einem im Ausland geborenen Elternteil signifikant bessere Leistungen als bei TIMSS 2007. Auch in den Naturwissenschaften konnten die Schülerinnen und Schüler mit Migrationshintergrund aufholen. Hier erreichen nicht nur die Viertklässlerinnen und Viertklässler aus Familien, bei denen ein Elternteil im Ausland geboren wurde bessere Leistungen, sondern auch die Schülerinnen und Schüler aus Familien, bei denen beide Elternteile im Ausland geboren wurden.

Ob diese positiven Veränderungen zwischen TIMSS 2007 und 2011 einen Hinweis auf einen möglichen Trend darstellen, wird sich erst unter Berücksichtigung eines weiteren Erhebungszyklus – wie er für TIMSS 2015 in Vorbereitung ist – untersuchen lassen.

Die in diesem Kapitel berichteten Ergebnisse und die in IGLU 2011 für das Leseverständnis gezeigten Befunde (Schwippert et al., 2012) verdeutlichen, dass die Förderung von Kindern mit Migrationshintergrund auch weiterhin ein wichtiges Handlungsfeld darstellt. Es zeigt sich jedoch, dass sich die Unterschiede in den mathematischen und naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern aus Familien mit Migrationshintergrund im Vergleich zu Kindern ohne Migrationsgeschichte nicht allein auf den familiären Sprachgebrauch zurückführen lassen. Vielmehr liegen Unterschiede auch im sozialen Hintergrund von Schülerfamilien begründet. Förderprogramme sollten entsprechend – auch um Stigmatisierungen zu vermeiden – konzeptionell primär an den konkreten Förderbedarfen beziehungsweise Leistungsschwächen der individuellen Schülerinnen und Schüler orientiert sein. Mögliche Teilnahmehemmnisse, die in der besonderen ökonomischen und sozialen Situation von Schülerfamilien begründet sind, sollten gleichwohl ernst genommen werden.

Die Ergebnisse weisen ebenfalls darauf hin, dass es für die Kompetenzentwicklung der Kinder sehr förderlich sein kann, Migrantenfamilien bei ent-

sprechender Qualifikation den gesellschaftlichen Aufstieg zu erleichtern. Dieses Problem kann allerdings nicht allein von der Schule gelöst werden.

Literatur

- Autorengruppe Bildungsbericht. (2008). *Bildung in Deutschland 2008. Ein indikatoren-gestützter Bericht mit einer Analyse zu Übergängen im Anschluss an den Sekundarbereich II*. Bielefeld: Bertelsmann.
- Esser, H. (2006). *Sprache und Integration. Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt a. M.: Campus.
- Gogolin, I. (1994). *Der monolinguale Habitus der multilingualen Schule*. Münster: Waxmann.
- Gogolin, I., Neumann, U. & Roth, H.-J. (2003). *Förderung von Kindern und Jugendlichen mit Migrationshintergrund. BLK-Gutachten*. Bonn: Bundesministerium für Bildung und Forschung.
- 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. 209–235). Münster: Waxmann.
- Konsortium Bildungsberichterstattung. (2006). *Bildung in Deutschland. Ein indikatoren-gestützter Bericht mit einer Analyse zu Bildung und Migration*. Bielefeld: Bertelsmann.
- Martin, M. O., Mullis, I. V. S. & Kennedy, A. M. (Hrsg.). (2003). *PIRLS 2001 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S. & Kennedy, A. M. (Hrsg.). (2007). *PIRLS 2006 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Minnich, C. A., Stanco, G. M., Arora, A., Centurino, V. A. S. & Castle, C. E. (Hrsg.). (2012). *TIMSS 2011 Encyclopedia. Education policy and curriculum in mathematics and science, Volumes 1 and 2*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Reich, H. H. & Roth, H.-J. (2007). HAVAS 5 – das Hamburger Verfahren zur Analyse des Sprachstands von Fünfjährigen. In H. H. Reich, H.-J. Roth & U. Neumann (Hrsg.), *Sprachdiagnostik im Lernprozess* (S. 71–94). Münster: Waxmann.
- Schwippert, K., Bos, W. & Lankes, E.-M. (2003). Heterogenität und Chancengleichheit am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 265–302). Münster: Waxmann.
- Schwippert, K. & Schnabel, K. U. (2000). Mathematisch-naturwissenschaftliche Grundbildung ausländischer Schulausbildungsabsolventen. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit* (Bd. 1, S. 282–300). Opladen: Leske + Budrich.
- Schwippert, K., Wendt, H. & Tarelli, I. (2012). Lesekompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Stanat, P. (2008). Heranwachsende mit Migrationshintergrund im deutschen Bildungswesen. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (S. 685–743). Reinbek: Rowohlt.
- Stanat, P., Rauch, D. & Segeritz, M. (2010a). 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. (2010b). Der Einfluss des Migrantenanteils in Schulklassen auf den Kompetenzerwerb: Längsschnittliche Überprüfung eines umstrittenen Effekts. *Zeitschrift für Pädagogik* (55. Beiheft), 147–164.

- Valtin, R., Bos, W., Hornberg, S. & Schwippert, K. (2007). Zusammenschau und Schlussfolgerungen. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 329–348). Münster: Waxmann.
- Walter, O. (2008). Herkunftsassoziierte Disparitäten im Lesen, der Mathematik und den Naturwissenschaften: ein Vergleich zwischen PISA 2000, PISA 2003 und PISA 2006. *Zeitschrift für Erziehungswissenschaft, Sonderheft 10*, 149–168.
- Walter, O. & Taskinen, P. (2007). Kompetenzen und bildungsrelevante Einstellungen von Jugendlichen mit Migrationshintergrund in Deutschland: Ein Vergleich mit ausgewählten OECD-Staaten. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 337–366). Münster: Waxmann.

Kapitel IX

Leistungsprofile von Viertklässlerinnen und Viertklässlern in Deutschland

Wilfried Bos, Heike Wendt, Ali Ünlü, Renate Valtin, Benjamin Euen, Daniel Kasper und Irmela Tarelli

1 Einleitung

Mit der *Internationalen Grundschul-Lese-Untersuchung (IGLU)/Progress in International Reading Literacy Study (PIRLS)* wird in einem fünfjährigen Zyklus das Leseverständnis von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe untersucht. Mit der *Trends in International Mathematics and Science Study (TIMSS)* werden in einem vierjährigen Zyklus die mathematischen und naturwissenschaftlichen Leistungen von Schülerinnen und Schülern – unter anderem am Ende der vierten Jahrgangsstufe – untersucht. Beide Studien werden von der *International Association for the Evaluation of Educational Achievement (IEA)* verantwortet. Im Jahr 2011 fiel der Erhebungszeitpunkt der beiden Studien IGLU und TIMSS erstmalig zusammen, so dass sich die Möglichkeit ergab, beide Studien gemeinsam durchzuführen. In Deutschland wurde – wie in 37 weiteren Staaten und Regionen (davon drei als Benchmark-Teilnehmer) – eine gemeinsame Stichprobe von Schülerinnen und Schülern realisiert (zu den Einzelheiten der Administration vgl. Kapitel 2 in diesem Band). Die gemeinsame Stichprobe in Deutschland umfasst 197 Grundschulen und 4229 Schülerinnen und Schüler. Die Testung für IGLU und TIMSS fand in der Regel an zwei aufeinanderfolgenden Testtagen statt, wobei an 50 Prozent der Schulen am ersten Testtag TIMSS und an den anderen 50 Prozent der Schulen am ersten Testtag IGLU durchgeführt wurde (vgl. Kapitel 2). Für insgesamt 3928 Schülerinnen und Schüler aus 197 Schulen liegen nun Daten vor, die Ergebnisse der Leistungstestung im Leseverständnis (IGLU), in Mathematik und in den Naturwissenschaften (TIMSS) umfassen. Diese besondere Datenlage eröffnet weiterführende Möglichkeiten der Auswertung.

In diesem Kapitel werden Leistungstestwerte und Kompetenzprofile (nachfolgend unter dem Begriff „Leistungsprofile“ subsumiert) von Grundschulkindern über die drei erfassten Kompetenzbereiche Leseverständnis (IGLU), Mathematik

und Naturwissenschaften (TIMSS) hinweg differenziert betrachtet. Dabei werden lediglich Ergebnisse für Deutschland berichtet, da der internationale Ergebnisbericht zu IGLU/TIMSS 2011 erst im Oktober 2013 erscheinen wird. Für die Analysen werden ausschließlich die Daten jener 3 928 Viertklässlerinnen und Viertklässler berücksichtigt, für die Leistungswerte in allen drei Leistungsbereichen vorliegen.

Bei der Betrachtung der drei Leistungsbereiche steht zum einen die Frage im Vordergrund, ob sich Gruppen von Schülerinnen und Schülern mit ähnlichen Mustern in den Leistungsausprägungen über die verschiedenen Domänen hinweg identifizieren lassen und in Bezug auf welche Charakteristika sich diese Gruppen unterscheiden. Zum anderen ist von Interesse, ob sich für Schülerinnen und Schüler am Ende der Grundschulzeit eher homogene Leistungsprofile oder starke fachspezifische Unterschiede erkennen lassen. In diesem Kapitel wird dabei die Verteilung der Leistungen der Schülerinnen und Schüler über die drei genannten Kompetenzdomänen sowie deren Inhaltsbereiche, für die Subskalen gebildet wurden, untersucht.

Für das Leseverständnis werden die beiden in IGLU differenzierten Leseintentionen herangezogen, die über zwei Textsorten erfasst werden (vgl. Bremerich-Vos, Valtin & Tarelli, 2012):

- das Lesen literarischer Texte,
- das Lesen informierender Texte.

Für die mathematischen Kompetenzen werden die in TIMSS unterschiedenen Inhaltsbereiche genutzt (vgl. Kapitel 3 in diesem Band):

- Arithmetik,
- Geometrie/Messen sowie
- Umgang mit Daten.

Für die naturwissenschaftlichen Kompetenzen werden die ebenfalls in TIMSS unterschiedenen Inhaltsbereiche berücksichtigt (vgl. Kapitel 4 in diesem Band):

- Biologie,
- Geographie sowie
- Physik/Chemie.

Eine differenzierte Betrachtung von weiterführenden Fragestellungen, beispielsweise den Zusammenhängen von Verstehensprozessen (IGLU) und kognitiven Anforderungen (TIMSS), ist vertiefenden Analysen vorbehalten.

Im Folgenden werden zunächst – als erster Ansatzpunkt zur Untersuchung der Leistungsprofile über die Kompetenzbereiche – die Verteilungen von Schülerinnen und Schülern auf den Kompetenzstufen in den jeweiligen Domänen in den Blick genommen (Abschnitt 2). Um den Zusammenhang zwischen den verschiedenen Skalen oder Subskalen von IGLU und TIMSS zu untersuchen, wurde auf eine weiterführende Methodik zurückgegriffen, die in Abschnitt 3 dargestellt wird: Damit die Leistungen der Viertklässlerinnen und Viertklässler in Deutschland auf einer gemeinsamen Skala domänenübergreifend beschrieben werden können, war es notwendig, eine multidimensionale *Item-Response-Modellierung* der Leistungstestdaten von IGLU und TIMSS vorzunehmen. In Abschnitt 3.1 wird dieses Vorgehen näher beschrieben. Für die Identifikation von Schülerinnen und Schülern mit unterschiedlichen Leistungsprofilen wurden anschließend *latente Profilanalysen* gerechnet, deren Ergebnisse in Abschnitt 3.2 skizziert werden. Eine Beschreibung der daraus resultierenden Leistungsprofile

anhand von profilspezifischen Informationen oder hintergrundbezogenen Variablen führt zur Diskussion von generellen Leistungstypen. In Abschnitt 4.1 werden diese Typen anhand der folgenden Fragestellungen diskutiert:

- Wie groß sind die Anteile jener Grundschülerinnen und Grundschüler, die über alle acht inhaltlichen Subdomänen der drei Kompetenzbereiche hinweg entweder sehr hohe oder sehr geringe Leistungsergebnisse erzielen beziehungsweise in allen Subdomänen im mittleren Leistungsbereich liegen?
- Wie groß sind die Anteile jener Grundschülerinnen und Grundschüler, die über alle acht inhaltlichen Subdomänen der drei Kompetenzbereiche hinweg verschiedene Leistungsprofile (z.B. hohe mathematische Leistungen, aber geringe Leistungen im Bereich Leseverständnis und Naturwissenschaften) aufweisen?

Eine weitergehende Diskussion der Leistungstypen wird in Abschnitt 4.2 und Abschnitt 4.3 anhand der folgenden Fragestellung vorgenommen:

- Welche Unterschiede zeigen sich zwischen bestimmten Leistungsprofilen in Bezug auf wichtige individuelle und familiäre Merkmale, von denen angenommen werden kann, dass sie mit schulischen Leistungen zusammenhängen (wie Geschlecht, kulturelle und soziale Merkmale der Familie, Migrationshintergrund und Familiensprache)?
- Welche Unterschiede zeigen sich zwischen bestimmten Leistungsprofilen in Bezug auf fachbezogene Einstellungen und Selbstkonzepte der Schülerinnen und Schüler?

2 Domänenübergreifende Betrachtung der Verteilung auf die Kompetenzstufen

In den Ausführungen zu den mathematischen (vgl. Kapitel 3 in diesem Band), naturwissenschaftlichen (vgl. Kapitel 4 in diesem Band) sowie den Lesekompetenzen (vgl. Bos, Bremerich-Vos, Tarelli & Valtin, 2012) im internationalen Vergleich ist die Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen nach Domänen getrennt dargestellt. In diesem Abschnitt soll zunächst beschrieben werden, wie sich die Verteilungen für die gemeinsame Stichprobe von IGLU/TIMSS 2011 darstellen.

Sowohl in IGLU als auch in TIMSS wurden die Leistungen auf einer Skala mit einem Skalenmittelwert von 500 und einer Standardabweichung von 100 dargestellt. Für alle drei Leistungsbereiche wurde die jeweilige Skala an bestimmten Referenzwerten, den sogenannten *Benchmarks*, in Skalenbereiche mit derselben Weite (jeweils 75 Punkte für die mittleren Stufen) aufgeteilt. Testwerte unter 400 Punkten entsprechen der Kompetenzstufe I, Testwerte zwischen 400 und 475 Punkten der Kompetenzstufe II, zwischen 475 und 550 Punkten der Kompetenzstufe III, zwischen 550 und 625 Punkten der Kompetenzstufe IV und Werte über 625 Punkten der Kompetenzstufe V (vgl. zur Bildung der Kompetenzstufen Kapitel 2 in diesem Band und Tarelli, Wendt, Bos & Zylowski, 2012). Die Kennzeichnung beziehungsweise die inhaltliche Beschreibung der Kompetenzstufen bezieht sich auf die Anforderungen der Aufgaben, die von den Schülerinnen und Schülern auf dieser Kompetenzstufe mit großer Wahrscheinlichkeit gelöst werden. Der konkreten Beschreibung und Entwicklung der Kompetenzstufen liegt jeweils eine spezifische Auswahl von Aufgaben zu-

grunde, die für das auf den vier Benchmarks gezeigte Leistungsniveau charakteristisch ist.

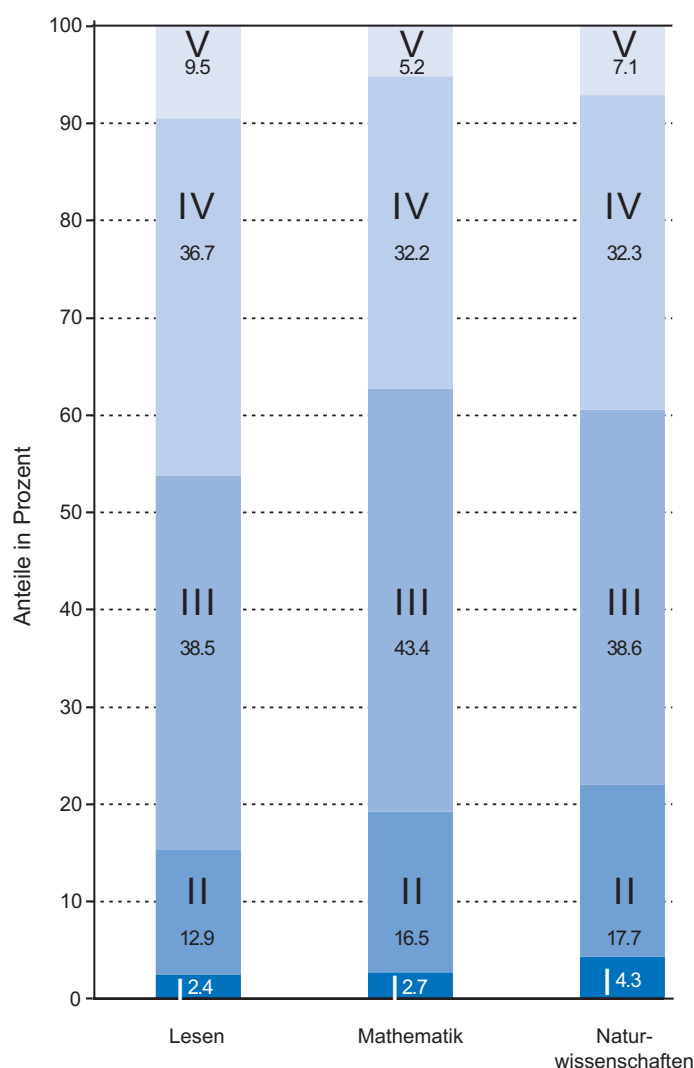
Die Definition der Kompetenzstufen ist auf die Aufgaben der unterschiedlichen Tests zur Erfassung von Leseverständnis, Mathematik und Naturwissenschaften ausgerichtet. In Tabelle 9.1 werden die wesentlichen Elemente der Kompetenzstufenbeschreibungen für das Leseverständnis, Mathematik und Naturwissenschaften aufgeführt (für eine ausführliche Beschreibung der Kompetenzstufen vgl. Kapitel 3 und 4 in diesem Band sowie Bremerich-Vos et al., 2012).

Tabelle 9.1: Kurzfassung der Kompetenzstufenbeschreibungen für Leseverständnis, Mathematik und Naturwissenschaften

Kompetenzstufe	Leseverständnis	Mathematik	Naturwissenschaften
V	Auf Textpassagen bzw. den Text als ganzen bezogene Aussagen selbstständig interpretierend und kombinierend begründen	Mathematische Fertigkeiten und Fähigkeiten verständig beim Lösen verhältnismäßig komplexer Probleme anwenden und Vorgehen erläutern	Grundlegendes Verständnis des Prozesses naturwissenschaftlichen Arbeitens und Wissen über naturwissenschaftliche Prozesse und Zusammenhänge anwenden
IV	Für die Herstellung von Kohärenz auf der Ebene des Textes relevante Aspekte erfassen und komplexe Schlüsse ziehen	Mathematische Fertigkeiten und Fähigkeiten für das Lösen von Problemen anwenden	Mit Wissen und Verständnis alltägliche Phänomene erklären
III	„Verstreute“ Informationen miteinander verknüpfen	Elementares mathematisches Wissen sowie elementare mathematische Fertigkeiten und Fähigkeiten in einfachen Situationen anwenden	Basiswissen und grundlegendes Verständnis auf naturwissenschaftsbezogene Situationen anwenden
II	Explizit angegebene Einzelinformationen identifizieren und benachbarte Informationen miteinander verknüpfen	Elementares mathematisches Wissen sowie elementare mathematische Fertigkeiten und Fähigkeiten	Elementares Wissen über Biologie und Physik/Chemie
I	Rudimentäres Leseverständnis	Rudimentäres schulisches Anfangswissen	Rudimentäres schulisches Anfangswissen

In Abbildung 9.1 werden die Anteile der Schülerinnen und Schüler auf den jeweiligen Kompetenzstufen dargestellt.

Abbildung 9.1: Verteilungen von Schülerinnen und Schülern auf die Kompetenzstufen in den Domänen Leseverständnis, Mathematik und Naturwissenschaften (Anteile in Prozent)



Abweichungen zu den Darstellungen in den Kapiteln 3 und 4 in diesem Band sowie in Bos, et al. (2012) ergeben sich durch die verkleinerte Stichprobe.

Es zeigt sich in Abbildung 9.1, dass die Anteile von Viertklässlerinnen und Viertklässlern in Deutschland, die Werte auf dem Niveau der höchsten Kompetenzstufe V erzielen, in allen drei Kompetenzbereichen verhältnismäßig gering sind: Im Lesen sind es 9.5 Prozent, in Mathematik 5.2 Prozent und in den Naturwissenschaften 7.1 Prozent (vgl. auch Kapitel 3 und 4 in diesem Band sowie Bos et al., 2012). Auf Kompetenzstufe IV sind es jeweils etwa ein Drittel der Schülerinnen und Schüler: Im Lesen 36.7 Prozent, in Mathematik 32.2 Prozent und in den Naturwissenschaften 32.3 Prozent. Diese Schülerinnen und Schüler verfügen mindestens über befriedigende Kompetenzen, die sie zu einer erfolgreichen Teilnahme am Unterricht an weiterführenden Schulen befähigen sollten. Der Anteil von Schülerinnen und Schülern, die Leistungen auf dem Niveau der Kompetenzstufe III erzielen, liegt für Lesen bei 38.5 Prozent, für Mathematik bei 43.4 Prozent und für Naturwissenschaften bei 38.6 Prozent. Diese Schülerinnen und Schüler werden vermutlich in der Sekundarstufe I eine gezielte Unterstützung benötigen, damit sie den Leistungsanforderungen in dieser Schulstufe gewachsen sind. Leistungen auf dem Niveau der Kompetenzstufe II

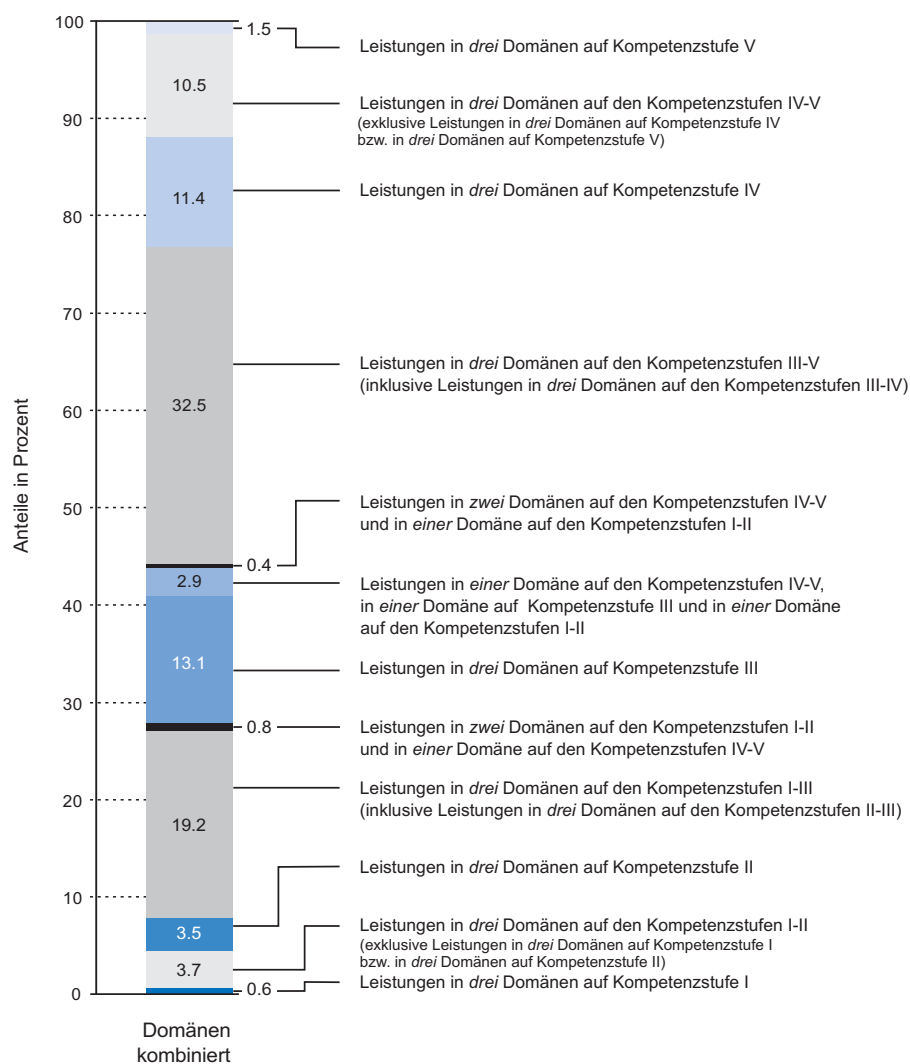
erzielen im Lesen 12.9 Prozent, in Mathematik 16.5 Prozent und in den Naturwissenschaften 17.7 Prozent der Viertklässlerinnen und Viertklässler in Deutschland. Der Anteil von Schülerinnen und Schülern, deren Leistungen nur dem Niveau der Kompetenzstufe I entsprechen, liegt im Lesen bei 2.4 Prozent, in Mathematik bei 2.7 Prozent und in den Naturwissenschaften bei 4.3 Prozent. Schülerinnen und Schüler auf den beiden untersten Kompetenzstufen verfügen nicht über ausreichende Basiskompetenzen. Für sie ist dringend eine zusätzliche schulische Förderung schon in der Grundschule, aber auch in der Sekundarstufe I notwendig.

Neben einer parallelen Betrachtung der Verteilungen auf die Kompetenzstufen getrennt nach den einzelnen Domänen ist von Interesse, wie sich die Verteilung der Schülerinnen und Schüler in Deutschland darstellt, wenn die Zuordnungen zu den Kompetenzstufen in ihren Kombinationen für alle drei Kompetenzdomänen zusammengeführt werden. Da sich bei fünf Kompetenzstufen und drei Domänen theoretisch 125 Kombinationsmöglichkeiten ergeben, wurden diese in 12 Gruppen zusammengefasst. Die Reduzierung der Kombinationsmöglichkeiten bezieht sich insbesondere darauf, dass sich die Gruppen lediglich auf die Unterschiede zwischen den Domänen beziehen, jedoch nicht differenzieren, in welcher der Kompetenzdomänen höhere beziehungsweise niedrigere Kompetenzstufen erreicht wurden. So wurde beispielsweise eine Gruppe von Schülerinnen und Schülern gebildet, deren Leistungen in allen drei Bereichen auf den Kompetenzstufen IV–V liegen – unabhängig davon, in welchen Domänen die Leistungen auf Kompetenzstufe IV und in welchen Domänen sie auf Kompetenzstufe V liegen. In Abbildung 9.2 ist die prozentuale Verteilung der Schülerinnen und Schüler in diesen Gruppen dargestellt.

Betrachtet man zunächst den Anteil der Schülerinnen und Schüler, die in allen drei Bereichen Leistungen auf derselben Kompetenzstufe erzielen, so ergibt sich Folgendes: In allen drei Domänen sind auf Kompetenzstufe I 0.6 Prozent, auf Kompetenzstufe II 3.5 Prozent, auf Kompetenzstufe III 13.1 Prozent, auf Kompetenzstufe IV 11.4 Prozent und auf Kompetenzstufe V 1.5 Prozent der Grundschülerinnen und -schüler. Mit einem Anteil von 30.1 Prozent aller Schülerinnen und Schüler erreicht also etwa ein Drittel der Viertklässlerinnen und Viertklässler ein ausgeglichenes Leistungsniveau über die drei Domänen hinweg. Dieser Anteil erhöht sich, wenn jeweils die aneinandergrenzenden Kompetenzstufen (I und II, II und III, III und IV, IV und V) für die Betrachtung zusammengefasst werden. In allen drei Domänen befinden sich auf den Kompetenzstufen I und II 3.7 Prozent, auf den Kompetenzstufen II und III 17.2 Prozent, auf den Kompetenzstufen III und IV 29.1 Prozent sowie auf den Kompetenzstufen IV und V 10.5 Prozent der Schülerinnen und Schüler. Werden zu diesen Anteilen diejenigen Kinder hinzugenommen, die in allen drei Bereichen dieselbe Kompetenzstufe erreichen (30.1%), erhöht sich der Anteil der Schülerinnen und Schüler mit einem einigermaßen ausgeglichenen Leistungsprofil auf 90.6 Prozent.

Insgesamt betrachtet zeigt sich, dass 55.9 Prozent der Schülerinnen und Schüler in allen drei Domänen Leistungen mindestens auf dem Niveau von Kompetenzstufe III und in mindestens einem Bereich auf einem höheren Niveau erzielen. Hingegen erreichen 40.1 Prozent der Schülerinnen und Schüler in allen drei Bereichen Leistungen, die maximal dem Niveau der Kompetenzstufe III entsprechen. Aus der Abbildung wird auch ersichtlich, dass es nur wenige Schülerinnen und Schüler gibt, deren Leistungen in ein oder zwei Bereichen den untersten und gleichzeitig in ein oder zwei Bereichen den obersten

Abbildung 9.2: Verteilung von Schülerinnen und Schülern auf die Kompetenzstufen, kombiniert für die Domänen Leseverständnis, Mathematik und Naturwissenschaften (Anteile in Prozent)



IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Kompetenzstufen zugeordnet werden und bei denen man eine ‚Inselbegabung‘ beziehungsweise ‚Lernschwäche‘ in nur einer Domäne annehmen könnte. Die Extremgruppen bilden hier 2.9 Prozent der Schülerinnen und Schüler, deren Leistungen über alle Kompetenzbereiche streuen, 0.4 Prozent der Schülerinnen und Schüler, die in zwei Domänen Leistungen auf den Kompetenzstufen IV oder V und in einer Domäne Leistungen auf den Kompetenzstufen I oder II erzielen, sowie weitere 0.8 Prozent der Schülerinnen und Schüler, die in einer Domäne Leistungen auf den Kompetenzstufen IV oder V und in zwei Domänen Leistungen auf den Kompetenzstufen I oder II erzielen. Zwei weitere Gruppen von Schülerinnen und Schülern mit heterogenen Leistungsprofilen, die jedoch nicht direkt aus der Abbildung ersichtlich sind, bilden jene, deren Leistungen über drei Kompetenzstufen streuen: von III bis V sind dies 3.4 Prozent sowie von I bis III 1.9 Prozent der Schülerinnen und Schüler.

Eine tiefgehende Interpretation der hier gewählten Darstellungen stößt in zweierlei Hinsicht an Grenzen. Zum einen bedingt das Verfahren der Zuordnung von Schülerinnen und Schülern zu den Kompetenzstufen (vgl. Kapitel 2), dass Schülerinnen und Schüler, die jeweils mit ihren Leistungspunkten nahe an den

Cut-Off-Points der Benchmarks liegen, trotz relativer Leistungshomogenität auf den stetigen Skalen unterschiedlichen diskretisierten Kompetenzstufen zugeordnet werden. Entsprechend besteht die Gefahr, den Anteil an Schülerinnen und Schülern mit heterogenen Leistungsprofilen zu überschätzen. Zum anderen ist eine simultane Betrachtung der drei Kompetenzdomänen (z.B. inklusive ihrer Abhängigkeiten) in der Berechnung von Testwerten und Leistungsprofilen der Schülerinnen und Schüler informativer.

Um die Schülerinnen und Schüler auf einer gemeinsamen Metrik verorten zu können und so unter Hinzunahme von wichtigen Hintergrundmerkmalen eine differenziertere Analyse der Leistungsstände über die Gesamtskalen und ihre Inhaltsbereiche hinweg zu realisieren, wird im Folgenden eine methodische Herangehensweise gewählt, die dies ermöglicht.

3 Multidimensionale Item-Response-Modellierung und latente Profilanalyse

Bisher erfolgte die Analyse der Leistungsergebnisse von IGLU und TIMSS nach Kompetenzbereichen getrennt. Da im vorliegenden Kapitel die Frage nach dem Abschneiden der Schülerinnen und Schüler über die verschiedenen Kompetenzdomänen hinweg im Vordergrund steht, kann nicht auf die in den vorangegangenen Kapiteln berichteten Leistungswerte unmittelbar zurückgegriffen werden, da diese separat für jede Domäne skaliert sind und keine kompetenzdomänenübergreifende Interpretation von Leistungsergebnissen (z.B. von Interaktionen) erlauben. Aus diesem Grund wurden die Testwerte der Schülerinnen und Schüler in einer Weise geschätzt, dass sich die Schülerleistungen in einem ersten Schritt in den Domänen Leseverständnis, Mathematik und Naturwissenschaften auf einer gemeinsamen Metrik verorten. Konkret wurde mit dem deutschen Datensatz eine multidimensionale *Item-Response-Modellierung* durchgeführt (Abschnitt 3.1). Die aus dieser gemeinsamen Skalierung generierten Leistungstestwerte, die eine domänenübergreifende Untersuchung von Schülerleistungen erlauben, werden in einem zweiten Schritt für eine Identifikation von Schülerinnen und Schülern mit unterschiedlichen Leistungsprofilen genutzt. Hierfür wurden latente Profilanalysen gerechnet, deren Ergebnisse in Abschnitt 3.2 erläutert werden. Eine Beschreibung der identifizierten Gruppen von Schülerinnen und Schülern erfolgt in Abschnitt 4.1 als Typen nach der durchschnittlichen Leistung (profilspezifisch) und in Abschnitt 4.2 als Leistungstypen nach relevanten Hintergrundmerkmalen (hintergrundmodellbezogen).

3.1 Simultane Skalierung der Daten von IGLU 2011 und TIMSS 2011

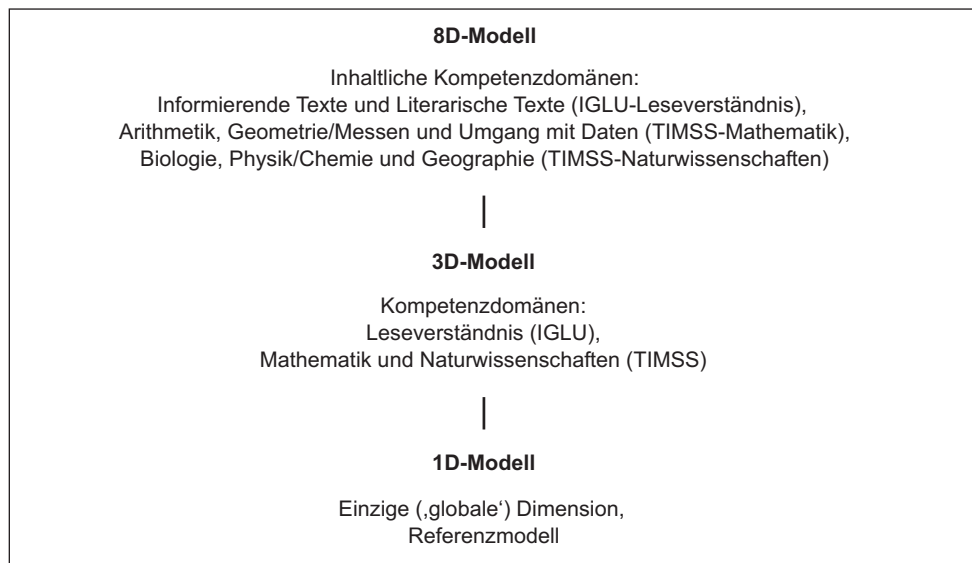
Um die Leistungen der Viertklässlerinnen und Viertklässler in Deutschland auf einer gemeinsamen Skala domänenübergreifend beschreiben zu können, wurde eine simultane Skalierung der für Deutschland vorliegenden IGLU/TIMSS-Daten vorgenommen. Um Verwechslungen mit den international vergleichbaren domänenspezifischen Leistungswerten zu vermeiden, ist eine Metrik gewählt, für die der Mittelwert auf 300 und die Standardabweichung auf 100 normiert wurden.

Die Datenaufbereitung und -skalierung erfolgte unter Nutzung der Software *SPSS* (IBM Corp., 2012) und der Software *ConQuest* (Wu, Adams, Wilson &

Haldane, 2007). Konkret wurde auf das im Kontext von TIMSS 1995 entwickelte *Mixed Coefficients Multinomial Logit Model* (MCMLM) von Adams, Wilson & Wang (1997) zurückgegriffen, welches ebenfalls der *Item Response Theory* (IRT; Boomsma, van Duijn & Snijders, 2001; van der Linden & Hambleton, 1997) zuzuordnen ist. Dieses Skalierungsmodell integriert die international in PIRLS und TIMSS verwendeten Skalierungsmodelle (Foy, Brossman & Galia, 2012) und wird beispielsweise auch zentral in PISA eingesetzt (OECD, 2012). Das MCMLM stellt eine mehrdimensionale Verallgemeinerung des Rasch-Modells (Fischer & Molenaar, 1995; Rasch, 1980) dar, welches in seiner Formulierung auch eine *latente Regression* von Leistungstestwerten auf Hintergrundvariablen mit einschließt (OECD, 2012). Diese Vorgehensweise ermöglicht es, die Kovarianzen zwischen den verschiedenen Kompetenzen und latente Zusammenhänge mit weiteren Merkmalen der Schülerinnen und Schüler für die Schätzung von Personenparametern (*Plausible Values*, PV) zu berücksichtigen (Mislevy, 1991; Mislevy, Beaton, Kaplan & Sheehan, 1992; vgl. auch Ünlü, Kasper & Trendtel, eingereichtes Manuskript). Für die Bildung des nationalen Hintergrundmodells wurden sämtliche Informationen aus den Fragebögen der Schülerinnen und Schüler, dem Elternfragebogen, der von den Lehrkräften ausgefüllten Schülerteilnahmeliste und dem Schulfragebogen sowie Kennwerte aus einem kognitiven Fähigkeitstest (Heller & Perleth, 2000) verwendet. Um die Fülle der Informationen beziehungsweise Variablen zu aggregieren, wurden die Hintergrunddaten für die latente Regression, wie in *Large Scale Assessments* im Allgemeinen üblich, mit der Hauptkomponentenmethode reskaliert, so dass die verbleibenden Variablen etwa 75 Prozent der Varianz der Ausgangsdaten erklären.

3.1.1 Modelle mit verschiedenen Dimensionen

Um die Datenstruktur der IGLU- und TIMSS-Tests abzubilden, wurden ein drei- und ein achtdimensionales MCMLM verwendet (vgl. Abbildung 9.3). Das dreidimensionale Modell bildet die Dimensionen Leseverständnis, Mathematik und Naturwissenschaften ab. Das achtdimensionale Modell differenziert die Kompetenzdomänen nach Textsorten (IGLU) und Inhaltsbereichen (TIMSS) und repräsentiert die Dimensionen (1) Lesen: *Informierende Texte*, (2) Lesen: *Literarische Texte*, (3) Mathematik: *Arithmetik*, (4) Mathematik: *Geometrie/Messen* und (5) Mathematik: *Umgang mit Daten* sowie (6) Naturwissenschaften: *Biologie*, (7) Naturwissenschaften: *Physik/Chemie* und (8) Naturwissenschaften: *Geographie*. Als einfachste Modellvariante wurde zudem ein eindimensionales Modell (mit latenter Regression) auf alle Aufgaben der Tests ohne Differenzierung nach Kompetenzdomänen betrachtet. Bei den anschließenden Modellvergleichen wird eher eine Präferenz für die mehrdimensionalen Modellvarianten zu erwarten sein, da davon auszugehen ist, dass sich die Kompetenzdimensionen in den Daten getrennt abbilden (Schwippert, Bos & Lankes, 2003). Vorerst wird das eindimensionale Modell aber in erster Linie als Referenzmodell für die Modelltestung genutzt. Abbildung 9.3 ist zu entnehmen, in welchem Verhältnis die unterschiedlichen Modellvarianten zueinander stehen: Die Linien verweisen darauf, dass die Modelle in einer hierarchischen Beziehung zueinander stehen, wobei ein spezielleres Modell (weiter unten) durch Vereinfachung (Modellrestriktionen) aus dem allgemeinen Modell abgeleitet werden kann.

Abbildung 9.3: Hierarchie der ein-, drei- und achtdimensionalen Kompetenzmodelle

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Die drei Modelle können anhand von Modellauswahlkriterien und, sofern die Voraussetzungen erfüllt sind, anhand eines *Likelihood Ratio Tests* (LRT) auf ihre Passung zur Datengrundlage miteinander verglichen werden. Sollten die Voraussetzungen nicht erfüllt sein, sind alternativ *Resampling*-Varianten dieser Teststatistik denkbar. Die Ergebnisse des Vergleichs der angepassten ein-, drei- und achtdimensionalen Modelle mittels Devianz und Informationskriterien finden sich in Tabelle 9.2.

Als Ergebnis lässt sich zunächst festhalten, dass unter Betrachtung der Auswahlkriterien *Consistent Akaike Information Criterion* (CAIC) und *Bayesian inference criterion* (BIC) weder das drei- noch das achtdimensionale Modell gegenüber dem eindimensionalen Modell bevorzugt werden können. Lediglich nach dem *Akaike Information Criterion* (AIC) ist das dreidimensionale Modell dem eindimensionalen vorzuziehen, wobei das achtdimensionale Modell für die übliche Modellauswahl zu komplex ist. Diese Feststellung ist jedoch dahingehend zu bewerten, dass die üblichen Auswahlkriterien wie AIC, CAIC und BIC (Burnham & Anderson, 2002; Myung, Forster & Browne, 2000) für die hier betrachtete Modellfamilie nicht unmittelbar einsetzbar zu sein scheinen. Die mit wachsender latenter Dimensionalität relativ schnell ansteigende

Tabelle 9.2: Ergebnisse des Modellvergleichs (Devianz und Informationskriterien AIC, CAIC und BIC)

Modell	Likelihood ¹	Anzahl Modellparameter	AIC	CAIC	BIC	n
1D-Modell Domänen und Subdomänen übergreifend	339004	472	339948	343382	342910	3928
3D-Modell Domänen bezogen	335726	1419	338564	348889	347470	3928
8D-Modell Subdomänen bezogen	334272	3804	341880	369558	365754	3928

¹ Hier als *Final Deviance*, $-2\log\{\text{maximierte Likelihood}\}$, im Sinne einer *ConQuest*-Ausgabe wiedergegeben.

Modellkomplexität (z.B. 472 zu schätzende Parameter versus 3 804 Parameter beim Übergang von ein- zu achtdimensionalem MCMLM), welche durch die Kriterien gegen den Modellfit ausbalanciert wird, fällt in dieser Modellfamilie sehr stark ins Gewicht. Der „Gewinn“ an *Likelihood* unter den allgemeinen Modellen wird durch eine zu große Anzahl an zu schätzenden Parametern unverhältnismäßig stark „bestraft“. Weitere Forschung wäre an dieser Stelle nötig, um geeignete Auswahlmaße in dieser Modellfamilie zu entwickeln. Auch empirisch plausible Restriktionen in den Modellparametern sind eine denkbare Möglichkeit, die zu einem späteren Zeitpunkt untersucht werden sollte.

Alternativ zu den Auswahlkriterien AIC, CAIC und BIC kann die Frage der Modellpassung jedoch auch „absolut“ durch Betrachtung von LRT-Statistiken bewertet werden. Während die Informationskriterien AIC, CAIC und BIC lediglich einen „relativen“ Vergleich auf das am besten geeignete Modell geben, hat der allgemeine LRT folgenden Vorteil: Er berücksichtigt die Differenzen der Devianzen und ermöglicht – sofern die Voraussetzungen erfüllt sind – auch eine Signifikanzprüfung (Burnham & Anderson, 2002; Myung et al., 2000; Rost, 2004). Vergleicht man die Modelle unter Nutzung von LRT-Statistiken, zeigt sich, dass sowohl das drei- als auch das achtdimensionale Modell gegenüber dem eindimensionalen Modell zu bevorzugen sind. Darüber hinaus verweist der LRT-Ansatz darauf, dass sowohl das drei- als auch das achtdimensionale Modell die Datenstruktur vergleichbar gut beschreiben.

3.1.2 Latente Korrelationen zwischen den Dimensionen

Neben der Etablierung einer gemeinsamen Metrik bietet die vorgestellte mehrdimensionale Modellierung der Leistungstests von IGLU und TIMSS den Vorteil, dass die Zusammenhänge beziehungsweise latenten Korrelationen zwischen den Kompetenzbereichen direkt ermittelt werden können. Die aus den Daten geschätzten latenten Korrelationen zwischen den Kompetenzdomänen und Inhaltsbereichen sind als Ergebnisse des drei- und des achtdimensionalen Modells in Tabelle 9.3 und in Tabelle 9.4 dargestellt.

Tabelle 9.3: Latente Korrelationen für das dreidimensionale domänenbezogene MCMLM: Leseverständnis, Mathematik und Naturwissenschaften

	Lesen	Mathematik	Naturwissenschaften
Lesen			
Mathematik	.54		
Naturwissenschaften	.74	.66	

Aus der Höhe der Korrelationen ist abzulesen, dass zwischen allen drei Bereichen substantielle Zusammenhänge bestehen und dass die Leistungen im naturwissenschaftlichen Test (.74) noch stärker als die im Mathematiktest (.54) mit dem Leseverständnis zusammenhängen. Dies ist auch auf das Aufgabenformat zurückzuführen, denn das Verständnis und die Lösung der naturwissenschaftlichen Aufgaben setzen das Leseverständnis in stärkerem Maße voraus. Der ermittelte Zusammenhang zwischen den Dimensionen Mathematik und Naturwissenschaften liegt bei .66.

Eine sinnvolle Vergleichsmöglichkeit dieser Ergebnisse ergibt sich durch die berichteten latenten Zusammenhänge in anderen Studien. Die im Rahmen von IGLU 2001 (Bos et al., 2003) ermittelten Korrelationen zwischen den verschiedenen Kompetenzdimensionen liegen auf einem ähnlichen Niveau: für das Leseverständnis mit den mathematischen und naturwissenschaftlichen Kompetenzen bei .60 beziehungsweise .64 und für mathematische und naturwissenschaftliche Kompetenzen bei .74. Vergleicht man die Korrelationen mit den in PISA ermittelten Zusammenhängen, zeigt sich, dass die für IGLU und TIMSS ermittelten Zusammenhänge geringer ausfallen als die Korrelationen in PISA. Der für PISA 2000 ermittelte Zusammenhang zwischen den fachübergreifenden Skalen Lesen, Mathematik und Naturwissenschaften fiel mit Korrelationen zwischen .82 und .89 vergleichsweise hoch aus (Adams & Carstensen, 2002). In PISA 2009 zeigen sich für die internationale Skalierung ebenfalls hohe latente Korrelationen zwischen den drei Domänen (Lesekompetenz/Mathematik: .84; Lesekompetenz/Naturwissenschaften: .87; Mathematik/Naturwissenschaften: .89) (OECD, 2012).

Ein differenzierteres Bild der IGLU/TIMSS-Daten ergibt sich, wenn man die Korrelationen zwischen den einzelnen Subdomänen der Tests betrachtet, die in Tabelle 9.4 dargestellt sind.

Beim Lesetest besteht eine hohe Korrelation (.78) zwischen den beiden erfassten Bereichen des literarischen und informierenden Lesens, was erklärt, dass diese beiden Subskalen jeweils in ähnlicher Höhe mit den Subskalen des Mathematiktests (zwischen .52 und .60) und des naturwissenschaftlichen Tests (hier deutlich höher zwischen .67 und .79) korrelieren. Die Leistungen in den drei Subskalen in Naturwissenschaften korrelieren derart stark miteinander (.95),

Tabelle 9.4: Latente Korrelationen für das achtdimensionale subdomänenbezogene MCMLM: *Informierende Texte* und *Literarische Texte* (Leseverständnis), *Arithmetik*, *Geometrie/Messen* und *Umgang mit Daten* (Mathematik), *Biologie*, *Physik/Chemie* und *Geographie* (Naturwissenschaften)

	Lesen		Mathematik			Naturwissenschaften		
	Literarisch	Informierend	Arithmetik	Geometrie/ Messen	Umgang mit Daten	Physik/ Biologie Chemie Geographie		
Lesen: Literarisch								
Lesen: Informierend		.78						
Mathematik: Arithmetik		.54	.52					
Mathematik: Geometrie/Messen		.55	.60	.88				
Mathematik: Umgang mit Daten		.54	.52	.83	.89			
Naturwissenschaften: Biologie		.75	.79	.62	.71	.72		
Naturwissenschaften: Physik/Chemie		.69	.67	.66	.76	.78	.95	
Naturwissenschaften: Geographie		.69	.68	.56	.71	.76	.95	.95

dass sie sich beinahe als Paralleltests füreinander interpretieren ließen. Auch die drei Subskalen des Mathematiktests korrelieren miteinander in einer beträchtlichen Höhe (zwischen .83 und .89). Betrachtet man die Zusammenhänge zwischen den mathematischen und den naturwissenschaftlichen Subskalen, so sind sie am geringsten in Kombination mit der Subskala *Arithmetik* (zwischen .56 und .66) und am größten in Kombination mit der Subskala *Umgang mit Daten* (zwischen .72 und .78). Die Betrachtung der Korrelationen der naturwissenschaftlichen Subskala *Biologie* zeigt, dass sich die größten Korrelationen mit den Subskalen im Lesen (.79 und .75) ergeben, während die Korrelationen mit den mathematischen Subskalen *Umgang mit Daten* (.72), *Geometrie/Messen* (.71) und *Arithmetik* (.62) geringer ausgeprägt sind. Die Subskala *Physik/Chemie* korreliert mit den mathematischen Subskalen *Umgang mit Daten* (.78) und *Geometrie/Messen* (.76) jeweils höher als mit den Subskalen im Lesen (.69 und .67). Auch für die naturwissenschaftliche Subskala *Geographie* zeigt sich, dass die Korrelationen zu den mathematischen Subskalen *Umgang mit Daten* (.76) und *Geometrie/Messen* (.71) jeweils höher ausgeprägt sind als die Korrelationen mit den Subskalen im Lesen (.69 und .68).

Diese korrelativen Ergebnisse können in gewissem Sinne als Plausibilitätschecks für den Modellvergleich in Abschnitt 3.1.1 herangezogen werden. Die sehr hohen Korrelationen zwischen den einzelnen Subskalen einer jeden Domäne und die relativ großen Korrelationen zwischen den Domänen untereinander untermauern die Beobachtung, dass drei Dimensionen zur Beschreibung der IGLU/TIMSS-Daten vergleichbar gut sind wie acht Dimensionen und dass darüber hinaus, nebst der geschilderten Problematik der Modellkomplexität, lediglich das dreidimensionale Modell, und nur unter dem Auswahlkriterium AIC, gegenüber dem eindimensionalen Modell präferiert wird.

3.2 Ergebnisse der latenten Profilanalyse

Über die berichteten latenten Korrelationen hinausgehend bieten latente Profilanalysen auch die Möglichkeit, den Zusammenhang zwischen den verschiedenen Skalen oder Subskalen von IGLU und TIMSS zu untersuchen und Gruppen von Schülerinnen und Schülern mit unterschiedlichen Leistungs- oder Kompetenzprofilen zu ermitteln. Bei diesem Verfahren werden Kinder mit vergleichbaren Leistungsständen in den einzelnen Kompetenzdimensionen in homogene Typen gruppiert, in denen jeweils Schülerinnen und Schüler mit ähnlichen Leistungsprofilen vorzufinden sind; beispielsweise jene, die über alle acht inhaltlichen Subdomänen der drei Kompetenzbereiche hinweg sehr hohe beziehungsweise sehr geringe Leistungsergebnisse erzielen. Die statistische Methode, die hierbei zum Einsatz kommt, ist die *latente Profilanalyse* (LPA), die von der Idee ähnlich der *Latent Class Analyse* ist, jedoch mit stetigen manifesten Indikatoren und Verteilungen arbeitet (Bacher, Pöge & Wenzig, 2010; Muthén & Muthén, 2012; Vermunt & Magidson, 2005). Die LPA wurde basierend auf den geschätzten PV-Leistungsergebnissen der Grundschülerinnen und Grundschüler in den einzelnen Domänen beziehungsweise Subdomänen unter Nutzung des Softwareprogramms *Latent GOLD* (Vermunt & Magidson, 2005) durchgeführt. Zur Berechnung der Signifikanzen von Unterschieden zwischen den Typen mit ähnlichen Leistungsprofilen wurden die Standardfehler mit Hilfe des sogenannten *Jackknife*-Verfahrens (vgl. Kapitel 2 in diesem Band) bestimmt und anschließend *t*-Tests mit einer Irrtumswahrscheinlichkeit von $\alpha = .05$ zugrunde gelegt.

Tabelle 9.5: Vergleichswerte BIC und CAIC über verschiedene Anzahlen an identifizierten Profilen

Modell	Log-Likelihood	BIC	CAIC	Parameter
I. ein Profil	-70525	141099	141105	6
II. zwei Profile	-67848	135803	135816	13
III. drei Profile	-66436	133037	133057	20
IV. vier Profile	-65691	131605	131632	27
V. fünf Profile	-65346	130973	131007	34
VI. sechs Profile	-65179	130696	130737	41
VII. sieben Profile	-65066	130529	130577	48
VIII. acht Profile	-65039	130532	130587	55
IX. neun Profile	-65011	130536	130598	62
X. zehn Profile	-65012	130594	130663	69

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

In Tabelle 9.5 sind die Ergebnisse für die berechneten LPA-Modelle mit unterschiedlich vielen (bis zu zehn) spezifizierten Leistungsprofilen aufgeführt. Dabei erweist sich nach BIC als auch CAIC die 7-Leistungsprofile-Lösung (Modell VII) als die vergleichsweise beste, den Leistungswerten zugrunde liegende Lösung. Eine Überprüfung der Schwellenparameter (Mittelwerte M , berechnet in der Stichprobe) der sieben Leistungsprofile verdeutlicht, dass in allen sieben Profilklassen die Schwellen der einbezogenen Leistungswerte über jede der drei Domänen einheitlich die gleiche Ordnung aufweisen. Somit enthält diese Lösung keine Klasse, die von der erwartungsgemäßen Rangreihung in einer der Domänen abweicht. Es liegt eine Art „invariante Profilordnung“ über die Kompetenzbereiche hinweg vor (vgl. auch Croon, 1990, 1991; Sijtsma & Molenaar, 2002; Vermunt, 2001). Dies untermauert erneut die Niederdimensionalität der vorliegenden Daten als ein weiterer Plausibilitätscheck des vorgestellten Modellvergleichs. Die 7-Profil-Lösung weist gute Messeigenschaften auf und wird somit für die nachfolgenden interpretativen Betrachtungen genutzt. In der Beschreibung der resultierenden Leistungsprofile wird im Folgenden von Leistungstypen gesprochen.

In Tabelle 9.6 sind die Anteile und Leistungsmittelwerte der sieben Leistungstypen dargestellt. Die größten Anteile der Schülerinnen und Schüler in der Grundschule in Deutschland gehören mit etwa einem Viertel (25.5%) dem Leistungstyp 4 mit durchschnittlichen Leistungen um den Skalenmittelwert, mit

Tabelle 9.6: Die Leistungsprofile der Grundschülerinnen und Grundschüler in Deutschland, bezogen auf die drei Kompetenzdomänen (berechnet in der Stichprobe)

Typ	n	%	Gesamtskala Lesen				Gesamtskala Mathematik				Gesamtskala Naturwissenschaften			
			M	(SE)	SD	(SE)	M	(SE)	SD	(SE)	M	(SE)	SD	(SE)
Typ 7	165	4.3	471	(4.4)	46	(2.3)	482	(4.8)	55	(3.1)	473	(4.7)	50	(4.5)
Typ 6	533	13.5	410	(2.3)	45	(1.9)	400	(3.2)	58	(2.1)	409	(2.8)	50	(2.8)
Typ 5	900	22.8	355	(2.1)	43	(1.5)	344	(3.0)	59	(2.5)	350	(2.5)	51	(1.7)
Typ 4	1003	25.5	298	(2.0)	44	(1.4)	294	(2.6)	60	(1.8)	298	(2.7)	52	(1.6)
Typ 3	753	18.9	238	(1.7)	45	(1.4)	243	(3.0)	60	(2.0)	242	(2.2)	49	(1.8)
Typ 2	444	11.3	164	(3.3)	51	(2.9)	186	(3.6)	57	(3.1)	172	(3.9)	49	(2.8)
Typ 1	130	3.8	79	(5.3)	60	(3.8)	97	(8.0)	60	(4.8)	83	(6.0)	55	(5.2)

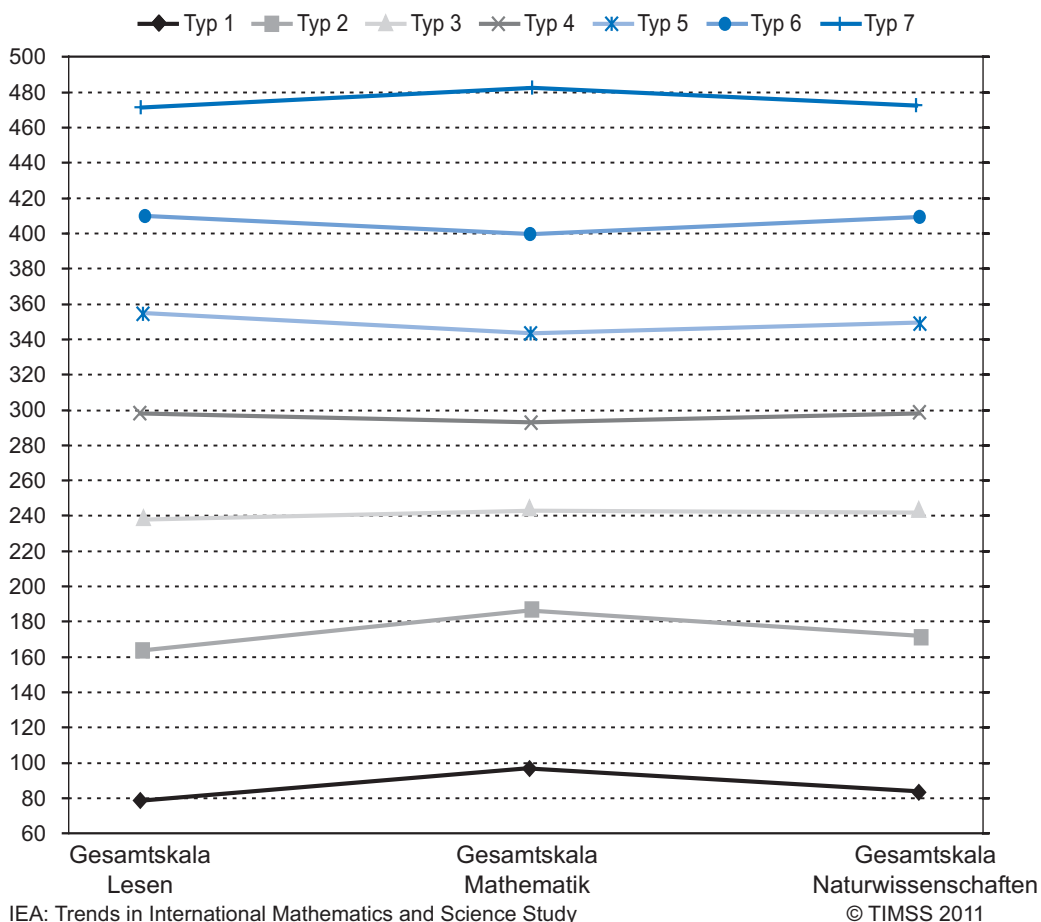
IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

etwas über einem Fünftel (22.8%) dem Leistungstyp 5 mit überdurchschnittlichen und mit etwas unter einem Fünftel (18.9%) dem Leistungstyp 3 mit unterdurchschnittlichen Leistungen an. Das verbleibende Drittel der Kinder teilt sich auf die Leistungstypen mit sehr hohen (Typ 6) beziehungsweise geringen Leistungen (Typ 2) sowie auf die Extremgruppen (Typ 7 bzw. Typ 1) auf. Bei Betrachtung der mittleren Leistungen fällt auf, dass sich diese für die einzelnen Leistungstypen deutlich abgrenzen lassen und sie von einem Leistungstyp zum nächsten ansteigen. Wie in Abbildung 9.4 illustriert, ergibt sich daraus ein eher paralleler Verlauf der Leistungen über die verschiedenen Kompetenzdomänen hinweg, das heißt höher liegende Leistungstypen weisen jeweils höhere mittlere Leistungen in allen Kompetenzbereichen auf.

Aus Tabelle 9.6 geht weiterhin hervor, dass die Unterschiede zwischen den Leistungsmittelwerten in den drei betrachteten Domänen innerhalb eines Leistungstyps jeweils geringer sind als die Unterschiede im Vergleich zu den Leistungsmittelwerten anderer Typen. Insgesamt ergibt sich damit hinsichtlich der Leistungsmittelwerte der Kompetenzdomänen ein eher homogenes Bild innerhalb der jeweiligen Leistungstypen, während zwischen den Typen Heterogenität vorliegt. Dennoch bestehen Unterschiede innerhalb der Leistungstypen zwischen den verschiedenen Kompetenzmittelwerten, die je nach Leistungstyp unterschiedlich stark ausgeprägt sind. In Abbildung 9.4 sind die Leistungsprofile der Grundschülerinnen und Grundschüler in Deutschland über die drei Kompetenzbereiche anhand der Leistungsmittelwerte (in der Stichprobe) dargestellt.

Abbildung 9.4: Die Leistungsprofile der Grundschülerinnen und Grundschüler in Deutschland, bezogen auf die drei Kompetenzbereiche (berechnet in der Stichprobe)



Die Leistungsmittelwerte des anteilig größten Leistungstyps (Typ 4) mit etwa einem Viertel der Schülerinnen und Schüler liegen sehr nah am Skalenmittelwert von 300 Punkten. Dem Leistungstyp 5 werden 22.8 Prozent der Kinder zugeordnet, mit Leistungswerten von etwa einer halben Standardabweichung über dem Skalenmittelwert (344 bis 355 Punkte). Eine weitere vergleichsweise große Gruppe, den Leistungstyp 3, bilden 18.9 Prozent der Viertklässlerinnen und Viertklässler mit Leistungswerten von mehr als einer halben Standardabweichung unter dem Skalenmittelwert (238 bis 243 Punkte). Die Leistungstypen 3 und 4 weisen ein eher ausgeglichenes Profil in den drei Leistungsbereichen auf, wenngleich sie sich im Niveau der Leistungen voneinander unterscheiden. Für Leistungstyp 5 mit über dem Durchschnitt liegenden Leistungen zeigt sich im Vergleich dazu ein eher heterogeneres Profil: Für diesen Leistungstyp ergeben sich höhere Mittelwerte für das Lesen als für Naturwissenschaften und insbesondere im Vergleich zum Bereich Mathematik.

An den beiden Extremen der Leistungsskala finden sich Leistungstypen mit deutlich unausgewogeneren Leistungsprofilen, wobei diese Gruppen vergleichsweise klein ausfallen: Leistungstyp 1 mit 3.8 Prozent und Leistungstyp 2 mit 11.3 Prozent der Schülerinnen und Schüler. Ihre Leistungen im Lesen liegen unter den Leistungen in Naturwissenschaften und – noch deutlicher – unter denen in Mathematik. Zu den leistungsstärksten Typen zählen 13.5 Prozent (Typ 6) und 4.3 Prozent (Typ 7) der Schülerinnen und Schüler. Schülerinnen und Schüler des Leistungstyps 7 erzielen herausragende Leistungen, wobei hier relative Stärken im Bereich Mathematik bestehen. Schülerinnen und Schüler des Leistungstyps 6 erzielen insgesamt sehr hohe Leistungen, jedoch stellt Mathematik eine relative Schwäche dar.

Eine differenziertere Betrachtung als die vorstehende erlauben die post-LPA deskriptiven Befunde, die sich basierend auf dieser 7-Leistungsprofile-Lösung für die acht Subskalen aus dem Lese-, Mathematik- und naturwissenschaftlichen Test ergeben und die nachfolgend in Abschnitt 4 diskutiert werden.

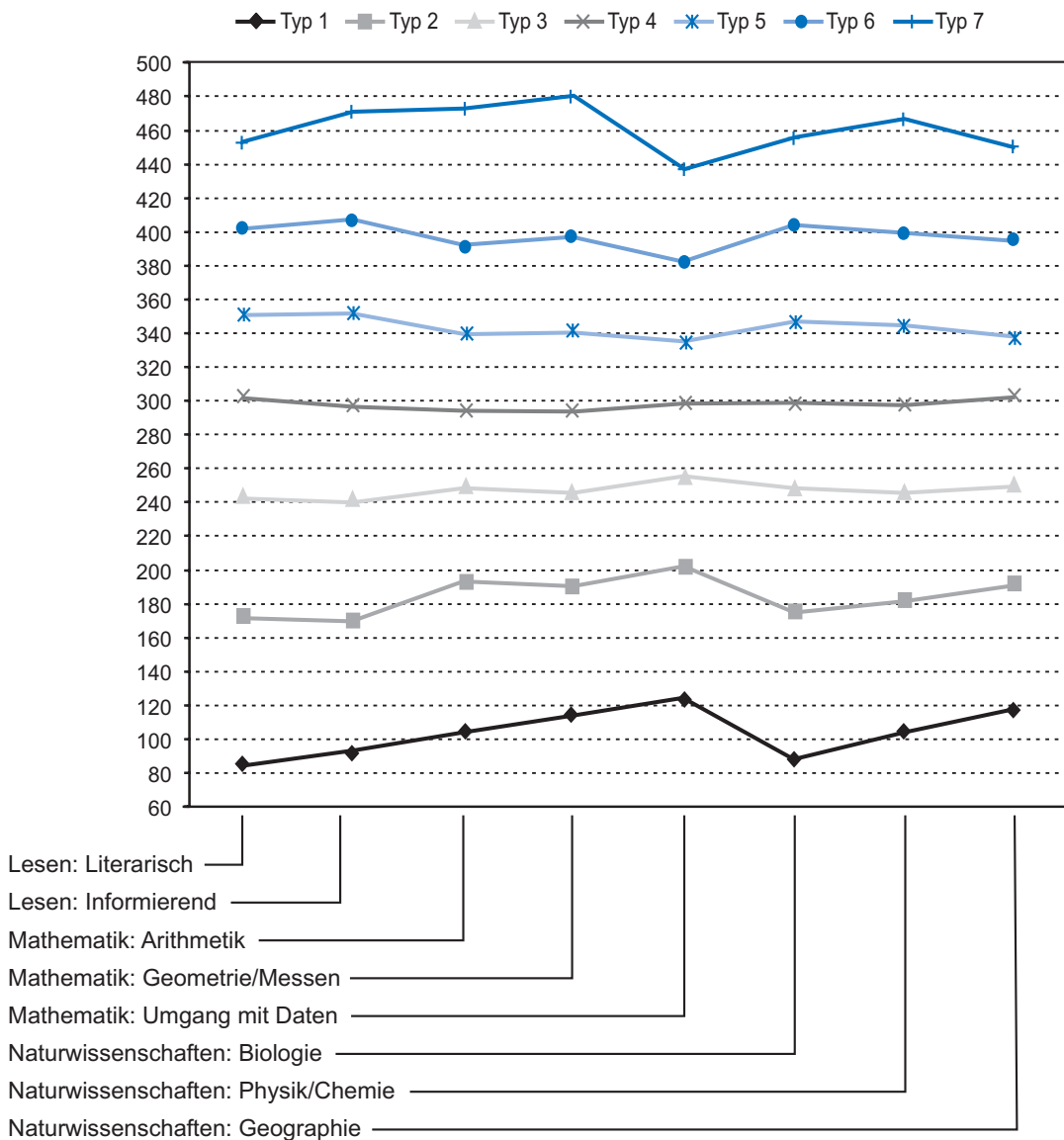
4 Beschreibung der Leistungstypen

Die sieben Leistungstypen, die in der latenten Profilanalyse identifiziert werden konnten, sollen im Folgenden genauer beschrieben werden. Dabei interessiert zunächst das Profil der Leistungen in den drei Kompetenzdomänen Leseverständnis, Mathematik und Naturwissenschaften sowie in den inhaltsbezogenen Subdomänen (Abschnitt 4.1). Nachfolgend werden die hier gebildeten Leistungstypen in Bezug auf relevante Hintergrundmerkmale der Schülerinnen und Schüler betrachtet, die mit schulischen Leistungen im Zusammenhang stehen (Abschnitt 4.2). Abschließend werden die fachbezogenen Einstellungen und Selbstkonzepte der Schülerinnen und Schüler in den Blick genommen (Abschnitt 4.3).

4.1 Beschreibung der Leistungstypen nach der durchschnittlichen Leistung in den drei Kompetenzdomänen und ihren inhaltsbezogenen Subdomänen

Die Beschreibung der durchschnittlichen Leistung in den drei Kompetenzdomänen Leseverständnis, Mathematik und Naturwissenschaften sowie in ihren inhaltsbezogenen Subdomänen erfolgt auf Grundlage von Leistungswerten, die in der oben beschriebenen dreidimensionalen Skalierung mit einem nationalen Hintergrundmodell geschätzt wurden (vgl. Abschnitt 3.1). Wie bereits erwähnt, wurde dabei eine neue Metrik verwendet (mit einem Mittelwert von 300 und einer Standardabweichung von 100), um die so ermittelten Leistungswerte von den Werten aus der internationalen Skalierung (Mittelwert von 500, Standardabweichung von 100) abzugrenzen. In Abbildung 9.5 sind die Leistungsmittelwerte der sieben Profile in den acht Inhaltsbereichen der

Abbildung 9.5: Die Leistungsprofile der Grundschülerinnen und Grundschüler in Deutschland, bezogen auf die acht inhaltlichen Subdomänen der drei Kompetenzbereiche (berechnet in der Stichprobe, deskriptiv basierend auf der in Abschnitt 3 gewonnenen 7-Profile-Lösung)



drei Kompetenzdomänen dargestellt. Die Beschreibung der Typen beginnt mit der leistungsstärksten Gruppe und endet mit der Gruppe mit den geringsten Leistungswerten.

4.1.1 Schülerinnen und Schüler mit hohen Leistungen

Typ 7: Schülerinnen und Schüler mit herausragenden Leistungen, insbesondere in Mathematik

Der Leistungstyp 7 repräsentiert etwa vier Prozent der Schülerinnen und Schüler in der vierten Jahrgangsstufe in Deutschland. Die Schülerinnen und Schüler dieses Leistungstyps erreichen herausragende mittlere Leistungen von 471 Punkten im Lesen, 482 Punkten in Mathematik und 473 Punkten in den Naturwissenschaften. Bei Betrachtung der Leistungen in den acht Subdomänen wird ersichtlich, dass sich vergleichsweise große Unterschiede zeigen. Damit ergibt sich für die Gruppe der Schülerinnen und Schüler mit herausragenden Leistungen ein heterogeneres Profil über die acht inhaltlichen Subdomänen als für die Leistungstypen mit unter- bis überdurchschnittlichen Leistungen (Typen 3, 4, 5). Allerdings gilt es zu beachten, dass sich aufgrund der geringeren Schüleranzahl für diesen Typ höhere Variationen einstellen können. Relative Leistungsstärken ergeben sich für die Schülerinnen und Schüler dieses Leistungstyps in den mathematischen Inhaltsbereichen *Geometrie/Messen* und *Arithmetik*, gefolgt vom Leseverständnis bei *informierenden Textsorten* und dem naturwissenschaftlichen Inhaltsbereich *Physik/Chemie*, während die Leistungen im Inhaltsbereich *Umgang mit Daten* im Verhältnis niedrigere Werte aufweisen.

Typ 6: Schülerinnen und Schüler mit sehr hohen Leistungen, insbesondere im Lesen und den Naturwissenschaften

Der Leistungstyp 6 repräsentiert 13,5 Prozent der Schülerinnen und Schüler der vierten Jahrgangsstufe in Deutschland. Die Schülerinnen und Schüler dieses Typs erreichen mittlere Leistungen von 410 Punkten im Lesen, 400 Punkten in Mathematik und 409 Punkten in den Naturwissenschaften. Ihre Leistungen sind damit im Vergleich zum Skalenmittelwert von 300 als sehr hoch einzuordnen, liegen jedoch deutlich unterhalb der Mittelwerte des Leistungstyps mit herausragenden Leistungen (Typ 7). Auch für diesen Leistungstyp zeigen sich Unterschiede zwischen den acht inhaltlichen Subdomänen, auch wenn diese nicht so deutlich sind, wie bei den Leistungstypen mit herausragenden beziehungsweise sehr geringen Leistungen (Typ 7 und Typ 1). Damit ergibt sich auch für die Gruppe der Schülerinnen und Schüler mit sehr hohen Leistungen ein heterogeneres Profil über die acht inhaltlichen Subdomänen als für die Leistungstypen mit unter- bis überdurchschnittlichen Leistungen (Typen 3, 4, 5). Die höchsten Leistungswerte erzielen die Schülerinnen und Schüler dieses Leistungstyps im Lesen und in den naturwissenschaftlichen Inhaltsbereichen; in den mathematischen Inhaltsbereichen sind ihre Leistungen vergleichsweise schwächer, liegen aber immer noch deutlich über dem Skalenmittelwert.

4.1.2 Schülerinnen und Schüler mit mittleren Leistungen

Typ 5: Schülerinnen und Schüler mit überdurchschnittlichen Leistungen auf ähnlichem Niveau

Der Leistungstyp 5 repräsentiert etwa ein Fünftel (22.8%) der Schülerinnen und Schüler in der vierten Jahrgangsstufe in Deutschland und zählt damit zu den drei größten Leistungsgruppen. Schülerinnen und Schüler dieses Typs erreichen mittlere Leistungen von 355 Punkten im Lesen, 344 Punkten in Mathematik und 350 Punkten in den Naturwissenschaften. Ihre Leistungen sind damit im Vergleich zum Skalenmittelwert von 300 noch als überdurchschnittlich einzuordnen, liegen jedoch deutlich unterhalb derer der Leistungstypen mit sehr hohen beziehungsweise herausragenden Leistungen (Typ 6 und Typ 7). Hinsichtlich der acht inhaltlichen Subdomänen ergibt sich für den Leistungstyp mit überdurchschnittlichen Leistungen ein relativ homogenes Bild. Am unteren Ende der Leistungsspanne liegt der Inhaltsbereich *Umgang mit Daten* und am oberen Ende befinden sich die Lesekompetenzen bei *literarischen* und *informierenden* Texten.

Typ 4: Schülerinnen und Schüler mit durchschnittlichen Leistungen in allen Bereichen

Der Leistungstyp 4 repräsentiert etwa ein Viertel der Schülerinnen und Schüler in der vierten Jahrgangsstufe in Deutschland und bildet damit die größte Gruppe. Schülerinnen und Schüler dieses Typs erreichen mittlere Leistungen von 298 Punkten im Lesen, 294 Punkten in Mathematik und 298 Punkten in den Naturwissenschaften. Ihre Leistungen sind damit im Vergleich zum Skalenmittelwert von 300 als durchschnittlich einzustufen. Hinsichtlich der acht inhaltlichen Subdomänen ergibt sich für den Leistungstyp mit durchschnittlichen Leistungen ein sehr homogenes Bild. Am unteren Ende der Leistungsspanne liegen die Inhaltsbereiche *Arithmetik* und *Geometrie/Messen* und am oberen Ende die Inhaltsbereiche *Geographie* sowie das Leseverständnis beim Lesen *literarischer Texte*.

Typ 3: Schülerinnen und Schüler mit unterdurchschnittlichen Leistungen auf ähnlichem Niveau

Der Leistungstyp 3 repräsentiert etwa ein Fünftel der Schülerinnen und Schüler der vierten Jahrgangsstufe in Deutschland und zählt damit zu den drei größten Leistungsgruppen. Schülerinnen und Schüler dieses Typs erreichen mittlere Leistungen von 238 Punkten im Lesen, 243 Punkten in Mathematik und 242 Punkten in den Naturwissenschaften. Ihre Leistungen sind damit im Vergleich zum Skalenmittelwert von 300 als unterdurchschnittlich einzuordnen, liegen jedoch deutlich über denen der Leistungstypen mit geringen beziehungsweise sehr geringen Leistungen (Typ 2 und Typ 1). Hinsichtlich der acht inhaltlichen Subdomänen ergibt sich für den Leistungstyp 3 mit unterdurchschnittlichen Leistungen ein relativ homogenes Bild. Am unteren Ende der Leistungsspanne liegt die Lesekompetenz beim Lesen von *informierenden* und *literarischen* Texten und am oberen Ende der Leistungsspanne befinden sich die Inhaltsbereiche *Geographie* und *Umgang mit Daten*.

4.1.3 Schülerinnen und Schüler mit geringen Leistungen

Typ 2: Schülerinnen und Schüler mit geringen Leistungen, insbesondere im Lesen

Der Leistungstyp 2 repräsentiert etwa 11 Prozent der Schülerinnen und Schüler der vierten Jahrgangsstufe in Deutschland. Die Schülerinnen und Schüler dieses Typs erreichen geringe mittlere Leistungen von 164 Punkten im Lesen, 186 Punkten in Mathematik und 172 Punkten in den Naturwissenschaften. Ihre Leistungen sind damit im Vergleich zum Skalenmittelwert von 300 als niedrig einzuordnen, liegen jedoch deutlich über denen des Leistungstyps mit sehr geringen Leistungen (Typ 1).

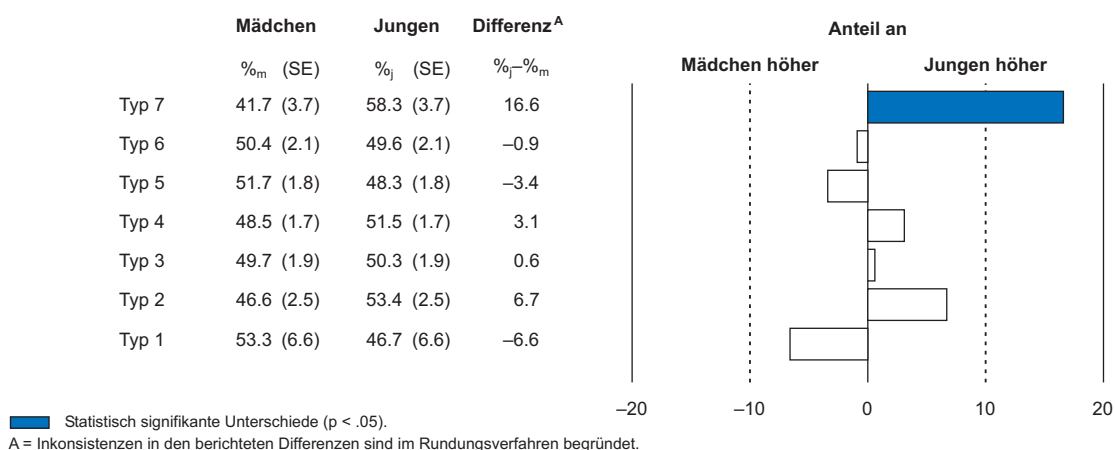
Für Leistungstyp 2 zeigen sich Unterschiede zwischen den acht inhaltlichen Subdomänen, auch wenn diese nicht so deutlich sind, wie bei den Leistungstypen mit herausragenden beziehungsweise sehr geringen Leistungen (Typ 7 und Typ 1). Für die Gruppe der Schülerinnen und Schüler mit geringen Leistungen ergibt sich ein heterogeneres Profil über die acht inhaltlichen Subdomänen als für die Leistungstypen mit unter- bis überdurchschnittlichen Leistungen (Typen 3, 4, 5). Die höchsten Leistungswerte erzielen die Schülerinnen und Schüler dieses Leistungstyps in den mathematischen Inhaltsbereichen, gefolgt von den naturwissenschaftlichen Inhaltsbereichen. Die niedrigsten Leistungswerte erzielen Schülerinnen und Schüler dieses Typs im Lesen.

Typ 1: Schülerinnen und Schüler mit sehr geringen Leistungen, insbesondere im Lesen

Der Leistungstyp 1 repräsentiert etwa vier Prozent der Schülerinnen und Schüler der vierten Jahrgangsstufe in Deutschland. Schülerinnen und Schüler dieses Typs erreichen mittlere Leistungen von 79 Punkten im Lesen, 97 Punkten in Mathematik und 83 Punkten in den Naturwissenschaften. Ihre Leistungen sind damit als außerordentlich schwach einzuordnen. Bei Betrachtung der Leistungen in den acht Subdomänen wird ersichtlich, dass sich vergleichsweise große Unterschiede zeigen. Damit ergibt sich auch für die Gruppe der Schülerinnen und Schüler mit sehr geringen Leistungen ein heterogeneres Profil über die acht inhaltlichen Subdomänen als für die Leistungstypen mit unter- bis überdurchschnittlichen Leistungen (Typen 3, 4, 5). Allerdings gilt es auch hier zu beachten, dass sich aufgrund der geringeren Schüleranzahl für diesen Leistungstyp höhere Variationen einstellen können. Relative Leistungsstärken ergeben sich für die Schülerinnen und Schüler dieses Typs im mathematischen Inhaltsbereich *Umgang mit Daten* und im naturwissenschaftlichen Inhaltsbereich *Geographie*, während der Leistungstyp mit sehr geringen Leistungen im Inhaltsbereich *Biologie* und im Lesen verhältnismäßig die niedrigsten Leistungswerte aufweist.

4.2 Beschreibung der Leistungstypen nach relevanten Hintergrundmerkmalen

In diesem Abschnitt wird der Frage nachgegangen, ob und inwieweit sich die unterschiedlichen Typen in Bezug auf die Geschlechterverteilung sowie hinsichtlich ausgewählter Merkmale des kulturellen und sozialen Hintergrunds, die gemeinhin mit schulischen Leistungen korrelieren, unterscheiden. Abbildung 9.6 sind die Anteile von Mädchen und Jungen in den sieben Leistungstypen zu entnehmen. In der Graphik im rechten Teil der Abbildung sind die Differenzen jeweils durch

Abbildung 9.6: Zusammensetzung der sieben Leistungstypen nach Geschlecht (berechnet in der Stichprobe, deskriptiv basierend auf der in Abschnitt 3 gewonnenen 7-Profil-Lösung)

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

einen Balken dargestellt, wobei die Färbung des Balkens auf einen signifikanten Unterschied verweist.

Signifikante Unterschiede zwischen den Anteilen an Mädchen und Jungen zeigen sich einzig mit einer Differenz von 16.6 Prozentpunkten beim Leistungstyp 7, der sich durch herausragende Leistungen in den drei Bereichen und besonders in Mathematik auszeichnet. Der Anteil an Jungen beträgt hier 58.3 Prozent, der der Mädchen nur 41.7 Prozent. Dieses Ergebnis (basierend auf simultaner Skalierung der Dimensionen) spiegelt sich auch in der Kompetenzstufenzugehörigkeit (basierend auf separater Skalierung der Dimensionen) wider: Auf der obersten Kompetenzstufe V ist der Anteil der Jungen in Mathematik und Naturwissenschaften deutlich höher als der der Mädchen (vgl. Kapitel 6 in diesem Band). Im Lesen zeigen sich hingegen keine signifikanten Unterschiede, gleichermaßen viele Mädchen und Jungen erzielten Leistungen, die dem Niveau der obersten Kompetenzstufe entsprechen. Bei den unteren Leistungstypen 1 und 2 ergeben sich zwar nominelle Unterschiede hinsichtlich des Geschlechterverhältnisses, allerdings sind die Unterschiede nicht signifikant.

Im Folgenden werden die kulturellen und sozialen Hintergrundmerkmale betrachtet, die in IGLU und TIMSS durch den Eltern- und den Schülerfragebogen erhoben wurden. Die kulturellen und sozioökonomischen Merkmale wurden anhand der folgenden Indikatoren beschrieben: Anteil von Kindern aus Familien der dritten Sozialschicht (Akademiker, Techniker und Führungskräfte), Anteil von Kindern aus Familien, in denen mindestens ein Elternteil einen tertiären Bildungsabschluss erreicht hat (ISCED-Level 5A und 6), Anteil von Kindern aus Familien mit mehr als 100 Büchern sowie der Anteil von Kindern, die von Armut betroffen sind (vgl. zur Bildung der Indikatoren Kapitel 7 in diesem Band und Wendt, Stubbe & Schwippert, 2012). Migrationshintergrund und Familiensprache werden anhand des Anteils der Kinder, deren Eltern im Ausland geboren sind, sowie des Anteils der Kinder, die zu Hause nur manchmal oder nie Deutsch sprechen, dargestellt (vgl. vertiefend Kapitel 8 in diesem Band und Schwippert, Wendt & Tarelli, 2012).

Tabelle 9.7 zeigt die Ergebnisse bezüglich der Zusammensetzung der sieben Leistungstypen nach kulturellen und sozioökonomischen Merkmalen, Migrationshintergrund und Familiensprache. Aufgenommen in die Tabelle

Tabelle 9.7: Zusammensetzung der 7 Leistungstypen nach kulturellen und sozioökonomischen Merkmalen, Migrationshintergrund und Familiensprache sowie besonderem Förderbedarf (berechnet in der Stichprobe, deskriptiv basierend auf der in Abschnitt 3 gewonnenen 7-Profil-Lösung)

	Typ 1	Typ 2	Typ 3	Typ 4	Typ 5	Typ 6	Typ 7
	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)
Kulturelle und sozioökonomische Merkmale							
Familie mit hohem sozioökonomischem Status ^A	11.4 (3.7)	12.4 (2.0)	16.0 (1.5)	26.0 (1.8)	39.2 (2.0)	51.6 (2.5)	62.0 (4.5)
Familie mit hohem Bildungsniveau ^B	3.0 (1.8)	9.9 (1.8)	11.3 (1.4)	20.6 (1.8)	33.2 (2.2)	53.8 (2.4)	63.9 (4.5)
Bedeutsamer Buchbesitz in der Familie ^C	18.1 (5.2)	14.4 (2.3)	20.0 (1.5)	31.5 (2.2)	43.3 (1.9)	53.0 (2.7)	72.5 (3.8)
armutsgefährdete Familie	50.3 (5.1)	46.4 (3.4)	39.1 (2.3)	25.3 (1.9)	19.1 (1.6)	10.0 (1.6)	6.0 (2.0)
Migrationshintergrund und Familiensprache							
Migrationshintergrund: beide Elternteile	36.8 (5.5)	36.4 (3.1)	22.0 (1.8)	15.1 (1.6)	10.7 (1.1)	5.9 (1.2)	0.7 (0.7)
Familiensprache ist nicht deutsch ^D	31.5 (5.4)	39.1 (2.9)	25.5 (2.1)	18.1 (1.5)	14.7 (1.3)	10.2 (1.3)	4.8 (1.7)
Besonderer Förderbedarf							
Lernschwächen oder körperliche Schwächen	39.4 (7.3)	13.5 (2.2)	4.8 (1.0)	3.0 (0.6)	0.9 (0.3)	0.9 (0.4)	1.0 (0.9)

A = Sozialschicht nach Angabe der Eltern (0= Manuell Tätige, Angestellte und kleinere Unternehmer; 1= Akademiker, Techniker und Führungskräfte)

B = Bildungsniveau nach höchstem Bildungsabschluss der Eltern (0= kein Elternteil mit (Fach-)Hochschulabschluss; 1= mindestens ein Elternteil mit (Fach-)Hochschulabschluss)

C = Heimischer Buchbesitz nach Angabe der Eltern (0= maximal 100 Bücher; 1= mehr als 100 Bücher)

D = Familiensprache nach Angabe der Schülerinnen und Schüler (0= Deutsch: immer oder fast immer; 1= Deutsch: manchmal oder nie)

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

sind auch die Anteile von Kindern mit besonderem Förderbedarf im Bereich des Lernens oder der körperlichen und motorischen Entwicklung, die über die Angaben der Lehrkräfte ermittelt wurden.

Insgesamt gesehen ist Tabelle 9.7 zu entnehmen, dass so gut wie alle Werte für die erfassten Indikatoren von Leistungstyp 1 bis 7 kontinuierlich günstiger werden. Exemplarisch wird dies durch eine Gegenüberstellung der Extremgruppen aufgezeigt (Typ 7 vs. Typen 1 und 2).

Schülerinnen und Schüler mit herausragenden Leistungen (Typ 7) weisen in allen Hintergrundmerkmalen die günstigsten Werte auf und unterscheiden sich darin in beinahe allen Vergleichen signifikant von allen anderen Typen. In keinem anderen Leistungstyp ist der Anteil der Schülerinnen und Schüler, die mindestens einen Elternteil haben, welcher einen akademischen oder technischen Beruf oder einen Beruf in einer Position als Führungskraft ausübt, höher (62.0%). Am höchsten ist dementsprechend auch der Anteil von Eltern, von denen mindestens ein Elternteil einen Fachhochschul- oder höheren Abschluss erlangt hat (63.9%), und mit 6 Prozent ist der Anteil der Kinder, die von Armut gefährdet sind, in diesem Typ am geringsten. Auch der Anteil der Schülerinnen und Schüler, die laut Schülerangabe aus Familien stammen, die mehr als 100 Bücher zu Hause besitzen (72.5%), ist in diesem Leistungstyp am höchsten.

In allen hier genannten Merkmalen werden die Werte ungünstiger, je niedriger das Leistungsniveau der Schülerinnen und Schüler in den unterschiedlichen Leistungstypen ist. Die ungünstigsten Werte finden sich in den Typen 1 und 2, den Gruppen von Kindern mit geringen beziehungsweise sehr geringen Leistungen in allen Domänen und Inhaltsbereichen. Nur bei 11.4 Prozent (Typ 1) und 12.4 Prozent (Typ 2) der Schülerinnen und Schüler gehören Vater oder Mutter der dritten Sozialschicht an (Akademiker, Techniker und Führungskräfte). Lediglich ein geringer Anteil dieser Kinder hat einen Elternteil, welcher mindestens einen Fachhochschul- oder höheren Abschluss erlangt hat (3.0% bei Typ 1 und 9.9% bei Typ 2). Mit 50.3 Prozent (Typ 1) beziehungsweise

se 46.4 Prozent (Typ 2) ist der Anteil der Kinder, die von Armut gefährdet sind, in diesen Gruppen am höchsten. Mehr als 100 Bücher zu Hause haben lediglich 14.4 Prozent (Typ 2) und 18.1 Prozent (Typ 1) der Schülerinnen und Schüler. Während sich Leistungstyp 1 und 2 zusammen betrachtet in einigen Vergleichen der prozentualen Verteilung der Schülerinnen und Schüler nach den oben beschriebenen Hintergrundmerkmalen nicht signifikant vom Typ 3, dem nächstgelegenen Leistungstypen mit höheren Leistungswerten unterscheiden, weisen sie insgesamt signifikant ungünstigere Hintergrundmerkmale im Vergleich zu allen anderen Typen mit höheren Leistungswerten auf.

Auch in Bezug auf den Migrationshintergrund und die Familiensprache unterscheiden sich die sieben Leistungstypen, wobei auch fast alle Unterschiede (mit Ausnahme von Typ 1 vs. Typ 2; und für die Familiensprache auch mit Ausnahme von Typ 1 vs. Typ 3) signifikant sind. Schülerinnen und Schüler, deren Eltern im Ausland geboren sind, sind mit einem Anteil von unter einem Prozent in Leistungstyp 7, der Gruppe der Schülerinnen und Schüler mit herausragenden Leistungen in allen Kompetenzbereichen, deutlich unterrepräsentiert. Auch in den Leistungstypen mit überdurchschnittlichen (Typ 5: 10.7%) und sehr hohen Leistungen (Typ 6: 5.9%) liegt der Anteil an Schülerinnen und Schülern, deren Eltern im Ausland geboren wurden, noch deutlich unter dem in IGLU und TIMSS ermittelten Durchschnitt von etwa 16 Prozent für Viertklässlerinnen und Viertklässler an Grundschulen in Deutschland (vgl. Kapitel 8 in diesem Band und Schwippert et al., 2012). Dieser Anteil findet sich in etwa für den Leistungstyp 4 mit durchschnittlichen Leistungen (15.1%), während der Anteil von Schülerinnen und Schülern, deren Eltern im Ausland geboren sind, in der Gruppe der Schülerinnen und Schüler mit unterdurchschnittlichen Leistungen (Typ 3) signifikant höher ist. Ebenfalls signifikant höher und deutlich über dem Durchschnitt ist dieser Anteil in den Leistungstypen mit sehr geringen (Typ 1) und geringen Leistungen (Typ 2): Mit jeweils über 36 Prozent sind Schülerinnen und Schüler mit Migrationshintergrund hier deutlich überrepräsentiert.

Eine weitere signifikante Abgrenzung von Leistungstyp 7 zu allen anderen Leistungstypen ergibt sich über die Familiensprache. Beim obersten Leistungstyp (Typ 7) sprechen lediglich weniger als fünf Prozent der Schülerinnen und Schüler zu Hause nur manchmal oder nie Deutsch. In den beiden Typen des untersten Leistungsniveaus finden sich hier jeweils höhere Anteile: Bei 31.5 Prozent (Typ 1) und 39.1 Prozent (Typ 2) wird Deutsch zu Hause nur manchmal oder nie gesprochen.

Insgesamt ist bei den Leistungstypen 1 und 2 somit eine soziale und eine Bildungsbenachteiligung zu konstatieren, die sich besonders deutlich beim Vergleich mit der begünstigten Gruppe von Kindern im Leistungstyp 7, aber auch im Leistungstyp 6 abzeichnet. Diese Bildungsbenachteiligung zeigt sich auch am Anteil der Kinder, die nach Angaben der Lehrkräfte einen besonderen Förderbedarf im Bereich des Lernens oder der körperlichen und motorischen Entwicklung haben. In den Leistungstypen mit überdurchschnittlichen, sehr hohen und herausragenden Leistungen (Typen 5, 6 und 7) sind Schülerinnen und Schüler, die nach Angaben der Lehrkräfte einen besonderen Förderbedarf haben, jeweils mit etwa einem Prozent vertreten. Es ist anzunehmen, dass es sich bei diesen Kindern in der Regel um Schülerinnen und Schüler mit Förderbedarf im Bereich der körperlichen und motorischen Entwicklung, nicht jedoch im Bereich des Lernens handelt. Eine weitere Ausdifferenzierung kann mit den vorliegenden Daten jedoch nicht erfolgen. In der Gruppe mit durchschnittlichen Leistungen (Typ 4) haben nach Angaben der Lehrkräfte 3 Prozent der Schülerinnen und

Schüler einen besonderen Förderbedarf. Dieser Anteil nimmt vom Leistungstyp 3 mit unterdurchschnittlichen (4.8%) zu den Typen mit geringen (Typ 2: 13.5%) und sehr geringen (Typ 1: 39.4%) Leistungen jeweils signifikant zu. Hierbei dürfte es sich vermutlich eher um Schülerinnen und Schüler mit Förderbedarf im Bereich des Lernens handeln.

4.3 Beschreibung der Leistungstypen nach fachbezogenen Einstellungen und Selbstkonzepten

Abschließend werden die Leistungstypen hinsichtlich der Einstellungen und Selbstkonzepte der Schülerinnen und Schüler beschrieben. In Tabelle 9.8 sind die Werte auf den Skalen zu den fachbezogenen Einstellungen und den Selbstkonzepten der Schülerinnen und Schüler in den sieben Leistungstypen dargestellt.

In Bezug auf die Einstellung zum Lesen verdeutlichen die Anteile in Tabelle 9.8 Folgendes: Je höher das Leistungsniveau, desto größer der Anteil von Schülerinnen und Schülern mit einer hohen positiven Einstellung zum Lesen. Beim Leistungstyp 1 mit nur sehr geringen Leistungen ist dieser Anteil signifikant kleiner als in allen anderen Leistungstypen. Bei leistungsstärkeren Typen ergeben sich jeweils signifikant größere Anteile an Schülerinnen und Schülern mit hohen positiven Einstellungen zum Lesen (mit Ausnahme von Typ 2 vs. Typ 3 sowie Typ 6 vs. Typ 7).

Das grundsätzliche Muster einer stetigen Zunahme des Anteils an Schülerinnen und Schülern mit günstigen Werten wird lediglich bei den hohen positiven Einstellungen zur Mathematik beziehungsweise zum Fach Sachunterricht durchbrochen. Die größten Anteile von Schülerinnen und Schülern mit einer hohen Ausprägung auf der Skala ‚Positive Einstellung zur Mathematik‘ finden sich in Leistungstyp 7, der Gruppe der Schülerinnen und Schüler mit herausragenden Leistungen (79.4%), und in Typ 1, der Gruppe mit sehr geringen Leistungen (75.0%), wobei die Differenz zwischen den Anteilen in diesen Leistungstypen nicht signifikant ist. Die Anteile von Schülerinnen und Schülern mit hohen positiven Einstellungen zur Mathematik sind in beiden Extremgruppen signifikant größer als in den übrigen Leistungstypen (mit Ausnahme von Typ 1 vs. Typ 4 sowie Typ 1 vs. Typ 6). Die Befunde deuten darauf hin, dass es in der Grundschule offenbar gelingt, sowohl Schülerinnen und Schüler mit herausragenden Leistungen als auch jene mit sehr geringen Leistungen gleichermaßen für Mathematik zu begeistern.

Auch mit Blick auf das Fach Sachunterricht unterscheiden sich die Anteile von Schülerinnen und Schülern mit besonders positiven Einstellungen zum Sachunterricht zwischen den beiden Leistungstypen mit extrem niedrigen beziehungsweise extrem hohen Leistungswerten nicht signifikant. In der Grundschule berichten sowohl Kinder mit herausragenden Leistungen als auch Schülerinnen und Schüler mit sehr geringen Leistungen eine gleichermaßen positive Einstellung zum Sachunterricht.

Bei Betrachtung der fachbezogenen Fähigkeitsselbstkonzepte ergibt sich wiederum das bereits aufgezeigte Muster für die Leistungstypen: Je höher das Leistungsniveau, desto günstiger die Werte, das heißt desto positiver das Selbstkonzept im jeweiligen Fach. Während sich die Leistungstypen der Schülerinnen und Schüler mit sehr geringen und geringen Leistungen (Typen 1 und 2) sowie die Gruppen der Kinder mit sehr hohen und herausragenden

Tabelle 9.8: Zusammensetzung der sieben Leistungstypen nach fachbezogener Einstellung und Selbstkonzept (berechnet in der Stichprobe, deskriptiv basierend auf der in Abschnitt 3 gewonnenen Profillösung) – Anteile von Schülerinnen und Schülern mit hohen positiven Einstellungen zum Lesen, zur Mathematik und zum Sachunterricht sowie Anteile mit hohem lese-, mathematik- und sachunterrichtsbezogenem Selbstkonzept nach den sieben Leistungstypen

	Typ 1	Typ 2	Typ 3	Typ 4	Typ 5	Typ 6	Typ 7
	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)
hohe positive fachbezogene Einstellungen							
zum Lesen	42.4 (5.4)	54.4 (2.8)	56.4 (2.1)	67.6 (1.9)	73.7 (1.6)	83.7 (1.7)	88.8 (2.6)
zur Mathematik	75.0 (3.6)	64.0 (2.4)	65.5 (2.0)	67.5 (1.5)	65.8 (1.9)	68.7 (2.2)	79.4 (3.3)
zum Fach Sachunterricht	77.7 (3.4)	74.2 (2.7)	74.6 (1.9)	80.2 (1.6)	81.0 (1.4)	84.2 (1.7)	86.3 (2.9)
hohe fachbezogene Selbstkonzepte							
Lesen	49.5 (4.3)	54.6 (2.4)	66.0 (2.0)	79.6 (1.5)	85.2 (1.4)	92.2 (1.3)	95.0 (1.6)
Mathematik	46.7 (4.4)	49.7 (2.7)	59.9 (2.0)	70.0 (1.7)	78.8 (1.6)	84.8 (1.6)	97.5 (1.2)
Sachunterricht	50.3 (3.8)	67.1 (2.9)	67.6 (2.1)	78.8 (1.7)	86.2 (1.4)	92.8 (1.5)	98.7 (0.9)

IEA: Trends in International Mathematics and Science Study

© TIMSS 2011

Leistungen (Typen 6 und 7) nicht signifikant hinsichtlich des Anteils von Kindern mit hohen lesebezogenen Selbstkonzepten unterscheiden, lässt sich für alle anderen Vergleiche festhalten, dass in Leistungstypen mit höherer Leistung auch der Anteil an Schülerinnen und Schülern mit hohen lesebezogenen Selbstkonzepten signifikant größer ist. Der Anteil der Schülerinnen und Schüler mit hohen mathematikbezogenen Selbstkonzepten unterscheidet sich zwischen den Gruppen der Schülerinnen und Schüler mit sehr geringen und geringen Leistungen (Typen 1 und 2) nicht signifikant, während für alle anderen Vergleiche zu konstatieren ist, dass sich für die leistungsstärkeren Typen auch signifikant größere Anteile an Schülerinnen und Schülern mit hohen mathematikbezogenen Selbstkonzepten ergeben. Dieser Zusammenhang gilt grundsätzlich auch für das sachunterrichtsbezogene Selbstkonzept. Nur die Leistungstypen mit geringen und unterdurchschnittlichen Leistungen (Typen 2 und 3) unterscheiden sich hinsichtlich des Anteils von Schülerinnen und Schülern mit hohen sachunterrichtsbezogenen Selbstkonzepten nicht signifikant voneinander.

5 Zusammenfassung und Diskussion

In diesem Kapitel wurden die Leistungen der Schülerinnen und Schüler in den Domänen Lesen, Mathematik und Naturwissenschaften vergleichend untersucht. Für die gemeinsame deutsche IGLU/TIMSS-2011-Schülerstichprobe wurden mitunter auch Leistungsprofile berechnet, wobei die Fragestellung im Vordergrund stand, inwieweit sich eher homogene oder eher heterogene Leistungsprofile zeigen. Im Folgenden werden die einzelnen Analyseschritte und Ergebnisse zusammengefasst und die Befunde nach Möglichkeit mit anderen Studien verglichen.

In einem ersten Schritt wurden die Anteile von Schülerinnen und Schülern auf den verschiedenen Kompetenzstufen für Lesen (IGLU) und Mathematik beziehungsweise Naturwissenschaften (TIMSS) gegenüber gestellt. Es wurde überprüft, wie viele Schülerinnen und Schüler sich in den drei Domänen auf dem Niveau der gleichen oder auf unterschiedlichen Kompetenzstufen befinden. Es konnte gezeigt werden, dass sich für etwa ein Drittel der Viertklässlerinnen und

Viertklässler ein ausgeglichenes Leistungsniveau ergibt, das heißt, dass sie in den drei Domänen jeweils Leistungen erreichen, die der gleichen Kompetenzstufe entsprechen. Fasst man jeweils die aneinandergrenzenden Kompetenzstufen zusammen, erhöht sich dieser Anteil deutlich (90.6%).

Ein heterogenes Leistungsprofil – also Leistungen in Lesen, Mathematik und Naturwissenschaften, die über drei Kompetenzstufen streuen – zeigt sich lediglich für ungefähr 3 Prozent der Kinder im oberen Leistungsbereich (III bis V) und für etwa 2 Prozent im unteren Leistungsbereich (I bis III). Ebenfalls nur sehr wenige Schülerinnen und Schüler (etwa 4%) zeigen extrem heterogene Leistungen, die in den drei Domänen sowohl auf den beiden untersten als auch auf den beiden obersten Kompetenzstufen liegen, so dass eine Inselbegabung beziehungsweise eine Lernschwäche in einer Domäne angenommen werden könnte. Dies ist durchaus erwartungsgemäß; schließlich sind isolierte Lese- oder Rechenschwächen eher in den beiden ersten Schuljahren zu erwarten. Plausibel ist, dass eine isolierte Leseschwäche spätestens dann die Leistungen in den anderen Fächern beeinträchtigt, wenn die Lerninhalte in diesen Fächern in schriftlichen Texten präsentiert werden und die Aufgaben das Lesen und Verarbeiten schriftlicher Informationen erfordern, was vor allem beim selbstständigen Erarbeiten eines Themengebiets der Fall ist.

In einem zweiten Schritt wurden die Testleistungen in den drei Bereichen Lesen, Mathematik und Naturwissenschaften für Deutschland gemeinsam skaliert, um so die Leistungen der Schülerinnen und Schüler direkt miteinander vergleichen zu können. In die Skalierung wurde ein nationales Hintergrundmodell eingeschlossen. Um die Zusammenhänge in den erhobenen Leistungsdaten vertiefend betrachten zu können, wurde ein eindimensionales (Referenzmodell) gegen ein drei- (IGLU- und TIMSS-Gesamtskalen) beziehungsweise achtdimensionales (IGLU- und TIMSS-Subskalen) Skalierungsmodell getestet. Der Vergleich der Modelle unter Nutzung der LRT-Statistik zeigt, dass sowohl das drei- als auch das achtdimensionale Modell gegenüber dem eindimensionalen Modell zu bevorzugen sind. Darüber hinaus verwies der LRT-Ansatz darauf, dass beide Modelle (das drei- und das achtdimensionale Modell) die Datenstruktur vergleichbar gut beschreiben. Es wurde auf das Problem des *Trade-Off* zwischen Fit und Komplexität dieser Modelle hingewiesen und konstatiert, dass weitere Forschung an dieser Stelle nötig ist, um geeignete informationstheoretische Auswahlmaße in dieser Modellfamilie zu entwickeln.

Diese mehrdimensionale IRT-Modellierung ermöglicht auch die Bestimmung latenter Korrelationen zwischen den Kompetenzdomänen und Inhaltsbereichen. Hier zeigt sich, dass zwischen den Testleistungen substantielle Zusammenhänge bestehen. Während die Leistungen in Mathematik und Naturwissenschaften mit .66 korrelieren, hängen Leistungen in den naturwissenschaftlichen Tests (.74) noch stärker als die in den Mathematiktests (.54) mit dem Leseverständnis zusammen. Die Betrachtung der Zusammenhänge weist darauf hin, dass zwar substantielle Korrelationen zwischen den drei domänenspezifischen Leistungstests bestehen, jedoch eine nach Kompetenzdomänen und Inhaltsbereichen differenzierte Betrachtung von Schulleistungen zu bevorzugen ist.

In einem dritten Schritt wurden anhand einer latenten Profilanalyse sieben Typen von Schülerinnen und Schülern ermittelt. Die gefundenen Leistungstypen unterscheiden sich vornehmlich im erzielten Leistungsniveau. Zudem zeigt sich ein relativ paralleles Muster der Leistungsprofile über die verschiedenen Kompetenzdomänen hinweg. Schülerinnen und Schüler in Deutschland weisen also überwiegend ausgeglichene Profile über die Kompetenzdomänen Lesen,

Mathematik und Naturwissenschaften auf, die sich insgesamt durch die Höhe des Niveaus der Kompetenzen von anderen Leistungstypen unterscheiden. Für die Leistungstypen mit unterdurchschnittlichen und durchschnittlichen Leistungen ergeben sich eher ausgeglichene Profile in den drei Leistungsbereichen, während sich für den Leistungstyp mit über dem Durchschnitt liegenden Leistungen ein etwas heterogeneres Profil zeigt. Dieser Typ schneidet besser im Bereich Lesen ab als in den Naturwissenschaften und Mathematik. Vergleichsweise unausgewogene Leistungsprofile zeigen sich für die Typen an den beiden Extrema der Leistungsskala: Die Leseleistung der Leistungstypen im unteren Leistungsbereich liegt unter den Leistungen in Naturwissenschaften und – noch deutlicher – unter den Mathematikleistungen. Für den leistungsstärksten Typ mit herausragenden Leistungen sind ebenfalls die Mittelwerte für Mathematik höher als für die Leseleistung, die gleichauf liegt mit der Leistung in den Naturwissenschaften, während für den Leistungstyp mit sehr hohen Leistungen die mittlere Leistung im Bereich Mathematik geringer ausfällt als im Lesen und in den Naturwissenschaften.

Die hier vorgelegten Analysen unterstreichen die Ergebnisse, die sich schon bei dem Vergleich der Kompetenzstufenzugehörigkeit der Schülerinnen und Schüler in den jeweils erfassten Domänen Lesen, Mathematik und Naturwissenschaften abgezeichnet haben: Die Leistungen in den drei Domänen sind bei der Mehrheit der Kinder, wenn auch auf unterschiedlichen Niveaus, recht ausgeglichen. Große Unterschiede in den Leistungen sind die Ausnahme – am deutlichsten treten sie bei Schülerinnen und Schülern mit geringen und sehr geringen Leistungen auf, bei denen die Leseleistungen am schlechtesten ausfallen.

Das Ergebnis, dass die Leistungsprofile der Schülerinnen und Schüler eher ausgeglichen sind und die Leistungen im Lese-, Mathematik- und naturwissenschaftlichen Test sowie deren Subdomänen innerhalb der Leistungstypen nicht so weit auseinander liegen, sollte nicht überraschen, schließlich erfordert die Lösung der TIMSS-Aufgaben in Mathematik und Naturwissenschaften das sinnverstehende Lesen, wobei bei etwa der Hälfte der Aufgaben auch längere Textpassagen zu bewältigen sind. Für vertiefende Auswertungen wäre deshalb eine Aufgabenanalyse der TIMSS-Tests von Interesse, wobei die Leseschwierigkeiten der einzelnen Aufgaben zu bestimmen wäre. Eine von Lesekompetenzen unabhängige Bestimmung von Mathematik- beziehungsweise naturwissenschaftlichen Kompetenzen wäre nur möglich, wenn in den beiden letztgenannten Bereichen mündliche Aufgaben gestellt werden, die so strukturiert sind, dass sie das Arbeitsgedächtnis nicht überfordern, da beim Zuhören nur jeweils eine bestimmte Menge von Informationen aufgenommen und behalten werden kann. In der Schule werden Mathematik und Naturwissenschaften verstärkt über schriftliche Informationen vermittelt und in schriftlichen Arbeiten geprüft, so dass das Leseverständnis stets beteiligt ist.

Um wichtige Hintergrundmerkmale bei der Interpretation der Leistungstypen nicht unberücksichtigt zu lassen, wurden in einem weiteren Schritt relevante Hintergrundmerkmale betrachtet, die mit schulischen Leistungen korreliert sind. Zunächst wurde das Geschlecht der Kinder, die den verschiedenen Leistungstypen zugeordnet wurden, herangezogen. Dabei zeigte sich, dass der einzige Leistungstyp, in dem sich ein signifikanter Unterschied in den Anteilen der Geschlechter zeigt, die Gruppe der Schülerinnen und Schüler mit herausragenden Leistungen ist. Dieser Leistungstyp repräsentiert mit 4 Prozent allerdings einen nur sehr kleinen Anteil der Schülerinnen und Schüler in Deutschland. Dieser Befund reflektiert ein generelles Phänomen, wonach in der Gruppe der

Schülerinnen und Schüler mit herausragenden Leistungen zumeist mehr Jungen als Mädchen vertreten sind. Jungen sind in diesem Leistungstyp mit einem Anteil von 58.3 Prozent signifikant häufiger vertreten als Mädchen. Es scheint der Grundschule folglich nicht zu gelingen, Mädchen in allen Kompetenzbereichen so zu fördern und zu motivieren, dass sie einen vergleichbaren Anteil in der Gruppe der Schülerinnen und Schüler mit extrem hohen Kompetenzen in allen Bereichen haben. In den übrigen Leistungstypen finden sich keine signifikanten Unterschiede zwischen den Anteilen von Mädchen und Jungen. Diese Ergebnisse liefern hinsichtlich des Aspekts von Kompetenzunterschieden keinen Beleg für die gegenwärtig in der Öffentlichkeit diskutierte These der Benachteiligung von Jungen in der Schule (zur Diskussion vgl. Aktionsrat Bildung, 2009; König & Valtin, 2011; Valtin, Wagner & Schwippert, 2005).

In Bezug auf kulturelle und soziale Hintergrundmerkmale unterscheiden sich viele der sieben Leistungstypen signifikant voneinander. So weist die Zugehörigkeit der Schülerinnen und Schüler zu den Leistungstypen eine sehr enge Kopplung mit der Zuordnung der Schülerfamilien zur Sozialschicht auf. Etwa 12 Prozent der Familien von Kindern der Leistungstypen mit geringen und sehr geringen Leistungen kann der dritten Sozialschicht (Akademiker, Techniker und Führungskräfte) zugeordnet werden. In den Leistungstypen mit sehr hohen und herausragenden Leistungen sind es mehr als die Hälfte der Schülerinnen und Schüler. Dies verdeutlicht, wie ausgeprägt der Zusammenhang von Leistung und sozioökonomischen Determinanten der Schülerfamilien ist. Zusätzlich zeigt sich, dass etwa jedes zweite Kind der Leistungstypen mit geringen und sehr geringen Leistungen von Armut gefährdet ist. In den Leistungstypen mit sehr hohen und herausragenden Leistungen betrifft dies maximal jedes zehnte Kind.

Hinsichtlich der Bildungsnähe der Elternhäuser zeigt sich, dass grundsätzlich für die höheren Leistungstypen auch der Anteil von Kindern aus bildungsnahen Elternhäusern höher ausfällt: Für fast alle Vergleiche gilt, dass bei Typen mit höherer Leistung auch der Anteil an Schülerinnen und Schülern, deren Mutter oder Vater einen Hochschul- oder höheren Abschluss besitzt, bedeutsam größer ist. Der Unterschied in den Anteilen von Elternteilen mit Hochschulabschluss von etwa 61 Prozentpunkten zwischen dem Typ mit sehr geringen Leistungen (3.0%) und dem Leistungstyp mit herausragenden Leistungen (63.9%) verdeutlicht die Bedeutung der Bildungsnähe, die möglicherweise auch mit der Bildungsaspiration verknüpft ist. Auch der Buchbesitz der Schülerfamilien, welcher als Indikator für das kulturelle Kapital der Schülerfamilien genutzt wird, weist auf deutliche Unterschiede zwischen den Leistungstypen hin. Während in den Leistungstypen mit sehr geringen, geringen und unterdurchschnittlichen Leistungen maximal 20 Prozent der Schülerinnen und Schüler aus Elternhäusern stammen, in denen es ihrer Angabe nach mehr als 100 Bücher gibt, nimmt dieser Anteil für die Leistungstypen mit höherer Leistung jeweils signifikant zu.

Auch bei Betrachtung des Migrationshintergrunds zeigen sich zumeist signifikante Unterschiede für die sieben Leistungstypen. Während rund 16 Prozent der Viertklässlerinnen und Viertklässler an Grundschulen in Deutschland aus Familien stammen, in denen die Eltern im Ausland geborenen wurden (vgl. Kapitel 8 in diesem Band und Schwippert et al., 2012), variiert dieser Anteil beträchtlich zwischen den einzelnen Leistungstypen: von 0.7 Prozent bei Schülerinnen und Schülern des Typs mit herausragenden Leistungen zu jeweils über 36 Prozent bei den Leistungstypen mit geringen und sehr geringen Leistungen. Ein ähnliches Bild ergibt sich für die Betrachtung der Schülerinnen und Schüler nach ihrem familiären Sprachgebrauch. Auch hier sind Schülerinnen und Schüler, die nur manchmal

oder nie zu Hause die Testsprache sprechen, in den oberen Leistungstypen unter- und in den unteren Leistungstypen überrepräsentiert.

Die Kopplung von ungünstigen sozioökonomischen sowie bildungsrelevanten Lernbedingungen und sehr schwachen Leistungen zeigte sich auch in einer Analyse im Rahmen von IGLU 2006, in welcher Kinder auf den untersten beiden Lesekompetenzstufen mit den lesestarken Viertklässlerinnen und Viertklässlern auf Kompetenzstufe IV und V verglichen wurden (Valtin, Hornberg, Buddeberg, Kowoll & Potthoff, 2010). Es zeigten sich deutliche Unterschiede beim beruflichen Status der Eltern, der Anzahl der Bücher zu Hause, dem Migrationsstatus (43 Prozent der Leseschwachen gegenüber 8 Prozent der Lesestarken hatten Eltern, die im Ausland geboren waren) und der Familiensprache (Deutsch war Familiensprache bei 80 Prozent der Lesestarken und nur bei 46 Prozent der Leseschwachen). Diese Befunde ergänzen die oben genannten Ergebnisse und machen deutlich, dass hier nach wie vor ein großer Bedarf an Maßnahmen zur Förderung der Unterrichtssprache besteht. Es kann jedoch nicht von einer monokausalen Beziehung zwischen Migrationsstatus und Kompetenzunterschieden ausgegangen werden. Wie in den Beiträgen zu Schülerinnen und Schülern mit Migrationshintergrund (vgl. Kapitel 8 in diesem Band und Schwippert et al., 2012) aufgeführt, zeigt sich ein vielschichtiger Zusammenhang unter anderem zwischen der Migrationsgeschichte und den Sprachkenntnissen in den Familien sowie dem kulturellen und sozioökonomischen Hintergrund der Schülerfamilien.

Insgesamt ist bei den Leistungstypen mit geringen und sehr geringen Leistungen eine Bildungsbenachteiligung zu konstatieren, die sich besonders deutlich im Vergleich zur begünstigten Gruppe von Kindern des Leistungstyps mit herausragenden Leistungen, aber auch zum Leistungstyp mit sehr hohen Leistungen abzeichnet. Diese Bildungsbenachteiligung zeigt sich auch am Anteil der Kinder, die nach Angaben der Lehrkräfte einen besonderen Förderbedarf im Bereich des Lernens oder der körperlichen und motorischen Entwicklung haben. Dieser Anteil beträgt bei den Leistungstypen mit überdurchschnittlichen bis herausragenden Leistungen etwa ein Prozent, beim Typ mit durchschnittlichen Leistungen 3 Prozent, bei Kindern des Typs mit geringen Leistungen 13.5 Prozent und beim Leistungstyp mit sehr geringen Leistungen 39.4 Prozent, wobei dieser Wert mit deutlichem Abstand signifikant größer ist als bei allen anderen Leistungstypen.

Die sieben Leistungstypen wurden nicht nur in Bezug auf kulturelle und soziale Hintergrundmerkmale untersucht, sondern auch hinsichtlich ihrer fachspezifischen Einstellungen und Selbstkonzepte. Für die meisten Vergleiche zwischen den Leistungstypen gilt, dass in den leistungsstärkeren Gruppen auch der Anteil an Schülerinnen und Schülern mit hohen positiven Einstellungen zum Lesen signifikant größer ist, allerdings zeigt sich dieser Zusammenhang nicht für die Einstellungen zur Mathematik und zum Fach Sachunterricht. Die Anteile von Schülerinnen und Schülern mit besonders positiven Einstellungen zur Mathematik beziehungsweise zum Sachunterricht unterscheiden sich nicht signifikant zwischen den beiden Leistungstypen mit extrem niedrigen beziehungsweise extrem hohen Leistungswerten. In der Grundschule scheinen sowohl Kinder mit herausragenden Leistungen als auch Schülerinnen und Schüler mit sehr geringen Leistungen gleichermaßen für Mathematik beziehungsweise den Sachunterricht begeistert zu sein.

Im Gegensatz zu den fachbezogenen Einstellungen hängen die Ausprägungen auf den Skalen zu den fachbezogenen Selbstkonzepten der Schülerinnen und Schüler wesentlich stärker mit der Zugehörigkeit zu einem bestimmten Leis-

tungstyp zusammen. In den leistungsstärkeren Gruppen ist auch der Anteil an Schülerinnen und Schülern mit hohen fachbezogenen Selbstkonzepten signifikant größer. Lediglich für wenige Vergleiche von aneinandergrenzenden Leistungstypen gilt dies nicht. Im Unterricht sollten die ungünstigen Lerneinstellungen und Selbstkonzepte der Schülerinnen und Schüler mit geringen beziehungsweise sehr geringen Leistungen berücksichtigt werden. Die ungünstigen motivationalen Überzeugungen lassen sich unter anderem als Reaktionen auf schulische Misserfolge und schlechte Zensuren deuten. Dass Zensurengebung bei leistungsschwachen Schülerinnen und Schülern zu größerer Leistungsangst und Misserfolgsorientierung führt, ist längsschnittlich belegt (Wagner & Valtin, 2003).

Obwohl der in diesem Kapitel gewählte methodische Ansatz, Typen von Schülerinnen und Schülern mit unterschiedlichen Leistungsprofilen zu bestimmen und näher zu untersuchen, deutlich differenziertere Erkenntnisse hervorbringt als auf Mittelwerte bezogene Analysen, ist die Vorgehensweise jedoch in deutschen Studien zu Schulleistungsvergleichen bisher nicht realisiert worden.

Es gibt eine Bandbreite an *Cluster- und Klassifikationsverfahren* in der Statistik, die hier zum Einsatz kommen könnten (ob auf Rohdatenebene oder basierend auf den skalierten Testwerten). Vor allem aber die „Profilierung“ der erhaltenen Typen (ob während oder nach der eigentlichen Analyse) stellt eine methodisch interessante Fragestellung dar, die untersucht und mit den hier erzielten Ergebnissen verglichen werden kann. Die statistisch fundierte Auswahl von Hintergrundvariablen und Ableitung eines Hintergrundmodells sowie die Bereitstellung von Gütekriterien für diese Art von Kovariatinformationen sind auch weiterer Forschung vorbehalten. Das Zusammenspiel zwischen Kompetenzstufenzuteilung und Leistungsprofilbildung sollte methodisch systematischer verfolgt werden. So ließen sich auch die Zuteilungen zu den Kompetenzstufen anhand diskreter latenter Variablen zum Beispiel mittels der *Latent Class Analyse* realisieren und mit der Typologie basierend auf den Leistungsprofilen vergleichen.

Ein Beispiel für eine Untersuchung, in der zahlreiche typologische Analysen durchgeführt wurden, ist die Studie AIDA (Adaptation in der Adoleszenz; König, Wagner & Valtin, 2011) für die Klassenstufen 7 bis 9. Unter anderem wurden Schulleistungen berücksichtigt, erfasst durch die Schulnoten in den sprachlichen Fächern (Deutsch mündlich bzw. schriftlich sowie Englisch mündlich bzw. schriftlich) und den mathematisch-naturwissenschaftlichen Fächern (Mathematik, Physik, Chemie). Mit dem Verfahren der *Latent Class Analyse* ließen sich in der AIDA-Studie sechs verschiedene Gruppen identifizieren, die sich – vergleichbar mit den hier geschilderten IGLU- und TIMSS-Ergebnissen – vor allem im Niveau der Leistungen beziehungsweise in der Höhe der Zensuren unterscheiden. Bei drei Typen war das Leistungsprofil eher ausgeglichen, die anderen drei Typen wiesen nach Domänen differenzierte Leistungen auf, wobei sich vor allem Unterschiede im Fach Englisch zeigten. Die mathematisch-naturwissenschaftlichen Leistungen lagen bei den verschiedenen Typen eher dicht beisammen. Auch die Zusammenhänge mit den Hintergrundmerkmalen sind in der AIDA-Studie ähnlich denen von IGLU und TIMSS: Je höher der mittlere Bildungs- und Berufsstatus der Eltern ist, desto höher ist auch die Schulleistung eines Typs. Außerdem zeigte sich ein Zusammenhang mit den erhobenen leistungsbezogenen Persönlichkeitsmerkmalen: Ich-Stärke, Leistungsvertrauen und auch Schulfreude waren höher in den Gruppen, die bessere Schulleistungen beziehungsweise Noten aufwiesen. Eine weitere Studie, in der mit *Clusteranalysen* der Zusammenhang zweier Kompetenzbereiche – der Rechen- und Schriftsprachleistung – unter-

sucht wurde, legt den Fokus auf die Betrachtung von Lernstörungen. Schwenck und Schneider (2003) identifizieren in ihrer empirischen Untersuchung mit einer Stichprobe von 115 Kindern in der ersten Jahrgangsstufe vier Subgruppen (1. gute Leistungen in beiden Bereichen: Rechen- und Schriftsprachleistung; 2. geringe Leistungen in beiden Bereichen; 3. geringe Leistungen nur in der Rechenleistung; 4. geringe Leistungen nur im Schriftsprachbereich) und vertreten die These, dass es eine eigenständige Gruppe mit kombinierten Defiziten im Rechen- und Schriftsprachbereich gibt, die spezifisch betrachtet werden sollte. Diese Ergebnisse weisen auf weiterführende Forschungsmöglichkeiten hin, indem zum Beispiel der Fragestellung nachgegangen werden könnte, ob sich bei ausschließlicher Betrachtung der in IGLU und TIMSS ermittelten Leistungstypen mit sehr geringen und geringen Leistungen sowie mit herausragenden und sehr hohen Leistungen noch weitere bedeutsame Subgruppen identifizieren lassen.

Die in diesem Kapitel berichteten Ergebnisse bezogen sich auf die Kompetenzdomänen Lesen, Mathematik und Naturwissenschaften sowie ihre inhaltsbezogenen Subdomänen, welche das Abschneiden bei Aufgaben aus unterschiedlichen inhaltlichen Bereichen erfassen, beispielsweise das Lesen unterschiedlicher Textsorten. Das Design von IGLU und TIMSS erlaubt jedoch auch weitere Auswertungsmöglichkeiten, zum Beispiel die Berücksichtigung der Verstehensprozesse (IGLU) und der kognitiven Anforderungen (TIMSS) sowie die Berechnung von Regressionsanalysen, in welche auch Resultate aus einem kognitiven Fähigkeitstest einbezogen werden können. Es sind entsprechend weiterführende Analysen zum Zusammenspiel der drei Kompetenzdomänen wünschenswert. Beispielsweise könnten parametrische und vor allem auch nicht-parametrische Dimensionalitätstests für die hier angerissene Diskussion der Kompetenzdimensionen und ihrer Zusammenhänge in IGLU und TIMSS weiterführend eingesetzt und in Hinblick auf ihre Befunde interpretiert und verglichen werden.

Literatur

- Adams, R. J. & Carstensen, C. H. (2002). Scaling outcomes. In R. Adams & M. Wu (Hrsg.), *PISA 2000 technical report*. Paris: OECD.
- Adams, R. J., Wilson, M. R. & Wang, W. L. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21 (1), 1-24.
- Aktionsrat Bildung. (2009). *Geschlechterdifferenzen im Bildungssystem. Jahresgutachten 2009* (1. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bacher, J., Pöge, A. & Wenzig, K. (2010). *Clusteranalyse. Anwendungsorientierte Einführung in Klassifikationsverfahren* (3. Aufl.). München: Oldenbourg.
- Boomsma, A., van Duijn, M. A. J. & Snijders, T. A. B. (Hrsg.). (2001). *Essays on item response theory*. New York: Springer.
- Bos, W., Bremerich-Vos, A., Tarelli, I. & Valtin, R. (2012). Lesekompetenzen im internationalen Vergleich. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Bremerich-Vos, A., Valtin, R. & Tarelli, I. (2012). Das Konzept von Lesekompetenz in IGLU 2011. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.

- Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach*. New York: Springer.
- Croon, M. A. (1990). Latent class analysis with ordered latent classes. *British Journal of Mathematical and Statistical Psychology*, 43 (2), 171-192.
- Croon, M. A. (1991). Investigating Mokken scalability of dichotomous items by means of ordinal latent class analysis. *British Journal of Mathematical and Statistical Psychology*, 44 (2), 315-331.
- Fischer, G. H. & Molenaar, I. W. (Hrsg.). (1995). *Rasch models. Foundations, recent developments, and applications*. New York: Springer.
- Foy, P., Brossman, B. & Galia, J. (2012). Scaling TIMSS and PIRLS 2011 achievement data. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 16.11.2012 unter <http://timssandpirls.bc.edu/methods/index.html>.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision – KFT 4-12+R : Manual*. Göttingen: Beltz Test.
- IBM Corp. (2012). *IBM SPSS Statistics for Windows* [Computer software]. Armonk, NY: IBM Corp.
- König, J. & Valtin, R. (2011). Gleichermaßen auf das Leben vorbereitet?: Ergebnisse der Jugendstudie AIDA. In A. Sasse & R. Valtin (Hrsg.), *Mädchen und Jungen in der Schule. Förderung von Lesekompetenz und Leseinteresse* (S. 54–72). Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- König, J., Wagner, C. & Valtin, R. (2011). *Jugend – Schule – Zukunft. Psychosoziale Bedingungen der Persönlichkeitsentwicklung. Ergebnisse der Längsschnittstudie AIDA*. Münster: Waxmann.
- Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56 (2), 177–196.
- Mislevy, R. J., Beaton, A. E., Kaplan, B. A. & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29 (2), 133–161.
- Muthén, L. K. & Muthén, B. O. (2012). *Mplus user's guide* (7. Aufl.). Los Angeles: Muthén & Muthén.
- Myung, I. J., Forster, M. R. & Browne, M. W. (2000). Guest editors' introduction: Special issue on model selection. *Journal of Mathematical Psychology*, 44 (1), 1–2.
- OECD. (2012). *PISA 2009 Technical Report*. Paris: OECD Publishing.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (2., erw. Aufl.). Chicago, IL: University of Chicago Press.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2., vollst. überarb. u. erw. Aufl.). Bern: Huber.
- Schwenck, C. & Schneider, W. (2003). Der Zusammenhang von Rechen- und Schriftsprachkompetenz im frühen Grundschulalter. *Zeitschrift für Pädagogische Psychologie*, 17 (3-4), 261-267.
- Schwippert, K., Bos, W. & Lankes, E.-M. (2003). Heterogenität und Chancengleichheit am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 265–302). Münster: Waxmann.
- Schwippert, K., Wendt, H. & Tarelli, I. (2012). Lesekompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Sijtsma, K. & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.
- Tarelli, I., Wendt, H., Bos, W. & Zylowski, A. (2012). Ziele, Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU 2011). In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Ünlü, A., Kasper, D. & Trendtel, M. (eingereichtes Manuskript). *The OECD's Programme for International Student Assessment (PISA) study: A review of its basic psychometric concepts*.
- Valtin, R., Hornberg, S., Buddeberg, M., Kowoll, M. E. & Potthoff, B. (2010). Schülerinnen und Schüler mit Leseproblemen – eine ökosystemische Betrachtungsweise. In W. Bos,

- S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, I. Tarelli & R. Valtin (Hrsg.), *IGLU 2006 – die Grundschule auf dem Prüfstand. Vertiefende Analysen zu Rahmenbedingungen schulischen Lernens* (S. 43–90). Münster: Waxmann.
- Valtin, R., Wagner, C. & Schwippert, K. (2005). Schülerinnen und Schüler am Ende der vierten Klasse – schulische Leistungen, lernbezogene Einstellungen und außerschulische Lernbedingungen. In W. Bos, E.-M. Lankes, M. Prenzel, R. Valtin & G. Walther (Hrsg.), *IGLU. Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien* (S. 187–238). Münster: Waxmann.
- van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.
- Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Applied Psychological Measurement*, 25 (3), 283-294.
- Vermunt, J. K. & Magidson, J. (2005). *Latent GOLD 4.0 User's Guide*. Belmont, MA: Statistical Innovations Inc.
- Wagner, C. & Valtin, R. (2003). Noten oder Verbalbeurteilungen? Die Wirkung unterschiedlicher Bewertungsformen auf die schulische Entwicklung von Grundschulkindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 45 (1), 27-36.
- Wendt, H., Stubbe, T. C. & Schwippert, K. (2012). Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Wu, M., Adams, R., Wilson, M. R. & Haldane, S. (2007). *ACER ConQuest 2.0. Generalised item response modelling software* [Computer software]. Camberwell: Acer Press.

Anhang A

Tabelle A.1: Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2007

TIMSS 2007		Teilnahme an TIMSS 2011	Besonderheiten bzgl. der nationalen Zielpopulation				Besonderheiten bzgl. der nationalen Schul- und Schülerteilnahmequoten				
			Getestete Jahrgangsstufe	Durchschnittsalter*	Ausschöpfungsgrad**	Ausschlüsse***	Schulteilnahmequote in %		Schülerteilnahmequote in %	Gesamteilnahmequote in %	
							ohne Ersatzschulen	mit Ersatzschulen		ohne Ersatzschulen	mit Ersatzschulen
Teilnehmer											
	Algerien		4	10.2	100	2.1	99	99	97	97	97
	Armenien	✓	4	10.6	100	3.4	93	100	96	90	96
	Australien	✓	4	9.9	100	4.0	99	100	95	94	95
³	Dänemark	✓	4	11.0	100	4.1	71	91	94	66	85
	Deutschland	✓	4	10.4	100	1.3	96	100	97	93	96
	El Salvador		4	11.0	100	2.3	99	100	98	97	98
^{1 3}	England ^A	✓	5	10.2	100	2.1	83	90	93	77	84
²	Georgien	✓	4	10.1	85	4.8	92	100	98	90	98
^{2 3}	Hongkong	✓	4	10.2	100	5.4	81	84	96	78	81
	Iran	✓	4	10.2	100	3.0	100	100	99	99	99
²	Italien	✓	4	9.8	100	5.3	91	100	97	88	97
	Japan	✓	4	10.5	100	1.1	97	99	97	94	95
	Jemen	✓	4	11.2	100	2.0	99	100	98	97	98
²	Kasachstan	✓	4	10.6	94	5.3	99	100	100	99	100
	Katar	✓	4	9.7	100	1.8	100	100	97	97	97
	Kolumbien		4	10.4	100	2.1	93	99	98	91	97
⁶	Kuwait	✓	4	10.2	100	0.0	100	100	85	85	85
²	Lettland		4	11.0	72	4.6	93	97	95	89	92
²	Litauen	✓	4	10.8	93	5.4	99	100	94	93	94
³	Marokko	✓	4	10.6	100	1.4	81	81	96	77	77
^{1 2}	Neuseeland ^A	✓	4.5-5.5	10.0	100	5.4	97	100	96	93	96
³	Niederlande	✓	4	10.2	100	4.8	48	95	97	46	91
²	Norwegen	✓	4	9.8	100	5.1	88	97	95	83	92
	Österreich	✓	4	10.3	100	5.0	98	99	98	96	97
	Russische Föderation	✓	4	10.8	100	3.6	100	100	98	98	98
^{1 3}	Schottland ^A		5	9.8	100	4.5	77	94	94	72	88
	Schweden	✓	4	10.8	100	3.1	98	100	97	94	97
	Singapur	✓	4	10.4	100	1.5	100	100	96	96	96
	Slowakei	✓	4	10.4	100	3.3	98	100	97	95	97
	Slowenien	✓	4	9.8	100	2.1	92	99	95	87	93
	Taiwan	✓	4	10.2	100	2.8	100	100	100	100	100
	Tschechische Republik	✓	4	10.3	100	4.9	89	98	94	83	92
	Tunesien	✓	4	10.2	100	2.9	100	100	99	99	99
	Ukraine		4	10.3	100	0.6	96	96	97	93	93
	Ungarn	✓	4	10.7	100	4.4	93	99	97	90	96
^{2 3}	USA	✓	4	10.3	100	9.2	70	89	95	66	84
Benchmark-Teilnehmer											
²	Alberta, Kanada	✓	4	9.8	100	7.6	99	99	96	94	94
²	Britisch-Kolumbien, Kanada		4	9.8	100	9.2	98	100	96	94	96
^{2 3 6}	Dubai, VAE	✓	4	10.0	100	5.4	75	75	91	67	67
²	Massachusetts, USA		4	10.3	100	10.4	92	96	96	88	92
^{2 3}	Minnesota, USA		4	10.3	100	8.3	53	100	97	52	97
²	Ontario, Kanada	✓	4	9.8	100	6.3	95	96	95	91	92
²	Québec, Kanada	✓	4	10.1	100	6.4	97	98	86	83	84

Tabelle A.2: Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2011

TIMSS 2011		Teilnahme an TIMSS 2007	Besonderheiten bzgl. der nationalen Zielpopulation				Besonderheiten bzgl. der nationalen Schul- und Schülerteilnahmequoten					Besonderheiten bzgl. der Leistungsmessung	
			Getestete Jahrgangsstufe	Durchschnittsalter*	Ausschöpfungsgrad**	Ausschlussse***	Schulteilnahmequote in %		Gesamteilnahmequote in %	Anteil der Schülerinnen und Schüler ohne gültige Leistungswerte in %	Naturwissenschaften	Mathematik	
							ohne Ersatzschulen	mit Ersatzschulen					
	Teilnehmer												
	Armenien	✓	4	10.0	100	2.0	100	100	98	98	98	8	8
^{2 3}	Aserbaidshon		4	10.2	100	7.2	84	100	100	84	100	5	7
	Australien	✓	4	10.0	100	4.4	96	98	95	91	93	2	3
	Bahrain		4	10.4	100	1.1	92	92	98	90	90	5	9
³	Belgien (Fläm. Gem.)		4	10.0	100	5.0	76	95	98	75	92	1	0
	Chile		4	10.1	100	3.7	86	99	96	82	95	3	6
^{2 3}	Dänemark	✓	4	11.0	100	6.3	79	92	95	75	87	1	1
	Deutschland	✓	4	10.4	100	1.9	96	99	96	92	95	1	1
^{1 3}	England ^B	✓	5	10.2	100	2.0	81	83	94	76	78	2	2
	Finnland		4	10.8	100	3.1	97	99	96	93	96	0	1
²	Georgien	✓	4	10.0	92	4.9	97	98	99	95	96	4	9
²	Hongkong	✓	4	10.1	100	8.6	87	88	93	81	82	1	0
	Iran	✓	4	10.2	100	4.5	100	100	99	99	99	5	11
	Irland		4	10.3	100	2.5	97	99	95	93	95	2	2
³	Italien	✓	4	9.7	100	3.7	81	98	97	78	95	1	2
	Japan	✓	4	10.5	100	3.2	96	99	97	93	97	0	0
^{4 5}	Jemen	✓	4	11.2	100	3.7	99	99	97	95	95	39	48
²	Kasachstan	✓	4	10.4	100	6.3	99	100	99	98	99	2	3
²	Katar	✓	4	10.0	100	6.2	100	100	99	99	99	11	15
²	Kroatien		4	10.7	100	7.9	99	100	95	94	95	1	3
^{2 4 5}	Kuwait	✓	4	9.7	78	0.3	99	99	94	91	91	18	28
²	Litauen	✓	4	10.7	93	5.6	94	100	94	89	94	1	1
¹	Malta ^B		5	9.8	100	3.6	100	100	95	95	95	6	4
^{4 5}	Marokko	✓	4	10.5	100	2.0	100	100	97	96	96	28	27
^{1 3}	Neuseeland ^B	✓	4.5-5.5	9.9	100	4.9	83	96	94	77	90	2	5
³	Niederlande	✓	4	10.2	100	4.0	49	82	97	47	79	0	0
³	Nordirland		4	10.4	100	3.5	62	85	93	58	79	2	2
³	Norwegen	✓	4	9.7	100	4.3	57	82	85	48	70	1	3
⁴	Oman		4	9.9	100	1.5	98	98	98	96	96	13	19
²	Österreich	✓	4	10.3	100	5.1	100	100	98	98	98	0	1
	Polen		4	9.9	100	3.8	100	100	96	96	96	2	4
	Portugal		4	10.0	100	2.5	87	98	94	81	92	1	1
	Republik Korea (Südkorea)		4	10.4	100	2.5	100	100	98	98	98	0	0
	Rumänien		4	10.9	100	4.0	99	100	98	97	97	5	8
²	Russische Föderation	✓	4	10.8	100	5.3	100	100	98	98	98	0	1
	Saudi-Arabien		4	10.0	100	1.6	95	100	99	94	99	6	14
	Schweden	✓	4	10.7	100	4.1	97	99	92	89	91	1	2
²	Serbien		4	10.8	100	9.4	97	100	97	94	97	2	4
²	Singapur	✓	4	10.4	100	6.3	100	100	96	96	96	1	1
	Slowakei	✓	4	10.4	100	4.6	95	99	96	91	96	1	3
	Slowenien	✓	4	9.9	100	2.6	96	97	97	93	94	1	2
²	Spanien		4	9.8	100	5.3	96	99	97	94	97	1	3
	Taiwan	✓	4	10.2	100	1.4	100	100	99	99	99	0	0
	Thailand		4	10.5	100	1.5	85	100	99	84	99	4	6
²	Tschechische Republik	✓	4	10.4	100	5.1	90	99	95	85	94	0	2
^{4 5}	Tunesien	✓	4	10.0	100	2.5	100	100	99	99	99	21	25
	Türkei		4	10.1	100	2.5	97	100	98	95	98	4	6
	Ungarn	✓	4	10.7	100	4.2	98	99	97	94	96	2	4
^{2 3}	USA	✓	4	10.2	100	7.0	79	84	95	76	80	1	1
	Vereinigte Arabische Emirate (VAE)		4	9.8	100	3.3	100	100	97	97	97	7	11
Teilnahme mit Jgst. 6													
¹	Botsuana		6	12.8	100	0.3	100	100	99	99	99	14	11
^{1 4}	Honduras		6	12.7	100	4.5	91	100	97	88	97	6	17
^{1 4}	Jemen		6	13.2	100	4.0	99	99	96	96	96	15	26
Benchmark-Teilnehmer													
	Abu Dhabi, VAE		4	9.7	100	2.7	99	99	98	97	97	8	14
²	Alberta, Kanada	✓	4	9.9	100	7.5	98	99	96	93	95	1	2
²	Dubai, VAE	✓	4	9.8	100	5.1	100	100	96	96	96	5	8
²	Florida, USA		4	10.4	89	12.1	96	96	95	91	91	1	1
²	North Carolina, USA		4	10.2	93	10.1	94	94	95	89	89	1	1
²	Ontario, Kanada	✓	4	9.8	100	5.3	97	98	96	93	94	1	2
	Québec, Kanada	✓	4	10.1	100	3.7	95	96	95	90	91	0	0

- = Abweichung von der internationalen Vorgabe
- = starke Abweichung von der internationalen Vorgabe
- 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
- 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
- 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
- 4 = Sehr hoher Anteil an Schülerinnen und Schülern mit nicht skalierbaren Leistungswerten in Mathematik.
- 5 = Sehr hoher Anteil an Schülerinnen und Schülern mit nicht skalierbaren Leistungswerten in Naturwissenschaften.
- 6 = Abweichender Testzeitpunkt (In Kuwait und Dubai, VAE, erfolgte die Testung zeitlich verzögert.)
 - * Durchschnittsalter berechnet als: Anzahl der Jahre + Anzahl der Monate/12 (z.B. 9 Jahre, 3 Monate = $9 + 3/12 = 9.25$ Jahre).
 - ** Ausschöpfungsgrad der nationalen Zielpopulation in Prozent bezogen auf die internationale Vorgabe (100 %)
 - *** Ausschlüsse von der nationalen Zielpopulation (Gesamtquote) in Prozent
- A = Da das Durchschnittsalter der Viertklässlerinnen und Viertklässler in England, Neuseeland und Schottland bedeutsam unter der internationalen Vorgabe (9.5 Jahre) lag, wurde gemäß der Kombination der Kriterien ‚formale Schulungszeit‘ und ‚Durchschnittsalter des Schülerjahrgangs‘ eine höhere Jahrgangsstufe als äquivalente nationale Zielpopulation interpretiert (vgl. Mullis, Martin & Foy, 2008, S. 378-381).
- B = Da das Durchschnittsalter der Viertklässlerinnen und Viertklässler in England, Malta und Neuseeland bedeutsam unter der internationalen Vorgabe (9.5 Jahre) lag, wurde gemäß der Kombination der Kriterien ‚formale Schulungszeit‘ und ‚Durchschnittsalter des Schülerjahrgangs‘ eine höhere Jahrgangsstufe als äquivalente nationale Zielpopulation interpretiert (vgl. Mullis, Martin, Foy & Arora, 2012; Martin, Mullis, Foy & Stanco, 2012).

Anhang B

Beschreibung der internationalen Benchmarks für das Mathematik- Kompetenzmodell in TIMSS 2011

Benchmark 1 für die Gesamtskala Mathematik (Kompetenzstufe II – ab 400 Punkten) – Niedrige Benchmark –

Zusammenfassung

Die Schülerinnen und Schüler verfügen über elementares mathematisches Wissen sowie über elementare mathematische Fertigkeiten und Fähigkeiten. Sie können einfache Additions- und Subtraktionsaufgaben mit natürlichen Zahlen lösen. Sie verfügen über anfängliches Wissen zu zueinander parallelen und rechtwinkligen Geraden, geometrischen Grundformen und Koordinatensystemen. Sie können einfachen Balkendiagrammen und Tabellen Informationen entnehmen und diese vervollständigen.

Schülerinnen und Schüler auf dieser Kompetenzstufe können Additions- und Subtraktionsaufgaben mit natürlichen Zahlen lösen. Beispielsweise können sie eine vierstellige und eine dreistellige Zahl addieren. Sie sind mit dem Zahlenraum bis 1000 vertraut.

Die Schülerinnen und Schüler haben Grundkenntnisse hinsichtlich zu einander parallelen und rechtwinkligen Geraden und zu einfachen geometrischen Grundformen. Sie können Positionen auf Karten oder Spielplänen angeben (z.B. A3) und Informationen aus einfachen Säulendiagrammen sowie Tabellen entnehmen und diese vervollständigen.

Benchmark 2 für die Gesamtskala Mathematik (Kompetenzstufe III – ab 475 Punkten) – Durchschnittliche Benchmark –

Zusammenfassung

Die Schülerinnen und Schüler können elementares mathematisches Wissen sowie elementare mathematische Fertigkeiten und Fähigkeiten in einfachen Situationen anwenden. Die Schülerinnen und Schüler auf dieser Kompetenzstufe zeigen Grundwissen von natürlichen Zahlen und anfängliches Wissen von Brüchen. Sie können sich dreidimensionale Figuren ausgehend von zweidimensionalen Darstellungen vorstellen. Sie entnehmen Säulendiagrammen, Piktogrammen und Tabellen Informationen, um einfache Probleme zu lösen.

Schülerinnen und Schüler auf dieser Kompetenzstufe zeigen grundlegende Zahl- und Operationsvorstellungen im Bereich der natürlichen Zahlen. Beispielsweise können sie den Wert einer Ziffer in einer vierstelligen Zahl bestimmen oder Aufgaben lösen, bei denen zwei einstellige Zahlen multipliziert werden. Die Schülerinnen und Schüler können Dezimalzahlen mit einer Nachkommastelle addieren und einen Term identifizieren, der eine Additions- oder eine Subtraktionsaufgabe repräsentiert. Sie können Darstellungen von Stammbrüchen und von Nicht-Stammbrüchen identifizieren und einfache Aufgaben zur Proportionalität lösen, bei denen halbiert werden muss. Sie können einfache geometrische Muster fortsetzen und deren nächste Glieder finden.

Die Schülerinnen und Schüler können sich dreidimensionale Figuren ausgehend von zweidimensionalen Darstellungen vorstellen und dabei einige Eigenschaften von bekannten Körpern erkennen. Sie können Winkel der Größe nach sortieren. Sie können eine Symmetrieachse erkennen und eine einfache Figur spiegeln. Sie können die Bewegungen auf einem Plan angeben, die erforderlich sind, um von einem Punkt zu einem anderen zu gelangen.

Die Schülerinnen und Schüler können Informationen in Säulendiagrammen, Piktogrammen und Tabellen interpretieren und damit einfache Aufgaben lösen. Sie können verschiedene Darstellungen derselben Daten lesen und interpretieren. Beispielsweise können sie Informationen aus Kreisdiagrammen und Tabellen miteinander abgleichen.

Benchmark 3 für die Gesamtskala Mathematik (Kompetenzstufe IV – ab 550 Punkten) – Hohe Benchmark –

Zusammenfassung

Die Schülerinnen und Schüler können ihre mathematischen Fertigkeiten und Fähigkeiten für das Lösen von Problemen anwenden. Sie können mit Hilfe der Grundrechenarten Textaufgaben lösen und dabei auch die Division sowie ihr Verständnis des Stellenwertprinzips in verschiedenen Situationen nutzen. Sie können Zahlenmuster fortsetzen und dabei auch weiter entfernte Werte bestimmen. Die Schülerinnen und Schüler zeigen ein Verständnis der Achsensymmetrie und geometrischer Eigenschaften. Sie können Daten aus Diagrammen und Tabellen entnehmen und interpretieren, um Probleme zu lösen. Sie können Informationen aus Piktogrammen und Strichlisten nutzen, um Säulendiagramme zu vervollständigen.

Schülerinnen und Schüler auf dieser Kompetenzstufe können Textaufgaben im Bereich der natürlichen Zahlen mit Hilfe der Grundrechenarten lösen. Sie können zwei zweistellige Zahlen multiplizieren und die Division in verschiedenen Situationen nutzen. Die Schülerinnen und Schüler können ihr Stellenwertverständnis nutzen, um Aufgaben zu lösen. Beispielsweise können sie fehlende Ziffern in Zahlen bestimmen, wenn deren Stellenwert angegeben ist. Sie können die beste Approximation für eine Summe auswählen oder Überschlagsergebnisse angeben. Sie zeigen anfängliches Verständnis von Vielfachen und Teilern.

Die Schülerinnen und Schüler können unbeschriftete Skalierungen auf einer Waage lesen und Textaufgaben lösen, bei denen gemessen und proportional gedacht werden muss. Sie können Aufgaben lösen, bei denen Uhrzeiten und Zeitspannen addiert werden müssen. Sie können Dezimalzahlen mit zwei Nachkommastellen addieren und Stammbrüche der Größe nach ordnen. Die Schülerinnen und Schüler können eine Zahl angeben, die zwischen zwei natürlichen Zahlen liegt. Sie können Muster fortsetzen, nicht benachbarte Folgenglieder ermitteln und eine zweiteilige Regel nutzen, um weitere Glieder zu finden.

Die Schülerinnen und Schüler demonstrieren ein Verständnis von Achsensymmetrie. Beispielsweise können sie eine Symmetrieachse zeichnen, Objekte an einer Symmetrieachse spiegeln und symmetrische Figuren identifizieren. Sie können Formen anhand vorgegebener Kriterien klassifizieren. Sie können rechte Winkel sowie zueinander parallele und rechtwinklige Geraden in unterschiedlichen Lagen erkennen. Die Schülerinnen und Schüler können den Umfang einfacher Figuren bestimmen. Sie können das Netz eines Würfels oder das Würfelgebäude mit dem größten Volumen identifizieren.

Die Schülerinnen und Schüler können Daten in Tabellen und Säulendiagrammen lesen und interpretieren, um Aufgaben zu lösen. Beispielsweise können sie Informationen vergleichen, die auf unterschiedliche Weise dargestellt sind, und daraus Schlussfolgerungen ziehen. Sie können Informationen aus Piktogrammen und Strichlisten nutzen, um Säulendiagramme zu vervollständigen.

Benchmark 4 für die Gesamtskala Mathematik (Kompetenzstufe V – ab 625 Punkten) – Fortgeschrittene Benchmark –

Zusammenfassung

Die Schülerinnen und Schüler können ihre mathematischen Fertigkeiten und Fähigkeiten verständlich beim Lösen verhältnismäßig komplexer Aufgabenstellungen anwenden und ihr Vorgehen erläutern. Sie können mehrschrittige Sachaufgaben in vielfältigen Kontexten mit natürlichen Zahlen lösen, auch durch Nutzen proportionaler Beziehungen. Sie lassen ein fortschreitendes Verständnis für Brüche und Dezimalbrüche erkennen. Die Schülerinnen und Schüler können ihr geometrisches Wissen und Können im Bereich zwei- und dreidimensionaler Figuren in vielfältigen Situationen anwenden. Sie können Schlussfolgerungen aus Daten in Tabellen ziehen und diese begründen.

Die Schülerinnen und Schüler können mehrschrittige Textaufgaben in vielfältigen Kontexten mit natürlichen Zahlen lösen. Sie können Aufgaben zur Proportionalität lösen, die natürliche Zahlen enthalten, und einfache Zahlensätze mit Unbekannten lösen. Sie lassen ein wachsendes Verständnis für Brüche und Dezimalzahlen erkennen. Sie können äquivalente Brüche bestimmen, die auf unterschiedliche Weise dargestellt sind. Sie können eine größere Bruchzahl mit einem anderen Nenner zu einer vorgegebenen Bruchzahl bestimmen. Sie können in einer Liste die kleinste Dezimalzahl mit einer beziehungsweise zwei Nachkommastellen bestimmen und ihre Kenntnisse und Fertigkeiten zum Lösen von zweiseitigen Aufgaben nutzen. Die Schülerinnen und Schüler können eine zweiseitige Regel für lineare Beziehungen zwischen zwei Zahlen in einer Tabelle identifizieren.

Die Schülerinnen und Schüler können ihr geometrisches Wissen und Können im Bereich zwei- und dreidimensionaler Figuren in vielfältigen Situationen einsetzen. Sie können die Länge einer gekrümmten Linie abschätzen. Sie nutzen ihr Wissen zum Umfang von Figuren, um mehrschrittige Aufgaben zu lösen. Die Schülerinnen und Schüler können den Flächeninhalt einfacher Figuren bestimmen, beispielsweise einer aus Rechtecken zusammengesetzten Figur oder eines gleichseitigen Dreiecks auf Kästchenpapier. Sie können den Flächeninhalt eines Rechtecks berechnen. Sie können die Anzahl der Würfel ermitteln, die benötigt wird, um einen vorgegebenen Quader auszufüllen.

Die Schülerinnen und Schüler können Daten strukturieren, interpretieren und darstellen, um zweiseitige Aufgaben zu lösen. Sie können Schlussfolgerungen aus Daten in Tabellen ziehen und diese begründen.

Anhang C

Beschreibung der internationalen Benchmarks für das Naturwissenschafts- Kompetenzmodell in TIMSS 2011

Benchmark 1 für die Gesamtskala Naturwissenschaften (Kompetenzstufe II – ab 400 Punkten) – Niedrige Benchmark –

Zusammenfassung

Die Schülerinnen und Schüler besitzen elementares Wissen über Biologie und Physik/Chemie. Sie besitzen Wissen über einfache Fakten zur menschlichen Gesundheit, zu Ökosystemen sowie zum Verhalten und Körperbau von Tieren. Die Schülerinnen und Schüler zeigen basales Wissen über Energie und kennen einige Eigenschaften von Materialien. Sie interpretieren einfache Diagramme, vervollständigen einfache Tabellen und geben kurze schriftliche Erläuterungen auf Fragen nach naturwissenschaftlichen Fakten.

In Biologie besitzen die Schülerinnen und Schüler Wissen über einfache Fakten zur menschlichen Gesundheit. Sie kennen die mögliche Wirkung der Sonne auf ungeschützte Haut und wissen, dass die Lunge das Organ des Menschen ist, das beim Rauchen am stärksten geschädigt wird. Sie verfügen über grundlegendes Wissen zum Verhalten und zum Körperbau von Tieren. Die Schülerinnen und Schüler wissen, dass Vögel ihre Eier bebrüten, um diese warm zu halten, und ordnen Flügel als bei Vögeln, Fledermäusen und Schmetterlingen üblich ein. Sie beginnen, Ökosysteme ansatzweise zu verstehen. Beispielsweise erkennen sie ein Tier, das in der Wüste lebt, klassifizieren einen Wolf als Raubtier und ordnen verschiedene Tiere den jeweiligen Ökosystemen zu.

In Physik/Chemie kennen die Schülerinnen und Schüler einige Materialeigenschaften. Sie wissen beispielsweise, dass Eis die feste Form von Wasser ist, Eisennägel rosten und Eisengegenstände allgemein schwerer sind als Gegenstände gleicher Form und Größe aus Holz oder Styropor. Darüber hinaus besitzen sie ein Grundverständnis von Kräften. In einer schematischen Darstellung erkennen sie die Richtung der Erdanziehungskraft und Wind als Ursache der Bewegung eines Segelbootes. Sie wissen, dass der Ton einer Gitarre durch Schwingungen von den Saiten ausgeht. In einem Diagramm erkennen sie den heißesten Thermometerwert.

Die Schülerinnen und Schüler interpretieren beschriftete Bilder und einfache Diagramme (z.B. Kräfte, die auf einen Quader wirken, und Thermometerwerte), vervollständigen einfache Tabellen (ordnen z.B. Tiere Ökosystemen zu) und geben kurze schriftliche Erläuterungen auf Fragen nach Sachinformationen (z.B. eine mögliche Wirkung der Sonne auf ungeschützte Haut).

Benchmark 2 für die Gesamtskala Naturwissenschaften (Kompetenzstufe III – ab 475 Punkten) – Durchschnittliche Benchmark –

Zusammenfassung

Die Schülerinnen und Schüler können Basiswissen und grundlegendes Verständnis auf naturwissenschaftsbezogene Situationen anwenden. Sie kennen grundlegende Begriffe der belebten Natur und ihrer Interaktion mit der Umwelt und weisen ein basales Verständnis der Biologie und der Gesundheit des Menschen auf. Darüber hinaus verfügen sie über ein grundlegendes Wissen über Eigenschaften von Materialien sowie über Licht, elektrischen Strom, Energie und Kräfte. Außerdem kennen sie basale Fakten über das Sonnensystem und die Ressourcen der Erde. Sie verfügen über die Fähigkeiten, Informationen aus Bilddiagrammen zu interpretieren und Faktenwissen auf naturwissenschaftsbezogene Situationen anzuwenden.

In Biologie kennen die Schülerinnen und Schüler grundlegende Fakten über den Körperbau und die Gesundheit des Menschen. Sie wissen beispielsweise, dass der Magen ein Verdauungsorgan und der Sauerstoffbedarf des Körpers beim Sport erhöht ist. Darüber hinaus ist ihnen bekannt, dass Obst und Gemüse die besten Mineral- und Vitaminquellen sind. Sie können eine Methode beschreiben, wie Zähne zusätzlich zum Zähneputzen geschützt werden können, und können einen Weg nennen, wie Grippe übertragen werden kann. Die Schülerinnen und Schüler verfügen über grundlegende Kenntnisse über Eigenschaften von Lebewesen und deren Rolle in der Umwelt. Anhand von Tierbildern ordnen die Schülerinnen und Schüler beispielsweise jedem Tier die entsprechenden Eigenschaften zu (Skelett, Säugetier oder nicht, Beinanzahl). Sie kennen die Form der Füße eines Wasservogels und wissen, dass ein Walross von Fettschichten warm gehalten wird. Weiterhin können sie eine einfache Darstellung einer Nahrungskette deuten und erkennen auf einer Bildübersicht einen Vogel, der wahrscheinlich Säugetiere frisst. Die Schülerinnen und Schüler wissen, dass Bäume durch Sonnenlicht wachsen, und beschreiben im Zusammenhang mit einer Untersuchung des Pflanzenwachstums eine Behandlung, die es einer Pflanze ermöglicht, besser zu wachsen als eine andere. Darüber hinaus verfügen sie über ein grundlegendes Verständnis von Lebenszyklen und Lebewesen; sie wissen, dass Kaulquappen aus Froschlaich schlüpfen und sich Schlangen beim Wachsen häuten. Außerdem ist ihnen bekannt, dass aus Samen neue Pflanzen wachsen.

In Physik/Chemie verfügen die Schülerinnen und Schüler über ein grundlegendes Verständnis einfacher physikalischer/chemischer Phänomene. Sie wissen, dass ein Eisennagel einen Stromkreis schließen kann, um eine Glühbirne zum Leuchten zu bringen. Anhand einer schematischen Darstellung, die eine Person zeigt, die mit einem Strohhalm in ein Glas mit Wasser bläst, erklären die Schülerinnen und Schüler, warum Luftblasen nach oben steigen. Weiterhin wissen sie, dass ein in Wasser treibender Körper leichter ist als ein Körper gleicher Form und Größe, der nach unten sinkt. Die Schülerinnen und Schüler können die Farbe eines weißen Hemds unter blauem Licht vorhersagen. Darüber hinaus können sie Faktenwissen in naturwissenschaftsbezogenen Situationen

anwenden. Sie können beispielsweise Elektrizität als Energiequelle für drei Haushaltsgegenstände in einer schematischen Darstellung benennen sowie zwei Einsatzgebiete von Elektrizität im Alltag nennen. Weiterhin sind die Schülerinnen und Schüler in der Lage, jeweils ein Einsatzgebiet von Wasser in gefrorener oder flüssiger Form sowie brennbare Materialien zu nennen.

In Geographie kennen die Schülerinnen und Schüler grundlegende Fakten über das Sonnensystem. Sie kennen beispielsweise außer der Erde zwei weitere Planeten, die die Sonne umkreisen, und können einen Unterschied zwischen Sonne und Mond angeben. Außerdem kennen sie einen Unterschied im Wetter der verschiedenen Jahreszeiten und die Wirkung der Windstärke auf ein an einem Pfosten befestigtes Band. Darüber hinaus verfügen die Schülerinnen und Schüler über erste Einsichten in Rohstoffe der Erde. Sie können zum Beispiel zwei verschiedene Einsatzgebiete von Holz nennen und erklären, warum Menschen kein Wasser direkt aus dem Meer trinken sollten.

Die Schülerinnen und Schüler verfügen über die grundsätzliche Fähigkeit, Informationen aus Bilddiagrammen zu interpretieren und Faktenwissen auf naturwissenschaftsbezogene Situationen anzuwenden sowie einfache Erklärungen physikalischer/chemischer Phänomene zu geben.

Benchmark 3 für die Gesamtskala Naturwissenschaften (Kompetenzstufe IV – ab 550 Punkten) – Hohe Benchmark –

Zusammenfassung

Die Schülerinnen und Schüler können mit ihrem Wissen und Verständnis alltägliche Phänomene erklären. Sie besitzen ein grundlegendes Verständnis des Aufbaus von Pflanzen, des Körperbaus von Tieren sowie von Fortpflanzung und Lebenszyklen. Sie besitzen Wissen über natürliche Prozesse und über die Umwelt und verfügen über grundlegende Kenntnisse über Eigenschaften von Materialien und physikalische/chemische Phänomene. Die Schülerinnen und Schüler besitzen ferner grundlegendes Wissen über das Sonnensystem und über die Struktur, die Veränderungen und Ressourcen der Erde. Sie zeigen elementare Kenntnisse und Fähigkeiten bezüglich naturwissenschaftlicher Denk- und Arbeitsweisen. Sie vergleichen, kontrastieren und ziehen einfache Schlussfolgerungen. Sie geben Erklärungen, bei denen sie ihr Wissen über naturwissenschaftliche Konzepte mit alltäglichen Erfahrungen und auch abstrakten Kontexten verbinden.

In Biologie besitzen die Schülerinnen und Schüler ein grundlegendes Verständnis des Aufbaus von Pflanzen und des Körperbaus von Tieren. Aus einer schematischen Darstellung erkennen sie den Teil einer Blütenpflanze, der Samen bildet, und wissen, dass Pflanzen durch Sonnenenergie wachsen. Die Schülerinnen und Schüler unterscheiden Fische und Meeressäuger anhand ihres Körperbaus und Verhaltens sowie Pflanzenfresser und Fleischfresser aufgrund ihrer Zahneigenschaften. Sie besitzen grundlegendes Wissen über Lebenszyklen

und Fortpflanzung. Beispielsweise können die Schülerinnen und Schüler eine Abbildung vervollständigen, die den Lebenszyklus einer Motte darstellt, und wissen, dass sich ein weibliches oder männliches Säugetier ohne das jeweils andere Geschlecht nicht fortpflanzen kann. Darüber hinaus weisen sie ein grundsätzliches Verständnis von Ökosystemen und der Umwelt auf. Sie vervollständigen eine Nahrungskette und erkennen das Verhältnis von Raubtier und Beute. In einer Abbildung eines Teichs als Ökosystem unterscheiden sie Lebewesen und nicht lebende Dinge. Weiterhin besitzen die Schülerinnen und Schüler Kenntnisse über menschliches Verhalten, das sich positiv oder negativ auf die Umwelt auswirkt.

In Physik/Chemie zeigen die Schülerinnen und Schüler grundlegendes Verständnis der Eigenschaften bekannter Materialien und physikalischer/chemischer Phänomene. Sie erkennen, dass Objekte größeren Volumens nicht notwendigerweise mehr wiegen und nennen Beispiele von Materialien, die bei Zimmertemperatur fest, flüssig oder gasförmig sind. Die Schülerinnen und Schüler beschreiben Materialveränderungen, etwa die Umwandlung einer Flüssigkeit in Gas, und verfügen über grundlegendes Verständnis von Gemengen und Lösungen. Beispielsweise nennen sie eine Methode zur Trennung eines Gemenges aus Eisenspänen und Sand und wissen, dass Salzwasser eine Lösung ist. Darüber hinaus ist den Schülerinnen und Schülern bekannt, dass sich ein Material in heißem Wasser schneller löst als in kaltem Wasser und dass sich in heißem Wasser eine größere Menge des Materials lösen lässt. Außerdem erklären sie, dass sich kleinere Stücke eines Materials schneller auflösen als größere Stücke. Die Schülerinnen und Schüler verfügen über ein grundlegendes Verständnis von Wärme und Leitfähigkeit und wissen beispielsweise, dass Metall wärmeleitfähiger ist als Holz und Eis in einem geschlossenen Behälter langsamer schmilzt als an der Luft. Sie haben Kenntnisse über Magnet- und Anziehungskräfte und über Elektrizität. In einer schematischen Darstellung vervollständigen die Schülerinnen und Schüler die Beschriftung von Magnetpolen und sagen die Bewegung zweier Magnete mit gekennzeichneten Polen voraus. Sie können ein Beispiel benennen, in dem sich ein Objekt durch Anziehungskraft bewegt, und wissen, dass die Erdanziehungskraft Objekte zu Boden fallen lässt. Weiterhin sind sie anhand der Abbildung eines Schaltkreises in der Lage zu erklären, warum eine intakte Glühbirne nicht aufleuchtet. Die Schülerinnen und Schüler verfügen über grundlegendes Verständnis der Eigenschaften von Licht. Aus einer schematischen Darstellung erkennen die Schülerinnen und Schüler die Richtung eines Schattens und können die Ursache benennen.

In Geographie besitzen die Schülerinnen und Schüler grundlegende Kenntnisse über das Sonnensystem und über Struktur, Veränderungen und Ressourcen der Erde. In einer schematischen Darstellung erkennen sie Erde, Mond und Sonne. In einer Tabelle, welche die Entfernungen zwischen Planeten angibt, benennen sie den Planeten, der der Sonne am nächsten ist, und den Planeten, auf dem wahrscheinlich die niedrigste Temperatur herrscht. Die Schülerinnen und Schüler wissen, dass der Großteil der Erdoberfläche von Wasser bedeckt ist, und können einen Vorteil von Agrarwirtschaft in Flussnähe beschreiben. Darüber hinaus können sie erklären, dass Wasser in feuchter Luft bei niedrigen Temperaturen kondensieren oder gefrieren kann und durch Kondensation Morgentau entstehen kann. Anhand von tabellarischen Daten über das Wetter erkennen die Schülerinnen und Schüler den Ort, an dem die Schneefallwahrscheinlichkeit hoch ist. Sie wissen, dass eine Gebirgsschicht, die Schalentierfossilien enthält, einst Teil des Meeresbodens war und Tierfossilien der beste Beweis dafür sind, dass es auf der Erde früher zahlreiche Tierarten gab, die heute ausgestorben sind.

Die Schülerinnen und Schüler verfügen über die grundsätzliche Kompetenz, wissenschaftliche Aufgaben auszuführen. Sie vergleichen, nennen Unterschiede und ziehen Schlüsse. Außerdem geben sie kurze Erläuterungen, für die sie ihr Wissen über naturwissenschaftliche Konzepte in Beziehung zu alltäglichen Erfahrungen mit biologischen und physikalischen/chemischen Vorgängen setzen.

Benchmark 4 für die Gesamtskala Naturwissenschaften (Kompetenzstufe V – ab 625 Punkten)

– Fortgeschrittene Benchmark –

Zusammenfassung

Die Schülerinnen und Schüler weisen ein grundlegendes Verständnis des Prozesses naturwissenschaftlichen Arbeitens auf und können ihr Wissen über naturwissenschaftliche Vorgänge und Zusammenhänge anwenden. Sie teilen ihr Wissen über Eigenschaften und Prozesse von Organismen, Fortpflanzung und Entwicklung, Ökosysteme und das Interagieren von Organismen mit Lebensräumen sowie über Faktoren gesundheitsbezogener Lebensführung aktiv mit. Sie zeigen ein grundlegendes Verständnis der Eigenschaften des Lichts, der Zusammenhänge verschiedener physikalischer oder chemischer Eigenschaften bekannter Materialien und Stoffe sowie der Kräfte im Zusammenhang mit Magnetismus und Erdanziehung. Sie zeigen ferner praktisches Wissen im Bereich der Elektrizität und weisen ein grundlegendes Verständnis des Sonnensystems sowie von Landschaftsmerkmalen und Veränderungen der Erde auf. Die Schülerinnen und Schüler besitzen elementare Fähigkeiten, Ergebnisse im Kontext einfacher Experimente zu interpretieren, Schlüsse zu ziehen, Argumente zu bewerten und diese zu vertreten.

In Biologie können die Schülerinnen und Schüler ihr Verständnis der Merkmale und der natürlichen Prozesse von Organismen sowie von Faktoren der menschlichen Gesundheit aktiv mitteilen. In Graphiken können die Schülerinnen und Schüler ein Tier mit Außenskelett, den Schutzpanzer eines Reptils und eine Gruppe von Säugetieren erkennen. Sie erklären anhand der Merkmale des Körperbaus oder des Verhaltens, warum sich Meeressäuger von Fischen unterscheiden, und erläutern anhand einer schematischen Darstellung eines Tierschädels die Funktion verschiedener Zahntypen. Die Schülerinnen und Schüler besitzen grundlegendes Wissen über die Fortpflanzung und kennen Beispiele von Tierarten, die ihre Brut pflegen. Sie sind in der Lage, eine mögliche körperliche Veränderung bei Säugetieren im Winter zu beschreiben, wissen, wie die Vogelwanderung die Überlebenschance der Vögel erhöht und kennen den Vorteil, der dem Monarchfalter dadurch entsteht, für Vögel giftig zu sein. Weiterhin beschreiben sie menschliches Verhalten, das zum Aussterben von Tierarten führen kann. Darüber hinaus können die Schülerinnen und Schüler ein Argument zur Notwendigkeit einer ausgewogenen Ernährung bewerten und vertreten, wählen aus einer Liste üblicher Speisen die beste Kalziumquelle aus und erläutern, warum Menschen viel Flüssigkeit zu sich nehmen sollten. Sie

kennen eine Vorbeugemaßnahme gegen Grippe und eine Ursache für erhöhte Körpertemperatur und wissen, dass Nahrung die nötige Energie für die Heilung einer Schnittwunde liefert.

In Physik/Chemie zeigen die Schülerinnen und Schüler ein Verständnis der Zusammenhänge verschiedener Eigenschaften bekannter Materialien und Stoffe. Sie wissen, dass Eisblöcke unabhängig von ihrer Größe im Wasser treiben, und erkennen die schematische Darstellung, die das Treibverhalten von Eis im Wasser am besten verdeutlicht. Mit Hilfe von Daten zu Eigenschaften bekannter Gegenstände erkennen die Schülerinnen und Schüler ein Objekt mit gleichen Eigenschaften; aus schematischen Daten ziehen sie Schlüsse über das relative Gewicht eines von vier Quadern. Sie können eine Möglichkeit nennen, die zum Trennen von Kugeln derselben Größe genutzt werden kann, die aus verschiedenen Metallen bestehen. Die Schülerinnen und Schüler nennen aus einer Liste bekannter Materialien den besten Wärmeleiter und können an einem Thermometer den Gefrierpunkt von Wasser zeigen. Sie kennen eine Möglichkeit zu zeigen, dass Sonnenlicht aus verschiedenen Farben besteht, und unterscheiden Gegenstände, die Licht erzeugen, von solchen, die dies nicht tun. Anhand der Beschreibung eines mehrstufigen Experiments können die Schülerinnen und Schüler die Ergebnisse beschreiben und daraus den Schluss ziehen, dass die Farbe eines Objektes unter verschiedenfarbigem Licht unterschiedlich aussieht. Sie besitzen praktisches Wissen im Bereich der Elektrizität. Anhand von zwei Schaltbildern mit unterschiedlich angeschlossenen Batterien erklären die Schülerinnen und Schüler, in welchem Schaltkreis eine Glühbirne aufleuchten kann. Außerdem können sie eine Energiequelle außer Kohle, Öl und Erdgas benennen, die zur Stromgewinnung genutzt werden kann.

In Geographie besitzen die Schülerinnen und Schüler ein grundlegendes Verständnis des Sonnensystems sowie von Landschaftsmerkmalen und Veränderungen der Erde. Sie wissen, wie lange eine Umdrehung der Erde um ihre eigene Achse und eine Umrundung der Erde um die Sonne dauert, und dass der Mond sichtbar ist, da er Sonnenlicht reflektiert. Die Schülerinnen und Schüler kennen die relative Verteilung von Land und Wasser auf der Erde und verfügen über ein grundlegendes Verständnis der Zusammensetzung der Erdoberfläche. Sie wissen, dass verrottende Pflanzen und verwesende Tiere den Boden fruchtbar machen und Pflanzen wachsen lassen. Außerdem beschreiben sie die Verwendung natürlicher Rohstoffe und Veränderungen des Bodens durch natürliche Ursachen. Die Schülerinnen und Schüler können eine Karte deuten, die die Fließrichtung eines Flusses vom Gebirge zum Meer darstellt, und einen Nachteil der Agrarwirtschaft in Flussnähe benennen.

Die Schülerinnen und Schüler weisen die grundsätzliche Kompetenz auf, Forschungsergebnisse zu interpretieren, Schlüsse zu ziehen, Argumente zu bewerten und diese zu vertreten.

Abbildungsverzeichnis

Abbildung 2.1:	Das TIMSS-Curriculum-Modell	34
Abbildung 2.2:	Staaten und Regionen, die an TIMSS 2011 mit der vierten Jahrgangsstufe teilnehmen.....	38
Abbildung 2.3:	Teilnehmer an TIMSS 2007 und TIMSS 2011	39
Abbildung 2.4:	Veränderung des Mittelwerts von Viertklässlerinnen und Viertklässlern in Deutschland beim Vergleich der Leistung auf der Gesamtskala Mathematik in Abhängigkeit von der Ausschlussquote	45
Abbildung 2.5:	Veränderung des Mittelwerts von Viertklässlerinnen und Viertklässlern in Deutschland beim Vergleich der Leistung auf der Gesamtskala Naturwissenschaften in Abhängigkeit von der Ausschlussquote.....	46
Abbildung 2.6:	Normalverteilung mit Perzentilen	62
Abbildung 2.7:	Beziehung von Benchmarks und Kompetenzstufen.....	64
Abbildung 2.8:	Darstellung von Perzentilbändern mit Konfidenzintervallen am Beispiel der Mathematikleistung in TIMSS 2011	65
Abbildung 3.1:	Zuordnung von Beispielaufgaben zu den Inhaltsbereichen und kognitiven Anforderungsbereichen.....	81
Abbildung 3.2:	Beispiele curricular nicht valider Testaufgaben.....	86
Abbildung 3.3:	Kompetenzstufen und Beispielaufgaben (Gesamtskala Mathematik) I	90
Abbildung 3.4:	Kompetenzstufen und Beispielaufgaben (Gesamtskala Mathematik) II.....	91
Abbildung 3.5:	Testleistung der Schülerinnen und Schüler im internationalen Vergleich – Gesamtskala Mathematik	96
Abbildung 3.6:	Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 – Gesamtskala Mathematik	100
Abbildung 3.7:	Prozentuale Verteilung der Schülerinnen und Schüler auf die fünf Kompetenzstufen (Mathematik) im internationalen Vergleich.....	103
Abbildung 3.8:	Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den Inhaltsbereichen <i>Arithmetik, Geometrie/Messen und Umgang mit Daten</i>	106
Abbildung 3.9:	Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 in den Inhaltsbereichen <i>Arithmetik, Geometrie/Messen und Umgang mit Daten</i>	108
Abbildung 3.10:	Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den kognitiven Anforderungsbereichen <i>Reproduzieren, Anwenden und Problemlösen</i>	110
Abbildung 3.11:	Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 in den kognitiven Anforderungsbereichen <i>Reproduzieren, Anwenden und Problemlösen</i>	111
Abbildung 3.12:	Mittlere positive Einstellung von Schülerinnen und Schülern zur Mathematik sowie prozentuale Verteilungen nach hoher, mittlerer und niedriger Einstellung – TIMSS 2007 und TIMSS 2011 im Vergleich	113
Abbildung 3.13:	Prozentuale Verteilungen und Testleistungen von Schülerinnen und Schülern in Mathematik auf den Kompetenzstufen nach positiver Einstellung	114
Abbildung 3.14:	Mittleres Selbstkonzept von Schülerinnen und Schülern in Mathematik sowie prozentuale Verteilungen nach hohem, mittlerem und niedrigem Selbstkonzept – TIMSS 2007 und TIMSS 2011 im Vergleich.....	115
Abbildung 3.15:	Prozentuale Verteilungen und Testleistungen von Schülerinnen und Schülern in Mathematik auf den Kompetenzstufen nach Selbstkonzept.....	116
Abbildung 4.1:	Zuordnung von Beispielaufgaben zu den Inhaltsbereichen und kognitiven Anforderungsbereichen.....	131
Abbildung 4.2:	Beispiele curricular nicht valider Testaufgaben.....	135
Abbildung 4.3:	Kompetenzstufen und Beispielaufgaben (Gesamtskala Naturwissenschaften) I	140
Abbildung 4.4:	Kompetenzstufen und Beispielaufgaben (Gesamtskala Naturwissenschaften) II	141

Abbildung 4.5:	Testleistung der Schülerinnen und Schüler im internationalen Vergleich – Gesamtskala Naturwissenschaften	145
Abbildung 4.6:	Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 – Gesamtskala Naturwissenschaften	149
Abbildung 4.7:	Prozentuale Verteilung der Schülerinnen und Schüler auf die fünf Kompetenzstufen (Naturwissenschaften) im internationalen Vergleich	152
Abbildung 4.8:	Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den Inhaltsbereichen <i>Biologie, Physik/Chemie</i> und <i>Geographie</i>	155
Abbildung 4.9:	Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 in den Inhaltsbereichen <i>Biologie, Physik/Chemie</i> und <i>Geographie</i>	156
Abbildung 4.10:	Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den kognitiven Anforderungsbereichen <i>Reproduzieren, Anwenden</i> und <i>Problemlösen</i>	158
Abbildung 4.11:	Vergleich der Testleistungen zwischen TIMSS 2007 und TIMSS 2011 in den kognitiven Anforderungsbereichen <i>Reproduzieren, Anwenden</i> und <i>Problemlösen</i>	159
Abbildung 4.12:	Mittlere positive Einstellung von Schülerinnen und Schülern zum Fach Sachunterricht sowie prozentuale Verteilung nach <i>hoher, mittlerer</i> und <i>niedriger</i> Einstellung – TIMSS 2007 und TIMSS 2011 im Vergleich	163
Abbildung 4.13:	Prozentuale Verteilung von Schülerinnen und Schülern auf den fünf Kompetenzstufen nach Einstellungen zum Fach Sachunterricht	163
Abbildung 4.14:	Mittleres sachunterrichtsbezogenes Selbstkonzept der Schülerinnen und Schüler sowie prozentuale Verteilung nach <i>niedrigem, mittlerem</i> und <i>hohem</i> Selbstkonzept – TIMSS 2007 und TIMSS 2011 im Vergleich	164
Abbildung 4.15:	Prozentuale Verteilung der Schülerinnen und Schüler auf den fünf Kompetenzstufen nach sachunterrichtsbezogenem Selbstkonzept	165
Abbildung 5.1:	Formen des Ganztags im Vergleich 2007 zu 2011 (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen in Prozent)	176
Abbildung 5.2:	Erweiterte Lern- und Förderangebote in Ganz- und Halbtagsgrundschulen (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen in Prozent)	177
Abbildung 5.3:	Ausstattungsprobleme im Mathematikunterricht (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen in Prozent)	181
Abbildung 5.4:	Ausstattungsprobleme im naturwissenschaftlichen Sachunterricht (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen in Prozent)	182
Abbildung 5.5:	Methodische Vorgehensweisen der Lehrkräfte (Anteile der Schülerinnen und Schüler nach Angaben der Lehrkräfte in Prozent)	185
Abbildung 5.6:	Computernutzung der Lehrkräfte (Anteile der Schülerinnen und Schüler nach Angaben der Lehrkräfte in Prozent)	188
Abbildung 5.7:	Altersverteilung der Lehrkräfte (Anteile der Schülerinnen und Schüler nach Angaben der Lehrkräfte in Prozent)	194
Abbildung 5.8:	Fortbildungsmöglichkeiten der Lehrkräfte (Anteile der Schülerinnen und Schüler nach Angaben der Lehrkräfte in Prozent)	196
Abbildung 5.9:	Zeit für Schulleiterfortbildung (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen, Kategorie „keine Zeit“)	198
Abbildung 6.1:	Mathematikleistung nach Geschlecht – Gesamtskala	209
Abbildung 6.2:	Leistungen in mathematischen Inhaltsbereichen nach Geschlecht im internationalen Vergleich	211
Abbildung 6.3:	Mittlere Leistungen und prozentuale Verteilungen von Jungen und Mädchen auf den Kompetenzstufen der Mathematik	212
Abbildung 6.4:	Vergleich der Leistungsunterschiede in Mathematik von Jungen und Mädchen in TIMSS 2007 und TIMSS 2011 (internationaler Vergleich)	213
Abbildung 6.5:	Naturwissenschaftsleistung nach Geschlecht – Gesamtskala	215

Abbildung 6.6:	Leistungen in naturwissenschaftlichen Inhaltsbereichen nach Geschlecht im internationalen Vergleich.....	216
Abbildung 6.7:	Mittlere Leistungen und prozentuale Verteilungen von Jungen und Mädchen auf den Kompetenzstufen der Naturwissenschaften.....	218
Abbildung 6.8:	Vergleich der Leistungsunterschiede in den Naturwissenschaften von Jungen und Mädchen in TIMSS 2007 und TIMSS 2011 (internationaler Vergleich).....	219
Abbildung 6.9:	Prozentuale Verteilung auf der Skala ‚positive Einstellung zur Mathematik‘ im Vergleich.....	221
Abbildung 6.10:	Prozentuale Verteilung auf der Skala ‚mathematikbezogenes Selbstkonzept‘ im Vergleich.....	222
Abbildung 6.11:	Prozentuale Verteilung auf der Skala ‚positive Einstellung zu dem Fach Sachunterricht‘ im Vergleich.....	223
Abbildung 6.12:	Prozentuale Verteilung auf der Skala ‚sachunterrichtsbezogenes Selbstkonzept‘ im Vergleich.....	223
Abbildung 7.1:	Leistungsvorsprung in der Mathematikkompetenz von Kindern aus Familien mit mehr als 100 Büchern vor denen mit maximal 100 Büchern.....	236
Abbildung 7.2:	Leistungsvorsprung in der Naturwissenschaftskompetenz von Kindern aus Familien mit mehr als 100 Büchern vor denen mit maximal 100 Büchern.....	237
Abbildung 7.3:	Unterschiede im Leistungsvorsprung in der Mathematikkompetenz von Kindern aus Familien mit mehr als 100 Büchern vor denen mit maximal 100 Büchern im Vergleich von TIMSS 2007 und TIMSS 2011.....	238
Abbildung 7.4:	Unterschiede im Leistungsvorsprung in der Naturwissenschaftskompetenz von Kindern aus Familien mit mehr als 100 Büchern vor denen mit maximal 100 Büchern im Vergleich von TIMSS 2007 und TIMSS 2011.....	239
Abbildung 7.5:	Armutsgefährdungsquoten nach Migrationshintergrund, höchstem Bildungsabschluss und EGP-Klasse (in Prozent).....	242
Abbildung 8.1:	Mathematische Kompetenzen der Schülerinnen und Schüler nach zu Hause gesprochener Sprache im internationalen Vergleich.....	250
Abbildung 8.2:	Naturwissenschaftliche Kompetenzen der Schülerinnen und Schüler nach zu Hause gesprochener Sprache im internationalen Vergleich.....	251
Abbildung 8.3:	Mathematische Kompetenzen von Schülerinnen und Schülern mit unterschiedlichem Migrationshintergrund in Deutschland im Vergleich.....	258
Abbildung 8.4:	Naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit unterschiedlichem Migrationshintergrund in Deutschland im Vergleich.....	259
Abbildung 8.5:	Unterschiede im Leistungsvorsprung in mathematischen Kompetenzen von Kindern mit unterschiedlichem Migrationshintergrund (nach Geburtsland der Eltern) in Deutschland – TIMSS 2007 und 2011 im Vergleich.....	259
Abbildung 8.6:	Unterschiede im Leistungsvorsprung in naturwissenschaftlichen Kompetenzen von Kindern mit unterschiedlichem Migrationshintergrund (nach Geburtsland der Eltern) in Deutschland – TIMSS 2007 und 2011 im Vergleich.....	260
Abbildung 8.7:	Anteile von Kindern mit und ohne Migrationshintergrund in Deutschland auf den unterschiedlichen Kompetenzstufen der Mathematik – TIMSS 2007 und 2011 im Vergleich (Angaben in Prozent).....	260
Abbildung 8.8:	Anteile von Kindern mit und ohne Migrationshintergrund in Deutschland auf den unterschiedlichen Kompetenzstufen der Naturwissenschaften – TIMSS 2007 und 2011 im Vergleich (Angaben in Prozent).....	261
Abbildung 9.1:	Verteilungen von Schülerinnen und Schülern auf die Kompetenzstufen in den Domänen Leseverständnis, Mathematik und Naturwissenschaften (Anteile in Prozent).....	273

Abbildung 9.2:	Verteilung von Schülerinnen und Schülern auf die Kompetenzstufen, kombiniert für die Domänen Leseverständnis, Mathematik und Naturwissenschaften (Anteile in Prozent)	275
Abbildung 9.3:	Hierarchie der ein-, drei- und achtdimensionalen Kompetenzmodelle	278
Abbildung 9.4:	Die Leistungsprofile der Grundschülerinnen und Grundschüler in Deutschland, bezogen auf die drei Kompetenzbereiche (berechnet in der Stichprobe)	283
Abbildung 9.5:	Die Leistungsprofile der Grundschülerinnen und Grundschüler in Deutschland, bezogen auf die acht inhaltlichen Subdomänen der drei Kompetenzbereiche (berechnet in der Stichprobe, deskriptiv basierend auf der in Abschnitt 3 gewonnenen 7-Profile-Lösung)	285
Abbildung 9.6:	Zusammensetzung der sieben Leistungstypen nach Geschlecht (berechnet in der Stichprobe, deskriptiv basierend auf der in Abschnitt 3 gewonnenen 7-Profile-Lösung)	289

Tabellenverzeichnis

Tabelle 2.1:	Inhaltsbereiche und kognitive Anforderungsbereiche in TIMSS 2011	36
Tabelle 2.2:	Vergleichsgruppen in TIMSS 2011	39
Tabelle 2.3:	Testheftdesign in TIMSS 2011	49
Tabelle 2.4:	Untersuchungsablauf von TIMSS 2011 und IGLU/PIRLS 2011	54
Tabelle 3.1:	IQB-Kompetenzstufenmodell für das Fach Mathematik in der Primarstufe (Jahrgangsstufe 4)	77
Tabelle 3.2:	Themengebiete der einzelnen mathematischen Inhaltsbereiche.....	79
Tabelle 3.3:	Kognitive Aktivitäten der einzelnen Anforderungsbereiche.....	80
Tabelle 3.4:	Verteilung der Testaufgaben auf die mathematischen Inhaltsbereiche.....	83
Tabelle 3.5:	Verteilung der Testaufgaben auf die kognitiven Anforderungsbereiche	83
Tabelle 3.6:	Merkmale von curricular nicht validen Testaufgaben	85
Tabelle 3.7:	Anteile curricular nicht valider Testaufgaben nach mathematischen Inhaltsbereichen	86
Tabelle 3.8:	Anteile curricular nicht valider Testaufgaben nach kognitiven Anforderungsbereichen.....	87
Tabelle 3.9:	Klassifikation curricular valider TIMSS-Mathematikaufgaben gemäß mathematischer Leitideen der Bildungsstandards.....	87
Tabelle 3.10:	Beschreibung der fünf Kompetenzstufen für die Gesamtskala Mathematik	89
Tabelle 3.11:	Testwerte und Rangplätze Mathematik – Vergleich der Skalierungsmodelle mit und ohne für Deutschland curricular validen Testaufgaben	98
Tabelle 3.12:	Mathematikleistungen nach Perzentilen im Vergleich: TIMSS 2007 und TIMSS 2011 – Gesamtskala Mathematik	101
Tabelle 3.13:	Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007 und in TIMSS 2011 (Angaben in Prozent)	104
Tabelle 4.1:	Themengebiete der einzelnen naturwissenschaftlichen Inhaltsbereiche.....	129
Tabelle 4.2:	Kognitive Aktivitäten der einzelnen Anforderungsbereiche	130
Tabelle 4.3:	Verteilung der Testaufgaben auf die naturwissenschaftlichen Inhaltsbereiche	133
Tabelle 4.4:	Verteilung der Testaufgaben auf die kognitiven Anforderungsbereiche	133
Tabelle 4.5:	Anteile curricular nicht valider Testaufgaben nach naturwissenschaftlichen Inhaltsbereichen	135
Tabelle 4.6:	Übergeordnete Themen von curricular nicht validen Testaufgaben	135
Tabelle 4.7:	Anteile curricular nicht valider Testaufgaben nach kognitiven Anforderungsbereichen.....	136
Tabelle 4.8:	Beschreibung der fünf Kompetenzstufen für die Gesamtskala Naturwissenschaften.....	138
Tabelle 4.9:	Testwerte und Rangplätze im Bereich Naturwissenschaften – Skala mit für Deutschland curricular validen Testaufgaben	147
Tabelle 4.10:	Naturwissenschaftliche Kompetenz nach Perzentilen im Vergleich: TIMSS 2007 und TIMSS 2011 – Gesamtskala Naturwissenschaften	150
Tabelle 4.11:	Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007 und in TIMSS 2011 (Angaben in Prozent)	153
Tabelle 5.1:	Anzahl der Computer (absolute Zahlen) für Unterrichtszwecke und die Verfügbarkeit im Fachunterricht (in Prozent)	179
Tabelle 5.2:	Aspekte der Schülerkomposition aus Sicht der Schulleitungen (Anteile der Schülerinnen und Schüler nach Angaben der Schulleitungen in Prozent)	191
Tabelle 5.3:	Aspekte der Schülerkomposition aus Sicht der Eltern und Schülerinnen und Schüler (Anteile der Schülerinnen und Schüler in Prozent)	192
Tabelle 5.4:	Fortbildungsinhalte der Mathematiklehrkräfte in den letzten zwei Jahren (Anteile der Schülerinnen und Schüler nach Angaben der Lehrkräfte in Prozent)	197

Tabelle 6.1:	Regression von Leistungsunterschieden in Mathematik und Naturwissenschaften auf das Geschlecht, das mathematikbezogene bzw. naturwissenschaftsbezogene Selbstkonzept, den höchsten Bildungsabschluss der Eltern und den sozialen Status.....	225
Tabelle 7.1:	Mittlere Mathematik- und Naturwissenschaftskompetenz je EGP-Klasse in Deutschland bei TIMSS 2007 und TIMSS 2011	241
Tabelle 7.2:	Mittlere Mathematik- und Naturwissenschaftskompetenz nach Armutgefährdung und EGP-Klasse	243
Tabelle 7.3:	Regressionsanalysen zum Interaktionseffekt zwischen HISEI und Armutgefährdung bei der Erklärung von Mathematik- und Naturwissenschaftskompetenz (TIMSS 2007 und 2011)	243
Tabelle 8.1:	Verteilung der Kinder mit und ohne Migrationshintergrund auf unterschiedliche Kompetenzstufen der Mathematik (Angaben in Prozent)	252
Tabelle 8.2:	Verteilung der Kinder mit und ohne Migrationshintergrund auf unterschiedliche Kompetenzstufen der Naturwissenschaften (Angaben in Prozent)	255
Tabelle 8.3:	Merkmale von Familien. Angaben zur Verteilung von Viertklässlerinnen und Viertklässlern, Familienstruktur, Kinderanzahl und zum Erwerbsstatus von Mutter und Vater sowie zum Kindergartenbesuch nach Migrationsstatus und familiärem Sprachgebrauch (Angaben in Prozent)	257
Tabelle 8.4:	Regressionsmodell zur Erklärung von Unterschieden in mathematischen und naturwissenschaftlichen Kompetenzen zwischen Kindern mit und ohne Migrationshintergrund (Angaben in Skalenpunkten der Gesamtskalen Mathematik und Naturwissenschaften) – Vergleich von TIMSS 2007 und 2011	263
Tabelle 9.1:	Kurzfassung der Kompetenzstufenbeschreibungen für Leseverständnis, Mathematik und Naturwissenschaften.....	272
Tabelle 9.2:	Ergebnisse des Modellvergleichs (Devianz und Informationskriterien AIC, CAIC und BIC).....	278
Tabelle 9.3:	Latente Korrelationen für das dreidimensionale domänenbezogene MCMLM: Leseverständnis, Mathematik und Naturwissenschaften	279
Tabelle 9.4:	Latente Korrelationen für das achtdimensionale subdomänenbezogene MCMLM: <i>Informierende Texte</i> und <i>Literarische Texte</i> (Leseverständnis), <i>Arithmetik</i> , <i>Geometrie/Messen</i> und <i>Umgang mit Daten</i> (Mathematik), <i>Biologie</i> , <i>Physik/Chemie</i> und <i>Geographie</i> (Naturwissenschaften)	280
Tabelle 9.5:	Vergleichswerte BIC und CAIC über verschiedene Anzahlen an identifizierten Profilen	282
Tabelle 9.6:	Die Leistungsprofile der Grundschülerinnen und Grundschüler in Deutschland, bezogen auf die drei Kompetenzdomänen (berechnet in der Stichprobe)	282
Tabelle 9.7:	Zusammensetzung der 7 Leistungstypen nach kulturellen und sozioökonomischen Merkmalen, Migrationshintergrund und Familiensprache sowie besonderem Förderbedarf (berechnet in der Stichprobe, deskriptiv basierend auf der in Abschnitt 3 gewonnenen 7-Profil-Lösung)	290
Tabelle 9.8:	Zusammensetzung der sieben Leistungstypen nach fachbezogener Einstellung und Selbstkonzept (berechnet in der Stichprobe, deskriptiv basierend auf der in Abschnitt 3 gewonnenen Profillösung) – Anteile von Schülerinnen und Schülern mit hohen positiven Einstellungen zum Lesen, zur Mathematik und zum Sachunterricht sowie Anteile mit hohem lese-, mathematik- und sachunterrichtsbezogenem Selbstkonzept nach den sieben Leistungstypen	293
Tabelle A.1:	Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2007	303
Tabelle A.2:	Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2011	304

Wilfried Bos, Irmela Tarelli,
Albert Bremerich-Vos, Knut Schwippert
(Hrsg.)

IGLU 2011

Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich

2012, 274 Seiten, br., 29,90 Euro,
ISBN 978-3-8309-2828-7



In diesem Band werden die Ergebnisse der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) für die deutsche Bildungsdiskussion aus 2011 vorgestellt. Die Lesekompetenzen der Grundschul Kinder werden anhand von zentralen Merkmalen wie soziale Herkunft, Migrationsstatus, Geschlecht und Lehr- und Lernbedingungen sowie mit Hinblick auf den Übergang von der Primar- in die Sekundarstufe I analysiert. Im Jahr 2011 wurde IGLU gemeinsam mit ihrer Schwesterstudie, der *Trends in International Mathematics and Science Study* (TIMSS 2011), durchgeführt. In diesem Band werden erste Ergebnisse einer gemeinsamen Auswertung von Zusammenhängen der drei erfassten Kompetenzbereiche Leseverständnis (IGLU), Mathematik und Naturwissenschaften (TIMSS) untersucht und kompetenzübergreifende Leistungsprofile von Schülerinnen und Schülern am Ende der Grundschulzeit beschrieben.



WAXMANN

Münster · New York · München · Berlin

Petra Stanat, Hans Anand Pant,
Katrin Böhme, Dirk Richter (Hrsg.)

Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik

Ergebnisse des
IQB-Ländervergleichs 2011

2012, 300 Seiten, br., 32,90 Euro,
ISBN 978-3-8309-2777-8



In diesem Band wird über den ersten Ländervergleich des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) im Primarbereich berichtet. Im Fokus der Studie stehen die Kompetenzen Lesen und Zuhören im Fach Deutsch sowie die mathematische Kompetenz, insgesamt und differenziert nach den fünf inhaltlichen Leitideen. Daneben werden auch Zusammenhänge zwischen den erreichten Kompetenzen und verschiedenen Hintergrundmerkmalen der Schülerinnen und Schüler analysiert. Ergänzend werden Informationen zu den im Primarbereich tätigen Lehrkräften sowie zu Maßnahmen der Sprach- und Leseförderung in der Grundschule berichtet. Die repräsentativen Erhebungen für den Ländervergleich fanden 2011 an insgesamt 1.349 Schulen statt.



WAXMANN

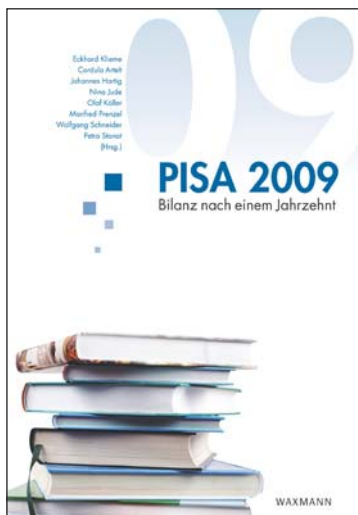
Münster · New York · München · Berlin

Eckhard Klieme, Cordula Artelt,
Johannes Hartig, Nina Jude,
Olaf Köller, Manfred Prenzel,
Wolfgang Schneider, Petra Stanat
(Hrsg.)

PISA 2009

Bilanz nach einem Jahrzehnt

2010, 310 Seiten, br., 24,90 Euro,
ISBN 978-3-8309-2450-0



Im Jahr 2009 hat Deutschland bereits zum vierten Mal am *Programme for International Student Assessment (PISA)* der OECD teilgenommen. Erfasst wurden die Kompetenzen von fünfzehnjährigen Schülerinnen und Schülern in den Bereichen Lesen, Mathematik und Naturwissenschaften in 65 Staaten weltweit.

Dieses Buch stellt den Ist-Stand im Jahr 2009 dar und verknüpft ihn mit den Entwicklungen zwischen den Jahren 2000 und 2009. Über welche Kompetenzen verfügen deutsche Schülerinnen und Schüler, und wie haben sich diese verändert? Wie haben sich Rahmenbedingungen und Ergebnisse von Bildungsprozessen, aber auch Schulen und außerschulische Faktoren verändert? Wie lässt sich diese Entwicklung beurteilen, wenn man sie in den internationalen Vergleich einordnet?

Eine unverzichtbare Grundlage für eine professionelle Debatte über PISA.

Jörg Schlömerkemper in: Pädagogik, 6/2011.



WAXMANN

Münster · New York · München · Berlin