

PISA 2012
Fortschritte und Herausforderungen in Deutschland

Manfred Prenzel, Christine Sälzer,
Eckhard Klieme, Olaf Köller (Hrsg.)

PISA 2012

Fortschritte und Herausforderungen
in Deutschland



Waxmann 2013
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 978-3-8309-3001-3

© 2013 Waxmann Verlag GmbH
Postfach 8603, 48046 Münster
Waxmann Publishing Co.
P.O. Box 1318, New York, NY 10028, USA

www.waxmann.com
info@waxmann.com

Buchumschlag: Christian Schorm, Berlin; Inna Ponomareva, Münster
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 des Präsidenten der Kultusministerkonferenz und der Bundesministerin für Bildung und Forschung	9
1 PISA 2012 – eine Einführung in die aktuelle Studie	11
<i>Christine Sälzer und Manfred Prenzel</i>	
1.1 Zielsetzung und Anlage des <i>Programme for International Student Assessment</i>	13
1.1.1 Funktionen eines internationalen Bildungsvergleichs	13
1.1.2 Grundlegende Merkmale von PISA	17
1.1.3 Missverständnisse und Kritik an PISA	20
1.2 Die Besonderheiten von PISA 2012	21
1.2.1 Teilnehmende Staaten	21
1.2.2 Testkonzeption und Testdesign	22
1.2.3 Zusatzprogramm und weitere internationale Optionen	26
1.3 Die Durchführung von PISA in Deutschland	26
1.3.1 Population und Stichprobe: Wer nimmt an PISA teil?	27
1.3.2 Teilnahmeverpflichtung	32
1.3.3 Internationale Vergleichbarkeit der Ergebnisse	33
1.3.4 Skalierung und Auswertung der Daten	37
1.4 Darstellung der Ergebnisse aus PISA	38
1.5 Interpretation der Befunde	42
1.6 Nationale und internationale Organisation der PISA-Studie	43
Literatur	44
2 Zwischen Grundlagenwissen und Anwendungsbezug: Mathematische Kompetenz im internationalen Vergleich	47
<i>Christine Sälzer, Kristina Reiss, Anja Schiepe-Tiska, Manfred Prenzel und Aiso Heinze</i>	
2.1 Einleitung	47
2.2 Mathematische Kompetenz in PISA	50
2.2.1 Elemente der Rahmenkonzeption: Inhalte, Prozesse und Kontexte	51
2.2.2 Fundamentale mathematische Fähigkeiten	53
2.3 Der Mathematiktest in PISA 2012 und seine Auswertung	57
2.3.1 Aufgabenformate	59
2.3.2 Skalierung und Kompetenzstufen	59
2.3.3 Aufgabenbeispiele	60
2.4 Ergebnisse des internationalen Vergleichs	70
2.4.1 Ergebnisse des internationalen Vergleichs auf der Gesamtskala	70
2.4.2 Verteilungen auf die Kompetenzstufen	73
2.4.3 Kompetenzunterschiede zwischen Jungen und Mädchen	75
2.4.4 Ergebnisse in den vier Inhaltsbereichen mathematischer Kompetenz	77

2.4.5	Veränderungen seit PISA 2003	84
2.5	Ergebnisse für Deutschland	86
2.5.1	Kompetenzen in verschiedenen Schularten	86
2.5.2	Mathematische Kompetenz in Deutschland im Trend betrachtet	89
2.6	Zusammenfassung und Diskussion	92
Literatur	95

3 Mathematikbezogene emotionale und motivationale Orientierungen, Einstellungen und Verhaltensweisen von Jugendlichen in PISA 2012

Anja Schiepe-Tiska und Stefanie Schmidtner

3.1	Die Rahmenkonzeption der PISA-Fragebögen 2012	100
3.2	Beschreibung der erhobenen Schülermerkmale	101
3.2.1	Emotionale und motivationale Orientierungen	102
3.2.2	Selbstbilder	103
3.2.3	Mathematikbezogene Einstellungen und Verhaltensweisen	104
3.3	Ergebnisse	105
3.3.1	Emotionale und motivationale Orientierungen im internationalen Vergleich	106
3.3.2	Selbstbilder im internationalen Vergleich	109
3.3.3	Mathematikbezogene Einstellungen und Verhaltensweisen im internationalen Vergleich	111
3.3.4	Veränderungen der Schülermerkmale seit PISA 2003	113
3.3.5	Zusammenhänge der Schülermerkmale untereinander	116
3.4	Zusammenfassung und Diskussion	117
Literatur	118

4 Mathematikunterricht in Deutschland: Befunde aus PISA 2012

Anja Schiepe-Tiska, Kristina Reiss, Andreas Obersteiner, Jörg-Henrik Heine, Tina Seidel und Manfred Prenzel

4.1	Merkmale guten Unterrichts	123
4.1.1	Befunde aus Metaanalysen	123
4.1.2	Unterrichtsmuster	125
4.1.3	Unterricht als Lernangebot	126
4.2	Fragen und Erwartungen	127
4.3	Erfassung des Unterrichts bei PISA	128
4.4	Ergebnisse	133
4.4.1	Charakterisierung des Mathematikunterrichts in Deutschland anhand ausgewählter Items	133
4.4.2	Mathematikunterricht in Deutschland im internationalen Vergleich	138
4.4.3	Zwischenfazit	143
4.4.4	Muster des Mathematikunterrichts	143
4.5	Zusammenfassung und Diskussion	149
Literatur	151

5	Schulische Rahmenbedingungen der Kompetenzentwicklung	155
	<i>Christine Sälzer, Manfred Prenzel und Eckhard Klieme</i>	
5.1	Institutionelle und organisatorische Rahmenbedingungen	158
5.1.1	Leistungsvarianz und Gliederung des Sekundarschulwesens	158
5.1.2	Alter und Klassenstufe der PISA-Kohorte	163
5.1.3	Lernzeit	168
5.2	Die Einzelschule als Ort des Lernens	171
5.2.1	Klassengröße	171
5.2.2	Schulentwicklung und Qualitätssicherung	173
5.2.3	Schülerverhalten und Schulklima	176
5.3	Zusammenfassung und Diskussion	183
	Literatur	184
6	Naturwissenschaftliche Kompetenz in PISA 2012: Ergebnisse und Herausforderungen	189
	<i>Anja Schiepe-Tiska, Katrin Schöps, Silke Rönnebeck, Olaf Köller und Manfred Prenzel</i>	
6.1	Naturwissenschaftliche Kompetenz in PISA 2012	191
6.2	Naturwissenschaftliche Kompetenz im internationalen Vergleich	197
6.3	Vertiefende Analysen der naturwissenschaftlichen Kompetenz in Deutschland ...	205
6.3.1	Unterschiede zwischen Schularten	205
6.3.2	Veränderung der naturwissenschaftlichen Kompetenz seit PISA 2006	208
6.4	Diskussion und Ausblick	211
	Literatur	212
7	Lesekompetenz in PISA 2012: Veränderungen und Perspektiven	217
	<i>Katharina Hohn, Anja Schiepe-Tiska, Christine Sälzer und Cordula Artelt</i>	
7.1	Lesekompetenz in PISA 2012	219
7.2	Lesekompetenz im internationalen Vergleich	227
7.3	Vertiefende Analysen der Lesekompetenz in Deutschland	234
7.3.1	Unterschiede zwischen Schularten	234
7.3.2	Unterschiede zwischen Mädchen und Jungen in der Lesekompetenz	238
7.3.3	Veränderung der Lesekompetenz seit PISA 2000	239
7.4	Diskussion und Ausblick	242
	Literatur	243
8	Soziale Herkunft als Bedingung der Kompetenzentwicklung	245
	<i>Katharina Müller und Timo Ehmke</i>	
8.1	Einleitung	245
8.2	Erfassung der sozialen Herkunft in PISA	247
8.3	Befunde zum Zusammenhang zwischen sozialer Herkunft und mathematischer Kompetenz im internationalen Vergleich	249
8.4	Die soziale Herkunft der Jugendlichen in Deutschland: vertiefende Analysen	262
8.5	Zusammenfassung und Diskussion	270
	Literatur	273

9	Mathematische Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund	275
	<i>Markus Gebhardt, Dominique Rauch, Julia Mang, Christine Sälzer und Petra Stanat</i>	
9.1	Methodische Vorbemerkungen	277
9.2	Jugendliche mit Zuwanderungshintergrund im internationalen Vergleich	280
9.3	Jugendliche mit Zuwanderungshintergrund in Deutschland	289
9.4	Zusammenfassung und Diskussion	304
	Literatur	307
10	Technische Grundlagen des fünften internationalen Vergleichs	309
	<i>Jörg-Henrik Heine, Christine Sälzer, Lars Borchert, Heiko Sibberns und Julia Mang</i>	
10.1	Die Stichprobe der PISA-Teilnehmer in Deutschland	309
10.1.1	Stichprobenplan und Ziehung der Stichprobe	310
10.1.2	Realisierte Stichprobe	319
10.1.3	Gewichtung	320
10.2	Skalierung der Leistungstests zur Modellierung von Kompetenzen	322
10.2.1	Das PISA-Multi-Matrix-Design und IRT-Skalierung	323
10.2.2	Latente Korrelationen als mehrdimensionale Modellierung	335
10.2.3	Latente Regression im Hintergrundmodell	336
10.3	Reliabilität und Stichprobenfehler	337
10.3.1	Designeffekte und Berechnung von Stichprobenfehlern	338
10.4	Bestimmung von Kompetenzstufen	339
10.4.1	Kompetenzstufen und Lösungswahrscheinlichkeiten	340
10.4.2	Kompetenzstufen für Mathematik, Lesen und Naturwissenschaften	340
10.5	Skalierung von Veränderungen zwischen PISA 2000 und PISA 2012	341
	Literatur	344
	Anhang	347
	Abbildungsverzeichnis	359
	Tabellenverzeichnis	361
	Die Autorinnen und Autoren dieses Berichtsbandes	365

Vorwort des Präsidenten der Kultusministerkonferenz und der Bundesministerin für Bildung und Forschung

Mit PISA 2012 wird der nunmehr fünfte PISA-Bericht für Deutschland veröffentlicht. Damit gewinnen wir erneut wichtige Erkenntnisse über die Kompetenzentwicklung 15-jähriger Schülerinnen und Schüler im internationalen Vergleich. Als Schwerpunkt greift PISA 2012 die Frage der Entwicklung der Leistungsfähigkeit der Schülerinnen und Schüler in den mathematischen Kompetenzen seit 2003 auf.

Für die Bildungspolitik sind die PISA-Studien zu einem wichtigen Instrument zur Verbesserung des Bildungssystems geworden. Die enttäuschenden Ergebnisse von PISA 2000 waren für Deutschland ein bildungspolitischer Weckruf.

Seitdem ist viel passiert: Maßnahmen zur Verbesserung der Sprachkompetenz, zur wirksamen Förderung bildungsbenachteiligter Kinder, zur Sicherung der Qualität von Unterricht und Schule, zur Verbesserung der Professionalität von Lehrkräften und der Ausbau von Ganztagsangeboten wurden realisiert. Die öffentlichen Investitionen in Bildung sind zwischen 2005 und 2012 um 23,6 Mrd. Euro gestiegen. Das bedeutet einen Anstieg von fast 30 Prozent.

Auch haben Bund und Länder die Bildungsforschung in Deutschland massiv ausgebaut. Unser gemeinsames Interesse ist es, mehr Erkenntnisse über Bildungsverläufe und über die Wirksamkeit von Maßnahmen zu gewinnen.

PISA 2012 zeigt, dass sich diese Anstrengungen lohnen. Die Kompetenzen der deutschen Schülerinnen und Schüler im Lesen, in der Mathematik und in den Naturwissenschaften sind seit Beginn der PISA-Studien in allen Bereichen kontinuierlich angestiegen.

Der vorliegende Bericht zu PISA 2012 wurde erstmals durch das von Bund und Ländern gemeinsam getragene Zentrum für internationale Vergleichsstudien (ZIB) erstellt, das zunächst bis 2016 für die nationale Durchführung von PISA verantwortlich zeichnet. Als Verbund der Technischen Universität München (TUM), des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) und des Leibniz-Instituts für die Pädagogik der Naturwissenschaften und Mathematik (IPN) vereint das ZIB drei herausragende Institutionen deutscher Bildungsforschung.

Wir danken den Autoren des ZIB, die diesen Bericht zu PISA 2012 nach hohen wissenschaftlichen Standards verfasst haben, und wünschen dieser Veröffentlichung viele Leserinnen und Leser. Unser Dank gilt auch den Schülerinnen und Schülern, den Lehrkräften sowie Schulleitungen, die an der Vorbereitung und Durchführung der Tests beteiligt waren.

Berlin, im Oktober 2013

Stephan Dorgerloh
Kultusminister des Landes Sachsen-Anhalt
Präsident der Ständigen Konferenz der
Kultusminister der Länder in der
Bundesrepublik Deutschland

Prof. Dr. Johanna Wanka
Bundesministerin für Bildung
und Forschung

1 PISA 2012 – eine Einführung in die aktuelle Studie

Christine Sälzer und Manfred Prenzel

Das *Programme for International Student Assessment* (PISA) hat 2012 die fünfte Erhebungsrunde erfolgreich gemeistert. Seitdem die *Organisation for Economic Co-operation and Development* (OECD) im Jahr 2000 die PISA-Studien startete, bearbeiten weltweit alle drei Jahre fünfzehnjährige Schülerinnen und Schüler ein umfangreiches Erhebungsprogramm, das aus Tests und Fragebogen besteht.

Die steigende Beteiligung von Staaten – an PISA 2012 nahmen 65 Staaten (gegenüber 32 bei PISA 2000) teil – bestätigt das nach wie vor hohe internationale Interesse an einem regelmäßigen Bildungsmonitoring, das zuverlässig und aussagekräftig über das erreichte Bildungsniveau informiert, auf Herausforderungen aufmerksam macht und im Sinne eines Benchmarkings Hinweise auf Verbesserungen gibt.

In Deutschland ist PISA ein wichtiger Bestandteil der Gesamtstrategie zum Bildungsmonitoring, die 2006 von der Ständigen Konferenz der Kultusminister der Länder verabschiedet wurde (KMK, 2006). PISA dient vor allem dazu, die Leistungsfähigkeit der Schülerinnen und Schüler am Ende der Sekundarstufe international zu verankern. Das Interesse der KMK richtet sich insbesondere auf Vergleichsperspektiven, die Entwicklungen über die Zeit (im Sinne von Trends) erfassen. Für die Steuerung von Bildungssystemen sind klar interpretierbare Trendinformationen von herausragender Bedeutung, denn sie erlauben es, Wirkungen von Maßnahmen abzuschätzen und problematische Entwicklungen frühzeitig zu erkennen. Das steuerungsrelevante Feedback aus trendfähigen Studien rechtfertigt letztlich den ambitioniert kurzen, nämlich dreijährigen Abstand zwischen den einzelnen PISA-Erhebungen. Vor dem Hintergrund zahlreicher und zum Teil umfassender Maßnahmen zur Sicherung und Verbesserung der Qualität der Schulen in Deutschland werden die Befunde der aktuellen Studie wieder mit großer Spannung erwartet. Unter anderem können sie Auskunft darüber geben, inwieweit die Ziele von Maßnahmen, die mit der Einführung länderübergreifender Bildungsstandards ab dem Jahre 2003 umgesetzt wurden (vgl. KMK, 2010), inzwischen erreicht worden sind: ein insgesamt höheres Niveau bei deutlicher Reduzierung schwacher Leistungen und die Verringerung von Disparitäten.

In der nun fünften Erhebungsrunde sollte die Rezeption der Berichterstattung auch in Deutschland mit weniger Aufregung als früher verbunden sein. Allerdings wird die Öffentlichkeit – wie auch schon bei PISA 2009 – von der OECD auf der einen Seite

mit mehreren (voraussichtlich sechs) sehr umfangreichen Berichtsbänden, auf der anderen Seite mit einer überschaubaren Anzahl einfacher und informationsarmer Rankings konfrontiert werden. Die Berichtsbände erfüllen wichtige Dokumentationszwecke aus einer weltweiten Perspektive. Auch wenn vermutlich an einzelnen Stellen in den OECD-Berichten Ergebnisse für Deutschland besonders markiert und kommentiert werden, dürfte es schwierig sein, einen kohärenten Gesamteindruck zu gewinnen, vor allem aber, die Ergebnisse aus einer deutschen Perspektive (vor dem Hintergrund der Besonderheiten des Schulsystems und der zwischenzeitlich ergriffenen Maßnahmen) zu interpretieren. In Abstimmung mit der KMK und dem BMBF dient der vorliegende Bericht dazu, einen systematischen Überblick über die Ergebnisse für Deutschland zu liefern. Er stellt nicht nur die Befunde für Deutschland in den Vordergrund, sondern soll durch die Erläuterung von Testkonzeptionen, Methoden und Ergebnissen sowie durch Hinweise auf spezielle Bedingungen und Maßnahmen helfen, die Ergebnisse und vor allem Trends zu interpretieren. Wie in den Runden vorher wurden auch diesmal umfassende Datenanalysen auf deutscher Seite durchgeführt und zum Teil zusätzliche Daten aus nationalen Ergänzungen einbezogen. Allerdings erschwert der den nationalen Teams erst sehr spät gewährte Zugriff auf die internationalen Datensätze die an sich wünschenswerte, zeitgleich mit der OECD durchgeführte Analyse der PISA-Daten durch die nationalen Arbeitsgruppen.

Über fünf Erhebungsrounden sind die grundlegenden Strukturen der Studie etabliert und viele Prozeduren zur Routine geworden. Auch wenn wichtige Merkmale von PISA bekannt sein dürften, werden in diesem einführenden Kapitel zunächst noch einmal die Zielsetzungen und das Grunddesign der Studie vorgestellt. Bei einem internationalen Vergleichsprogramm gilt es, den Aufwand und die damit verbundenen Belastungen aller Beteiligten mit dem zu erwartenden Ertrag in eine akzeptable Relation zu bringen. Deshalb soll in diesem Abschnitt erläutert werden, welche Gründe für den gewählten Zuschnitt von PISA sprechen, aber auch welche Grenzen das Design für die Interpretation von Befunden setzt. Der zweite Abschnitt wendet sich dann den Besonderheiten der aktuellen Studie und damit dem speziellen Erhebungsprogramm von PISA 2012 zu. Wichtige Aspekte der Durchführung der Studie in Deutschland (mit wenigen technischen Details) werden dann im dritten Abschnitt besprochen. Es folgen Abschnitte, die beim Lesen der Kapitel helfen sollen: Im vierten Teil werden einige Darstellungsformen (Kennzahlen, Abbildungen) erläutert, die in dem Band häufig verwendet werden, und der fünfte Teil weist auf einige Grenzen der Interpretierbarkeit bestimmter Typen von Befunden hin. In einer Art Abspannung werden schließlich die Organisationsstruktur und wichtige Akteure für die Durchführung von PISA 2012 in Deutschland vorgestellt.

1.1 Zielsetzung und Anlage des *Programme for International Student Assessment*

Die OECD errichtete mit PISA ein eigenes Indikatorensystem für Bildungsergebnisse. Mit PISA konnten die regelmäßigen OECD-Berichte (*Education at a Glance*, z. B. OECD, 2012a) Strukturdaten von Bildungssystemen mit Output-Daten verbinden und damit Zusammenhänge analysieren sowie gegebenenfalls Schlussfolgerungen über die Effektivität ermöglichen. PISA war zugleich ein Angebot an die OECD-Staaten (wie an weitere, sogenannte Partnerstaaten), dieses Indikatorenprogramm für ein kontinuierliches Bildungsmonitoring zu nutzen. Dieses Bildungsmonitoring sollte Daten bereitstellen, die für die Steuerung des Bildungssystems relevant sind, also Auskunft über relative Schwächen und Stärken, über Probleme, Herausforderungen und eventuell sogar Handlungsperspektiven geben.

Die Leitfrage bei PISA lautet entsprechend, wie gut es den Staaten gelingt, junge Menschen auf die Anforderungen einer Teilhabe an der Gesellschaft, auf das Erwachsenenleben und das weitere Lernen über die Lebensspanne vorzubereiten. Dementsprechend muss das Indikatorenprogramm Aussagen über das Niveau des Wissens und Könnens beziehungsweise wichtiger Fertigkeiten und Fähigkeiten zulassen, das in den einzelnen Staaten erreicht wird, wie auch über deren Verteilungen. Das Programm sollte aber auch Indikatoren für Disparitäten bereitstellen, die systematische Zusammenhänge zwischen Leistungsmaßen und zum Beispiel Merkmalen der sozialen Herkunft, der Zuwanderung oder dem Geschlecht berichten. Damit ungleiche Chancen auf Bildung nach Herkunftsmerkmalen festgestellt werden können, müssen weitere Merkmale von Schülerinnen und Schülern neben den Leistungen erhoben werden. Nicht zuletzt gewinnen Daten über lernrelevante Bedingungen in Unterricht und Schule an Bedeutung, wenn internationale Vergleiche über Schulsysteme und deren Ergebnisse informativ und anregend sein sollen.

1.1.1 Funktionen eines internationalen Bildungsvergleichs

Im Folgenden wird eine Reihe von Funktionen angesprochen, die das Indikatorensystem beziehungsweise das internationale Vergleichsprogramm erfüllen sollten und die letztlich dann auch umgesetzt wurden (OECD, 2013). Das Grundmodell von vergleichbaren Leistungserhebungen im dreijährlichen Abstand sollte mehrere Auswertungsperspektiven zulassen (Vergleiche zwischen Staaten, nationale und internationale Vergleiche über die Zeit, Vergleiche mit inhaltlich definierten Leistungsstandards). Die Vergleiche sollten auch dazu dienen, mehr oder weniger „gute“ Praxis zu identifizieren. In diesem Sinn können internationale Vergleiche für ein Benchmarking zwischen Staaten genutzt werden (Seidel & Prenzel, 2008).

Mit Blick auf die Aufwand-Ertrags-Relation stellt sich die Frage, an welcher Population die Erhebungen erfolgen sollen und welches inhaltliche Leistungsspektrum zu erfassen

sen ist. Im Sinne einer Bilanzierung des Einflusses der Schule bietet sich das Ende der allgemeinbildenden Pflichtschulzeit an. Da dieser Zeitpunkt zwischen vielen Staaten differiert, wurde für PISA entschieden, eine Altersgruppe zu untersuchen, die international normalerweise gerade noch die Schule besucht; damit fiel die Wahl auf die fünfzehnjährigen Schülerinnen und Schüler. Die Untersuchung einer Alterskohorte bringt den Vorteil, dass Wirkungen der Bildungssysteme international auf ein Lebensalter und damit eine fixe Zeitspanne bezogen werden. Es ist selbstverständlich, die Vergleiche an repräsentativen Stichproben durchzuführen (sowie entsprechende Kriterien für notwendige Testbeteiligung und Ausschöpfung etc. zu definieren).

Bei internationalen Vergleichen ist es hingegen schwierig zu entscheiden, welche Domänen untersucht werden sollen. Domänen oder Schulfächer müssten von ihren Lernangeboten international einigermaßen vergleichbar sein (schwierig z. B. bei Fächern wie Geschichte oder Religion) und Bildungsergebnisse sollten aussagekräftig mit Testverfahren erfasst und zuverlässig bewertet werden können (schwierig z. B. für Kunst). Außerdem sollten sie möglichst auch eine generalisierbare Einschätzung über die Qualität der Schulen zulassen und nicht zuletzt sollten sie auch für das Leben nach der Schule relevant sein. Aufgrund der Vorerfahrungen aus Vergleichsstudien der *International Association for the Evaluation of Educational Achievement* (IEA) wie zum Beispiel TIMSS – der *Third International Science and Mathematics Study* beziehungsweise *Trends in International Science and Mathematics Study* (z. B. Beaton et al., 1996) – zeichnete sich ab, dass die Curricula in den Bereichen Mathematik und Naturwissenschaften international beträchtliche Gemeinsamkeiten aufweisen und zuverlässige Testungen auch bei großen Stichproben realisiert werden können. In Hinblick auf die Machbarkeit schienen somit Mathematik und Naturwissenschaften geeignete Kandidaten zu sein. Zur Erweiterung des Spektrums wurde dann ein Leistungsbereich in Betracht gezogen, der in einer ganzen Reihe von Fächern eine wichtige Rolle spielt, nämlich die Lesekompetenz.

An dieser Stelle muss betont werden, dass die OECD mit PISA keine Duplizierung des Erhebungsansatzes der IEA für Mathematik und Naturwissenschaften anstrebte, sondern auch in diesen Bereichen ein originäres Konzept verfolgte. Während die IEA die Erhebungen in TIMSS und nachfolgenden Studien immer relativ eng auf die schulischen Curricula in Mathematik und Naturwissenschaften bezog, wollte die OECD überprüfen, inwieweit die Schülerinnen und Schüler in der Lage sind, mathematische und naturwissenschaftliche Kompetenzen auf Aufgabenstellungen und in Problemsituationen anzuwenden, die sich von traditionellen Schulbuchaufgaben abheben. Zwar sollte der schulische Kontext nicht ignoriert werden, aber die Testkonzeptionen sollten von Expertengruppen, ausgerichtet an aktueller mathematik- und naturwissenschaftsdidaktischer Forschung, ausgearbeitet werden. Dabei sollten relevante Inhaltsbereiche, Verfahrensweisen und Situationen geklärt und strukturiert werden, die für ein grundlegendes fachliches Verständnis, die Wissensanwendung in vielfältigen Problemsituationen und für das Weiterlernen im Verlauf der weiteren Bildungsbiographie essentiell sind. Die Strukturierung nach Inhalten, Prozessen und Situationen beziehungsweise Kontexten prägt bis

heute alle Testkonzeptionen für PISA. Im Vergleich zu den Studien der IEA wird die Anwendung des Wissens in vielfältigen Situationen betont. Deshalb werden bei PISA in gewisser Weise „situiertere“ Testaufgaben eingesetzt: Zu Beginn einer Aufgabe werden die Jugendlichen mit einer Situation und Problemstellung konfrontiert. Bezogen auf diesen Kontext folgen dann mehrere Testfragen (sogenannte Items). PISA unterscheidet damit systematisch zwischen *Aufgaben* (sogenannte Units) und *Items* (einzelne Testfragen als Teile einer Aufgabe).

Die OECD verwendet den Begriff der *Literacy*, der im englischsprachigen Raum seit einiger Zeit (und in der Mathematik wie in den Naturwissenschaften metaphorisch) für die Bezeichnung einer funktionalen Grundbildung gebraucht wird. In PISA wird funktional im Sinne von „anwendbar“ für die jetzige und spätere Teilhabe an einer Kultur und im Sinne von „anschlussfähig“ für nachfolgendes Lernen verstanden. Für den Anspruch, Bildungsergebnisse in einem bestimmten Alter messen zu wollen, ist diese Fokussierung entscheidend, denn es wird einerseits gefragt, inwieweit bis zum Testalter (schulische wie außerschulische) Lerngelegenheiten genutzt werden konnten, um in bestimmten Bereichen teilhabe- und handlungsfähig zu werden (im Englischen „literate“). Andererseits interessiert besonders das Wissen und Können in den untersuchten Domänen, welche benötigt werden, sich in diesen Bereichen lernend weiterentwickeln zu können. Auf den ersten Blick hat hier vor allem die Lesekompetenz einen sehr breiten Anwendungsbereich, denn in fast allen Domänen wird Wissen zum größten Teil über Texte (unterschiedlicher Sorten) bereitgestellt und angeeignet. Aber auch die Mathematik und die Naturwissenschaften gewinnen in unterschiedlichsten Fachkontexten und Bildungsprogrammen eine Anwendungsbreite, die weit über das Schulfach hinausweist. Mit dieser pragmatischen, zugleich anspruchsvollen Rahmenkonzeption kann PISA mit guten Gründen als „Bildungsstudie“ bezeichnet werden, denn mit diesem Ansatz können durchaus relevante Aspekte einer Grundbildung mit Blick auf Allgemeinbildung erfasst werden (vgl. Tenorth, 2004, 2005). Auf dieser Basis können ebenfalls Aussagen über die Qualität vorangegangener und gewissermaßen Prognosen für nachfolgende Bildungsprozesse getroffen werden.

Offensichtlich sind aber auch die Einschränkungen dieses Ansatzes: Es wird ein Ausschnitt aus dem Spektrum allgemeiner Bildung wie auch allgemeinbildender Schulfächer untersucht; die Tests fokussieren demnach kognitive Aspekte. Darüber hinausgehende Persönlichkeitsmerkmale werden allenfalls sehr selektiv und begrenzt durch Fragebögen erfasst. Entsprechende Einschränkungen, die notwendig zu empirischen Herangehensweisen gehören, sind zu beachten, um übergeneralisierende Aussagen über Bildung und die Bildungsqualität der Schule zu vermeiden. Für die Belastbarkeit der Aussagen wichtig ist vor allem die angemessene theoretische Modellierung der Erhebungen, die durch die Testkonzeption (theoretische Rahmenkonzeption, sogenanntes *Assessment Framework*) zu leisten ist.

Auch die Erhebung von Merkmalen, die in einem weiten Sinn als Bedingungen der Leistungsentwicklung in Bildungssystemen verstanden werden können, braucht eine

theoretische Modellierung als Grundlage. Ganz offensichtlich wäre ein Erhebungsprogramm überfordert, das alle möglichen Einflussgrößen auf die Testleistungen von Fünfzehnjährigen erfassen wollte – angefangen von Bedingungen in Elternhaus, Kindergarten und Schule bis zu Freundeskreisen und Medien, und das bezogen auf eine fünfzehnjährige Entwicklungs- und Lerngeschichte. Die erforderliche enge Fokussierung wird bei PISA wesentlich durch die Zielsetzung eines Bildungsmonitorings bestimmt, das insbesondere steuerungsrelevantes Wissen auf der Systemebene bereitstellen soll. Neben Informationen über die Leistungsfähigkeit in den getesteten Domänen werden hier zum Beispiel Daten wichtig, die Auskunft über weitere Merkmale der Schülerinnen und Schüler geben, wie etwa motivationale Orientierungen oder Einstellungen (z. B. Interesse, Selbstkonzept, Ängstlichkeit). Entsprechende Merkmale können als Bedingungs- oder Prozessvariablen mit Blick auf die Leistungsentwicklung verstanden werden, aber auch als ein Bildungsergebnis neben Leistung. Weiterhin können Kennwerte für Zusammenhänge zwischen Herkunftsmerkmalen und Testleistung als Indikator für Disparitäten genutzt werden, die auf Probleme der Chancengerechtigkeit oder auf nicht ausgeschöpfte Potenziale hinweisen. Für die Zielsetzung eines Benchmarkings braucht es außerdem Möglichkeiten, bestimmte strukturelle, institutionelle und organisatorische Merkmale vergleichen oder kontrollieren zu können. Schließlich liegt der Wunsch nahe, Unterschiede in den Testleistungen auf Bedingungen zurückführen zu können, indem zum Beispiel Anteile in der Leistungsvarianz statistisch durch mögliche Einflussfaktoren (und Varianzen darin) erklärt werden. Hier besteht die Hoffnung, Hinweise auf effektive Einflussfaktoren zum Beispiel schulischer Lernumgebungen zu erhalten. Die theoretische Strukturierung, die PISA von Anfang an für das übliche Fragebogenprogramm (Schülerfragenbogen, Schulfragebogen) zugrunde gelegt hat, orientiert sich erstens an einer basalen, unter anderem auch in der Schuleffektivitätsforschung verwendeten Differenzierung von Merkmalen, die als Input-, Prozess- oder Ergebnisvariablen eingeordnet werden. Zweitens werden unterschiedliche Aggregationsebenen differenziert, beginnend mit der Ebene *Individuum* und weiter aufgebaut über die Ebenen *Klassenzimmer*, *Schule* bis zur Ebene *Staat/System*. Diese beiden Aspekte werden als Dimensionen einer Matrix genutzt, die Ansatzpunkte für Erhebungen ordnet und zugleich auf bestimmte Nutzungs- und Interpretationsmöglichkeiten der Daten hinweist. Folgt man dem jüngsten *Questionnaire Framework* der OECD (2013) beziehungsweise den bei Klieme und Vieluf (2013) dargelegten Prinzipien der internationalen Fragebogenauswertung, dann wird diese Grundstruktur auch für künftige PISA-Erhebungen und -Auswertungen leitend sein.

1.1.2 Grundlegende Merkmale von PISA

Aus dieser Funktionsbeschreibung resultieren die Kernmerkmale von PISA: Die drei Domänen Lesen, Mathematik und Naturwissenschaften bilden gewissermaßen die Stichprobe aus den Inhalts- und Kompetenzbereichen, die weltweit in Schulen vermittelt werden und die als primär relevant für die weitere Bildungsbiographie, das Berufsleben wie auch die Partizipation an Kultur und Gesellschaft gelten. Was junge Menschen in diesen Bereichen wissen und können, wird an der Altersgruppe der Fünfzehnjährigen mit Testverfahren erfasst. Zusätzlich sollen Informationen über die Herkunft der Schülerinnen und Schüler, über ihre Einstellungen sowie Lern- und Freizeitaktivitäten erhoben und Merkmale von Unterricht und Schule erfasst werden. Die Erhebung findet alle drei Jahre an repräsentativen Stichproben statt und ist so angelegt, dass die Leistungen der jeweils untersuchten Kohorten im Zeitverlauf verglichen werden können. Bezugspunkt für Vergleiche der Staaten untereinander ist das mittlere Leistungsniveau der OECD-Staaten.

Weitere Charakteristika von PISA sind in den theoretischen Rahmenkonzeptionen (*Frameworks*) und *Technical Reports* für die einzelnen Studien beschrieben (z. B. OECD, 2013, 2012b). Von Beginn an wurde entschieden, die Domänen in den Erhebungsrounden unterschiedlich zu gewichten. Es gibt jeweils eine Hauptdomäne, für die mehr Testzeit eingeräumt wird und die über die Runden systematisch wechselt. In PISA 2012 ist wieder – wie bereits in PISA 2003 – Mathematik die Hauptdomäne. Auf diese Weise besteht alle neun Jahre die Gelegenheit, umfassende Tests durchzuführen und die Ergebnisse entsprechend detailliert im Trend zu betrachten. Zudem wird dabei auch die Möglichkeit genutzt, für die Hauptdomäne einer Erhebungsrunde die Testkonzeptionen zu überprüfen und gegebenenfalls zu modifizieren sowie den Aufgabenpool zu erneuern. Auch dies ist in PISA 2012 für die Mathematik geschehen.

Ein wesentliches Moment für die Trendfähigkeit ist ein Pool mit geheim gehaltenen Aufgaben (Anker- oder *Link-Units*), die in allen Erhebungsrounden eingesetzt werden und kriteriale (also inhaltsbezogene) Vergleiche der Leistungsfähigkeit über die Erhebungszeiträume erlauben. Aus methodisch-statistischer Sicht unterstützt die Modellierung und Skalierung der Tests mit Ansätzen der *Item Response Theory* (IRT) diese Auswertungsperspektive (z. B. Rost, 2004). Dieses Herangehen bringt weitere Vorteile mit sich: Aufgabenblöcke können über eine große Anzahl von Testheften rotiert und gemeinsam skaliert werden. Das bedeutet, dass jedes Testheft nur einen Teil des gesamten Aufgabenpools enthält und dass sich die Testhefte somit voneinander unterscheiden. Auf diese Weise wird nicht nur das Abschreiben im Test verhindert, sondern es kann insgesamt sehr viel mehr Aufgabenmaterial bei einer begrenzten Testzeit (von 120 Minuten) eingesetzt und zur Schätzung der Leistungsfähigkeit (auf Aggregations-ebenen wie Schule oder Staat) verwendet werden.

Außerdem nutzt PISA dieses Verfahren, um Kompetenzstufen zu differenzieren und zu beschreiben. Dabei werden Leistungskennwerte beziehungsweise -bereiche durch Aufgabenanforderungen illustriert, die typischerweise von den Schülerinnen und Schülern auf dieser Kompetenzstufe gemeistert werden. Kompetenzstufen veranschaulichen,

was Punktwerte auf der PISA-Skala bedeuten. Kompetenzstufen zeigen aber auch, an welchen Aufgaben die Jugendlichen der jeweiligen Stufe mit hoher Wahrscheinlichkeit scheitern würden. Auf diese Weise kann abgeschätzt werden, ob wichtige Bildungsziele eines Landes erreicht und grundlegende Voraussetzungen einer weiterführenden Bildungsstufe oder beruflichen Ausbildung entwickelt wurden. Wenn solche Voraussetzungen nicht entwickelt sind, wie etwa auf den untersten Kompetenzstufen, bestehen erhebliche Risiken für die weitere Biographie, für das Berufsleben und für die gesellschaftliche Teilhabe.

Vom Design her ist PISA international als Querschnittstudie angelegt: Es gibt nur einen Messzeitpunkt und damit Einschränkungen, Zusammenhänge zwischen Leistungsmaßen und anderen (theoretisch vielleicht als Einflussfaktor betrachteten) Merkmalen zu interpretieren. Sehr wichtig dabei ist, dass gefundene Zusammenhänge *nicht kausal* zu deuten sind; die in PISA auffindbaren Zusammenhänge sind zwar eine notwendige, nicht aber eine hinreichende Bedingung für kausale Aussagen. Das internationale Grunddesign sieht neben den verbindlichen Tests zwei weitere verpflichtende Erhebungsverfahren vor. Der sogenannte Schülerfragebogen enthält im wesentlichen Fragen zu Merkmalen der Person und ihrer Bildungsbiographie, zum Elternhaus (z. B. soziale Herkunft, Zuwanderungsgeschichte), zu Freizeitaktivitäten (etwa Mediennutzung) sowie zum Unterricht (vor allem in der jeweiligen Hauptdomäne) und zur Schule. Der sogenannte Schulfragebogen richtet sich an die Schulleitung und umfasst Merkmale der Schule (z. B. Größe, Ausstattung, Kollegium, Profil, Qualitätssicherung, Zusammenarbeit mit Eltern). Weitere Wahrnehmungen von Unterricht und Schule, etwa aus der Sicht von Lehrkräften und Eltern, gehören nicht zum standardmäßigen internationalen Erhebungsprogramm, können aber zum Teil als zusätzliche Option gewählt werden.

Das internationale Erhebungsdesign ist stark geprägt von der Zwecksetzung der OECD, durch PISA in erster Linie Wissen und Evidenzen für die Steuerung von Bildungssystemen bereitzustellen. Von dort aus betrachtet steht insbesondere die Aggregationsebene „Schule“ im Zentrum. Deshalb verlangt das internationale Stichprobendesign die Ziehung einer Zufallsstichprobe (in der Größenordnung von $n = 25$) von fünfzehnjährigen Schülerinnen und Schülern auf der Schulebene. Der „Normalfall“ mehrzügiger Schulen führt dazu, dass sich die Stichprobe der untersuchten Schülerinnen und Schüler auf mehrere Klassen und möglicherweise sogar mehrere Klassenstufen verteilt. Das internationale Grunddesign bietet damit keine Möglichkeit, systematisch Unterrichtsmerkmale auf der Klassenebene aus der Sicht mehrerer oder gar aller Schülerinnen und Schüler einer Klasse zu rekonstruieren. Die analytische Trennung einer Klassen- und einer Schulebene ist damit nicht möglich. Aufgrund dieser Stichprobenentscheidung ist das internationale Interesse an einer Befragung der Lehrkräfte – und insbesondere der Klassenlehrkräfte – verständlicherweise gering. Zugleich sind aber auf internationaler Ebene die Möglichkeiten von PISA sehr begrenzt, systematisches und objektivierbares Wissen über Unterrichtsmuster und die Qualität des Unterrichts zu erlangen (vgl. Prenzel & Lankes, 2013).

Für eine angemessene Interpretation von PISA-Ergebnissen ist es wichtig, neben den Stärken von vorneherein auch die Grenzen dieser Studie zu kennen. PISA weist besondere Stärken im *Assessment* auf, das auf drei Domänen konzentriert wird. Auf der Basis elaborierter Konzeptionen entwickelte Tests erfassen sowohl lehrplan- wie auch alltagsrelevantes Wissen und wichtige fachliche Voraussetzungen für weiteres Lernen in diesen Domänen. Die Tests sind so angelegt, dass weniger die simple Reproduktion, sondern vielmehr das flexible Anwenden des Wissens und das Lösen domänenspezifischer Probleme gefordert ist. Mit dem zugrunde liegenden psychometrischen Modell können bei PISA auf ökonomische Weise Facetten fachlicher Kompetenzen erhoben werden. Kompetenzstufen erlauben eine inhaltliche Interpretation von Testwerten und Prognosen der Anschlussfähigkeit für weiterführendes Lernen. Vor allem können mit dem Testansatz über die verschiedenen Erhebungsrunden nicht nur Vergleiche mit anderen Staaten getroffen, sondern Verbesserungen oder Verschlechterungen fachlicher Kompetenzen über die untersuchten Kohorten hinweg beschrieben werden. Diese Rückmeldefunktion ist in einem hohen Maße relevant für die bildungspolitische Steuerung. Die bei PISA verwendeten Schülerfragebögen gestatten es vor allem, Disparitäten der Leistung nach zahlreichen Merkmalen wie Geschlecht, soziale Herkunft, Migrationsstatus zu erfassen und gegebenenfalls auf Probleme ungleicher Bildungschancen aufmerksam zu machen. Die Fragebögen liefern auch Hinweise auf pädagogisch bedeutsame motivationale Orientierungen und Einstellungen. Die Informationen (insbesondere) über Unterrichtsbedingungen können als subjektive Einschätzungen prognostisch bedeutsam werden, gelangen aber an klare Grenzen, wenn Unterrichtsmuster beschrieben werden sollen, unter anderem auch aufgrund des Stichprobendesigns. Selbst wenn subjektive Unterrichtseinschätzungen maßgeblich zur Vorhersage von Leistungsunterschieden beitragen würden, erlaubt es das Erhebungsdesign nicht, Aussagen über die Unterrichtsqualität und mögliche Maßnahmen zu deren Verbesserung zu treffen.

Allerdings besteht die Möglichkeit, auf nationaler Ebene das internationale PISA-Design zu erweitern. Von dieser Möglichkeit hat Deutschland bisher immer Gebrauch gemacht und bei jeder, einschließlich dieser, Runde die Stichprobe der Fünfzehnjährigen durch zusätzliche Stichproben kompletter Klassen (pro Schule jeweils zwei der 9. Jahrgangsstufe) ergänzt. Die 9. Klasse ist in Deutschland diejenige Jahrgangsstufe, in der sich die meisten Schülerinnen und Schüler im Alter von fünfzehn Jahren befinden. Mit der Klassenstichprobe wurde auch der Einsatz von Lehrerfragebögen sinnvoll. Selbst das Querschnittsdesign wurde 2003 durch eine *Follow-up*-Erhebung erweitert (Prenzel et al., 2006), um solide Evidenz bezüglich relevanter Einflussfaktoren zu erhalten. Auch das Design von PISA 2012 wurde um ein *Follow-up* ergänzt, in dem alle Schulen, Klassen sowie Schülerinnen und Schüler aus der internationalen Erhebung im Sommer 2013 zu einer zweiten Erhebung gebeten wurden.

1.1.3 Missverständnisse und Kritik an PISA

PISA hat von Beginn an international hohe Aufmerksamkeit erfahren, wenn auch die öffentliche Resonanz in einigen Staaten sehr viel höher war als in anderen. Inzwischen zeichnet sich immer mehr ab, dass die öffentliche Diskussion auf nationaler Ebene dann heftiger wird, wenn PISA schlechte Botschaften überbringt und auf Probleme aufmerksam macht. Nicht immer ist das professionelle Bewusstsein bei Akteuren auf allen Ebenen so entwickelt, dass Informationen über Probleme als hilfreich betrachtet werden. Eine mögliche Reaktion besteht dann darin, die Studie zu ignorieren oder die Ergebnisse anzuzweifeln. Selbstverständlich muss sich PISA auch der Kritik stellen.

PISA erleichtert eine – auch kritische – Auseinandersetzung mit der Studie durch ein hohes Maß an Transparenz. Die internationalen Datensätze werden zusammen mit den offiziellen Berichten quasi zum Nachrechnen zugänglich gemacht. Die Auswertungsverfahren und die jeweiligen Schritte sind in den *Technical Reports* ausführlich beschrieben. Extern besetzte, hochkarätige Beiräte (z. B. *Technical Advisory Group*) prüfen das methodische Vorgehen und die Auswertungen bei jeder Runde gründlich. Ebenso engagieren sich viele Nationale Projektmanager bei der Überprüfung von Aufgaben und Prozeduren bis hin zu eigenen Re-Analysen der Daten. Ein internationales Qualitätsmonitoring wiederum stellt sicher, dass die Verfahren und Tests in allen Staaten regelgerecht durchgeführt werden. Allerdings sind die Testkonzeptionen und Auswertungsverfahren in PISA keineswegs trivial und allenfalls für einschlägig Sachverständige einigermaßen schnell nachzuvollziehen und zu beurteilen. Die Darstellung der Verfahren und Ergebnisse von PISA steht – auch in diesem Band – vor der Herausforderung, für eine breite Leserschaft verständlich zu sein und doch einige wichtige fachliche Begriffe und Details darzustellen. Den Dialog mit der Öffentlichkeit zu verbessern und Ergebnisse so aufzubereiten, dass Grenzen der Evidenz und Interpretierbarkeit ersichtlich werden, ist noch eine weitere Herausforderung für PISA (vgl. Bromme, Prenzel & Jäger, 2011).

Die in öffentlichen Diskursen (und insgesamt selten) geäußerte Kritik an PISA beruht häufig auf Missverständnissen und Unkenntnis. Das ist dann leicht der Fall, wenn die Information über PISA nur auf Zeitungsberichten, Rankings oder Internet-Blogs beruht und einschlägige wissenschaftliche Publikationen sowie die unmittelbare Berichterstattung aus dem Projekt vernachlässigt. Zu bestimmten Zeitpunkten – oft im Vorfeld einer neuen Veröffentlichung – tauchen in manchen Presseorganen Statements zu PISA auf, die deutliche und grundsätzliche Kritik äußern (zuletzt z. B. *TIMES Education Supplement*, 26. Juli 2013), zum Teil mit Verweis auf Stellungnahmen von Experten. Auch in der Geschichte der Berichterstattung über PISA in Deutschland gab es gelegentlich Stellungnahmen von einzelnen Wissenschaftlern, die fundamentale Kritik äußerten. Soweit die Stellungnahmen einigermaßen seriös waren, wurde von den nationalen Projektverantwortlichen Erwiderungen auf die Kritik veröffentlicht. Ein guter Maßstab für die Substanz einer Kritik ist, ob diese in einem einschlägigen internationalen wissenschaftlichen Journal mit Reviewverfahren publiziert wurde. Nimmt man dieses Kriterium, dann können die kritischen Stellungnahmen zu PISA, die nur auf Internetsei-

ten oder im Selbstverlag erschienen sind, mit gutem Gewissen ignoriert werden. In den einschlägigen internationalen *Journals* der Psychometrie oder der Lehr-Lern-Forschung jedenfalls findet man bis heute so gut wie keine grundlegende Kritik am methodischen Ansatz von PISA (mit Ausnahme etwa von Goldstein, 2004).

1.2 Die Besonderheiten von PISA 2012

Nach der Darstellung und Begründung der Kernmerkmale von PISA werden im Folgenden die Besonderheiten vorgestellt, durch die sich das internationale Erhebungsprogramm 2012 auszeichnete.

1.2.1 Teilnehmende Staaten

An PISA 2012 nahmen insgesamt 65 Staaten teil (OECD, 2013). Neben allen 34 OECD-Staaten beteiligten sich weitere 31 sogenannte Partnerstaaten an der aktuellen Erhebungsrunde. Im Vergleich zur ersten Erhebungsrunde im Jahr 2000 hat sich damit der Kreis der Teilnehmer von damals insgesamt 32 Staaten mehr als verdoppelt. Tabelle 1.1 gibt einen Überblick über alle teilnehmenden Staaten. Das Interesse von Staaten außerhalb der OECD nimmt weiter zu, sich mit PISA in einem Bildungsmonitoring einzuordnen, das die OECD-Ergebnisse als Bezugspunkt nimmt. Weltweit wurden in dem definierten Erhebungszeitraum (1. März bis 30. Juni 2012) ungefähr 500 000 Jugendliche getestet.

In PISA 2012 war Mathematik (nach 2003) wieder die Hauptdomäne. Die Nebendomäne Lesen fungierte in PISA 2000 und 2009 als Hauptdomäne; die Nebendomäne Naturwissenschaften bisher nur 2006 als Hauptdomäne.

Tabelle 1.1 An PISA 2012 teilnehmende Staaten

Albanien*	Jordanien*	Republik Serbien*
Argentinien*	Kanada	Rumänien*
Australien	Kasachstan*	Russische Föderation*
Belgien	Katar*	Schweden
Brasilien*	Kolumbien*	Schweiz
Bulgarien*	Korea	Shanghai (China)*
Chile	Kroatien*	Singapur*
Chinesisch Taipeh*	Lettland*	Slowakische Republik
Costa Rica*	Liechtenstein*	Slowenien
Dänemark	Litauen*	Spanien
Deutschland	Luxemburg	Thailand*
Estland	Macau (China)*	Tschechische Republik
Finnland	Malaysia*	Tunesien*
Frankreich	Mexiko	Türkei
Griechenland	Neuseeland	Ungarn
Hongkong (China)*	Niederlande	Uruguay*
Indonesien*	Norwegen	Vereinigte Arabische Emirate*
Irland	Österreich	Vereinigte Staaten
Island	Peru*	Vereinigtes Königreich
Israel	Polen	Vietnam*
Italien	Portugal	Zypern*
Japan	Republik Montenegro*	

Anmerkung: * OECD-Partnerstaaten

1.2.2 Testkonzeption und Testdesign

Die Testkonzeption, die im Jahr 2003 die Aufgabenentwicklung für die Mathematik als Hauptdomäne geleitet hatte, wurde für die Erhebung 2012 überprüft und an einigen Stellen modifiziert sowie ergänzt (näheres dazu in Kapitel 2 zur Mathematik in diesem Band). Die Grundstruktur der Testkonzeption wurde beibehalten, aber in verschiedener Hinsicht elaboriert. Die 2012 modifizierte Testkonzeption ist inhaltlich so angelegt, dass Vergleiche zum letzten Zeitpunkt, zu dem Mathematik Hauptdomäne war, und über die Erhebungsrunden möglich sind (OECD, 2013).

Tabelle 1.2 gibt einen Überblick über die Testkonzeptionen in PISA 2012. Jede Domäne ist in einer Spalte abgebildet und erhält zunächst in der obersten Zeile eine Definition der bereichsspezifischen Grundbildung. In den drei darunterliegenden Zeilen wird für jede Domäne zusammengefasst, welches die drei Wissensformen sind, über welche die Schüler verfügen sollten:

- die *Inhalte* beziehungsweise die Konzepte („Wissen, dass“), deren Kenntnis zum Erreichen der Grundbildung in den drei Domänen vorausgesetzt wird;

Tabelle 1.2 Organisation der theoretischen Rahmenkonzeptionen der drei kognitiven Domänen in PISA 2012

	Mathematik	Lesen	Naturwissenschaften
Definition	Die Fähigkeit, Mathematik in zahlreichen Kontexten anzuwenden, zu interpretieren und Formeln zu verwenden. Dazu gehört mathematisches Schlussfolgern ebenso wie die Anwendung mathematischer Konzepte, Vorgehensweisen, Fakten und Werkzeuge, um Phänomene zu beschreiben, zu erklären und vorherzusagen. Mathematische Grundbildung hilft Personen, die Rolle zu erkennen und zu verstehen, die Mathematik in der Welt spielt, fundierte mathematische Urteile abzugeben und Mathematik in einer Weise zu verwenden, die den Anforderungen des Lebens dieser Person als konstruktivem, engagiertem und reflektiertem Bürger entspricht.	Die Fähigkeit einer Person, geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, eigenes Wissen und eigene Potentiale zu entwickeln und um aktiv an der Gesellschaft teilzunehmen.	Die Fähigkeit, die charakteristischen Eigenschaften sowie die Bedeutung der Naturwissenschaften in unserer heutigen Welt zu verstehen, naturwissenschaftliches Wissen anzuwenden, um Fragestellungen zu erkennen, naturwissenschaftliche Phänomene zu beschreiben und aus Belegen Schlussfolgerungen zu ziehen, sowie die Bereitschaft, sich reflektierend mit naturwissenschaftlichen Ideen und Themen auseinanderzusetzen.
Inhalte	Vier Inhaltsbereiche: – Quantität – Raum und Form – Veränderung und Beziehungen – Unsicherheit und Daten	Textformate: – kontinuierliche Texte, z.B. Erzählungen, Beschreibungen, Argumentationen – nichtkontinuierliche Texte, z.B. Diagramme, Tabellen, Listen	Bereiche <i>naturwissenschaftlichen Wissens</i> : – Physikalische Systeme – Lebende Systeme – Erd- und Weltraumsysteme – Technologische Systeme Bereiche des (Meta-)Wissens <i>über die Naturwissenschaften</i> – naturwissenschaftliches Forschen – naturwissenschaftliche Erklärungen
Kontexte bzw. Situationen	Die Rahmenkonzeption unterscheidet folgende Kontexte: – persönliche – ausbildungs-/berufsbezogene – gesellschaftsbezogene – wissenschaftliche	Die Rahmenkonzeption unterscheidet folgende Lesesituationen bzw. Textarten: – Private Situationen (z.B. persönliche Briefe) – Öffentliche Situationen (z.B. offizielle Dokumente) – Bildungsbezogene Situationen (z.B. Lehrbuchtexte) – Berufsbezogene Situationen (z.B. Stellenausschreibungen)	Die Rahmenkonzeption unterscheidet fünf Kontexte, die jeweils einen persönlichen, sozialen oder globalen Bezug haben können: – Gesundheit – Natürliche Ressourcen – Umwelt – Risiken/Gefahren – Grenzen von Naturwissenschaft und Technik
Prozesse	– Formeln verwenden – Anwenden – Interpretieren	Leseaufgaben: – Informationen suchen und extrahieren – Textbezogen kombinieren und interpretieren – Reflektieren und bewerten	– das Erkennen naturwissenschaftlicher Fragestellungen, – das Beschreiben, Erklären und Vorhersagen naturwissenschaftlicher Phänomene sowie – das Nutzen naturwissenschaftlicher Evidenz, um zu Entscheidungen zu gelangen.

- die *Prozesse* und Vorgehensweisen („Wissen, wie“), welche die Jugendlichen bei der Lösung der PISA-Aufgaben zeigen sollten;
- die Situationen und *Kontexte*, in denen die Schülerinnen und Schüler ihr Wissen für eine angemessene Lösung anwenden können sollten.

Für PISA 2012 wurden 31 Mathematikaufgaben mit insgesamt 74 Items neu entwickelt. Daneben kamen 25 geheim gehaltene Aufgaben mit 36 Items aus früheren Erhebungen zum Einsatz, um die Trendberechnungen abzusichern. Die Aufgabenentwicklung erfolgte an fünf international einschlägig ausgewiesenen Instituten (*Australian Council for Educational Research* – ACER, Australien; *Leibniz Institut für die Pädagogik der Naturwissenschaften und Mathematik* – IPN, Deutschland; *Le service d'analyse des Systèmes et des Pratiques d'enseignement* – aSPe, Belgien; *Institute for Learning Science*, Norwegen; *National Institute for Educational Research* – NIER, Japan). Alle neu entwickelten Testaufgaben wurden in einem mehrstufigen Verfahren geprüft und bei Bedarf verbessert. Nach ersten Erprobungen der Aufgaben an Schülergruppen folgten Aufgabenbeurteilungen durch die internationale Expertengruppe für Mathematik sowie durch alle nationalen Projektmanager. Die Einschätzung wurde unter anderem nach Kriterien wie fachlicher Richtigkeit, Bezüge zum Lehrplan, Relevanz für die Lebenswelt Fünfzehnjähriger, möglicher kultureller oder geschlechtsbezogener Benachteiligung oder Aufgabenschwierigkeit vorgenommen. Alle neuen Aufgaben wurden im Feldtest, der 2011 in allen Staaten durchgeführt wurde, schließlich systematisch erprobt.

Die Erhebungen in den Nebendomänen Lesen und Naturwissenschaften erfolgte in PISA 2012 auf der Basis der früheren Testkonzeptionen, die anlässlich ihrer damaligen Eigenschaft als Hauptdomäne entwickelt worden waren. Entsprechend war für den Lesetest der Bezugspunkt PISA 2009 (OECD, 2009); für die Naturwissenschaften PISA 2006 (OECD, 2006). In diesen beiden Bereichen kamen keine neu entwickelten Aufgaben zum Einsatz. Für die Tests im Bereich Lesen wurden in PISA 2012 insgesamt 13 geheime Aufgaben mit 44 Items und im Bereich Naturwissenschaften 18 Aufgaben mit 53 Items verwendet.

Bei PISA 2012 wurde für die Mathematik Aufgabenmaterial für eine Testzeit von insgesamt 270 Minuten eingesetzt und auf die (rotierten) Testhefte verteilt; jede Schülerin und jeder Schüler erhielt einen Teil dieses Aufgabenmaterials. Für die beiden Nebendomänen Lesen und Naturwissenschaften wurde jeweils Testmaterial im Umfang von 90 Minuten verwendet, das ebenfalls auf verschiedene Testhefte verteilt wurde.

Der in PISA gewählte Ansatz zur Zusammenstellung der Aufgaben in Testhefte wird als *Multi-Matrix-Design* bezeichnet. Dabei wurden in PISA 2012 systematisch 13 verschiedene Testhefte mit variierten Aufgabengruppen (sogenannte *Cluster*) zusammengestellt. Die Testhefte werden den teilnehmenden Schülerinnen und Schülern nach einem Zufallsverfahren zugewiesen. Die Punktwerte, die mit Hilfe des Multi-Matrix-Designs anhand der von den Schülern gelösten Aufgaben erhoben werden, können mit speziellen Auswertungsverfahren (der *Item Response Theory*; vgl. Kapitel 10) ausgewertet, direkt miteinander verglichen und interpretiert werden.

Die Aufgabengruppen werden je Domäne in PISA so zusammengestellt, dass ein Cluster 30 Minuten Bearbeitungszeit entspricht. Jede Aufgabe kommt in mehreren Clustern und dort an unterschiedlichen Positionen vor. Ein Cluster besteht jeweils nur aus geheim gehaltenen, bereits eingesetzten Aufgaben (*Link-Aufgaben*) oder ausschließlich aus neu entwickelten Mathematikaufgaben. Aus dem Bereich der Mathematik wurden sieben Cluster gebildet, für das Lesen und die Naturwissenschaften jeweils drei. Jedes der 13 unterschiedlichen Testhefte enthielt vier Cluster, so dass eine Bearbeitungszeit von 120 Minuten (zuzüglich einer Pause nach 60 Minuten) anberaumt war. Im Anschluss an die Testung wurde den Schülerinnen und Schülern ein Fragebogen vorgelegt. Tabelle 1.3 veranschaulicht das Testheftdesign sowie den Ablauf der Testsitzung anhand der Bearbeitungszeiten.

Tabelle 1.3 Testheftdesign von PISA 2012

Testheft													
Dauer	1	2	3	4	5	6	7	8	9	10	11	12	13
30	M5	N3	L3	M6	M7	M1	M2	N2	L2	M3	M4	N1	L1
30	N3	L3	M6	M7	N1	M2	N2	L2	M3	M4	M5	L1	M1
30	M6	M7	N1	L1	M1	L1	M3	M4	M5	N3	L3	M2	N2
30	N2	L2	M3	M4	M5	M6	M7	N1	L1	M1	M2	N3	L3
45	Fragebogen												

Anmerkung: M = Mathematik, N = Naturwissenschaften, L = Lesen

Der im Anschluss an die Testung eingesetzte Schülerfragebogen wurde entsprechend dem bereits erwähnten Framework (OECD, 2013) entwickelt. In den 45 Minuten Bearbeitungszeit wurde von allen Schülerinnen und Schülern ein Kernbestand an Daten (die sogenannten Hintergrundmerkmale wie zum Beispiel soziale Herkunft, Migrationsstatus) erhoben. Weitere Teile des Fragebogens (z. B. Wahrnehmung von Unterricht und Schule) wurden in drei rotierten Varianten dargeboten, um mehr Informationen (für aggregierte Betrachtungen) in der Befragungszeit gewinnen zu können.

Zum Pflichtprogramm bei PISA 2012 zählte wiederum der an die Schulleitung adressierte Schulfragebogen. Ein Fragebogen für die Eltern der getesteten Jugendlichen war in 2012 wieder als internationale Option angeboten (und von Deutschland wahrgenommen) worden.

1.2.3 Zusatzprogramm und weitere internationale Optionen

Mit Blick auf die ab PISA 2015 geplante Umstellung des gesamten Erhebungsverfahrens von Papier- und Bleistift-Aufgaben auf computerbasierte Tests hat die OECD für PISA 2012 bereits entsprechende Testkomponenten angeboten. Auf diese Weise sollte Gelegenheit für Erprobungen und für das Sammeln von Erfahrungen gegeben werden.

Eine computerbasierte Testkomponente war obligatorisch und entsprechend auch in Deutschland durchgeführt worden. Diese Komponente war entwickelt worden, um individuelle Fähigkeiten zum Lösen (genereller, realistischer, interaktiv zu bearbeitender) Probleme zu erfassen (OECD, 2013). Die OECD wollte damit an frühere Ansätze anschließen (z. B. PISA 2003), bei denen Problemlösefähigkeiten im Sinne einer Querschnittskompetenz (*cross-curricular competence*) untersucht wurden. Geplant ist die Weiterentwicklung entsprechender Erhebungsansätze, um zukünftig computerbasiert Kompetenzen für das kollaborative Lösen von Problemen testen zu können. Die computerbasierte Erhebung individueller Problemlösefähigkeiten wurde in PISA 2012 an einer Teilstichprobe der Fünfzehnjährigen in einer zusätzlichen Testsitzung durchgeführt. Die internationalen Datenanalysen und Auswertungen zu diesem computerbasierten Problemlöse-Test sind derzeit noch nicht in Arbeit. Die OECD plant, die Daten und Berichte 2014 zugänglich zu machen. Aus diesem Grund sparen wir in diesem nationalen Berichtsband diese Domäne aus.

Die von der OECD angebotenen Optionen, bereits in PISA 2012 zusätzlich zu den herkömmlichen Papier- und Bleistift-Erhebungen computerbasierte Erhebungen zur Lese- und Mathematikkompetenz durchzuführen, waren von der KMK in Deutschland nicht gewählt worden. Das Zentrum für internationale Vergleichsstudien (ZIB), an dem auch die PISA-Erhebungen in Deutschland koordiniert werden, hat jedoch in sein Forschungsprogramm Studien aufgenommen, in denen Möglichkeiten und Grenzen sowie die Trendfähigkeit von computerbasierten Erhebungen in den PISA-Domänen systematisch untersucht werden.

Von deutscher Seite wurde auch auf die Wahrnehmung der von der OECD angebotenen Option eines zusätzlichen Tests zur „Financial Literacy“ verzichtet.

1.3 Die Durchführung von PISA in Deutschland

Die Arbeitsschritte, die im Verlauf einer PISA-Erhebungsrunde erledigt werden müssen, sind international verbindlich festgelegt und terminiert. Die Regelungen dienen insbesondere dazu, methodische Standards zu sichern und die internationale Vergleichbarkeit zu gewährleisten. In den folgenden Abschnitten werden Verfahrensschritte beschrieben, die für die Vorbereitung und Durchführung von PISA in Deutschland entscheidend waren.

1.3.1 Population und Stichprobe: Wer nimmt an PISA teil?

PISA soll den Leistungsstand von fünfzehnjährigen Schülerinnen und Schülern in den teilnehmenden Staaten *repräsentativ* abbilden. Deshalb muss eine repräsentative Stichprobe von Schulen und Jugendlichen gezogen werden. Zudem muss gewährleistet werden, dass möglichst alle Schulen und alle Schülerinnen und Schüler, die gezogen wurden, tatsächlich an der Erhebung teilnehmen. Im Folgenden wird skizziert, welche Maßnahmen zur Sicherung der Repräsentativität und Vergleichbarkeit der Ergebnisse aus PISA ergriffen werden. Genauere Angaben zu den Prozeduren findet man in den Technical Reports zu PISA (z. B. OECD, 2012b).

Ziehung der repräsentativen Stichprobe

Die Ziehung repräsentativer Stichproben ist nötig, um von den befragten beziehungsweise getesteten Schülerinnen und Schülern auf eine bestimmte Grundgesamtheit schließen zu können. Im Rahmen von PISA wird von einer repräsentativen Stichprobe fünfzehnjähriger Jugendlicher auf die Population der Fünfzehnjährigen, die in Deutschland zur Schule gehen, verallgemeinert. Die PISA-Stichproben werden so gewählt, dass Rückschlüsse auf die Ebene der teilnehmenden Staaten gezogen werden können. Entsprechend zugeschnitten ist die PISA-Stichprobe auch in Deutschland; sie erlaubt keine Vergleiche zwischen den Ländern innerhalb Deutschlands. Vergleiche zwischen den Bundesländern sind den standardbezogenen Leistungsvergleichen vorbehalten, die vom Institut zur Qualitätsentwicklung im Bildungswesen (IQB) in regelmäßigen Abständen erhoben werden. Der letzte bundesweite Leistungsvergleich auf der Ebene der Sekundarstufe wurde 2013 publiziert (Pant et al., 2013). Weltweit haben an PISA 2012 etwa 500 000 Jugendliche teilgenommen, in Deutschland waren es 5 001 Schülerinnen und Schüler an 230 Schulen. Die Stichprobe in Deutschland wurde im Sinne einer nationalen Option erweitert (für zusätzliche nationale Forschungsfragen und Auswertungen). Erstens wurden an jeder für PISA ausgewählten Schule zwei komplette Klassen der Jahrgangsstufe 9 gezogen. Diese Klassenstichprobe nahm am zweiten Testtag an den standardbezogenen Tests des IQB teil ($n = 9\,998$). Zweitens wurde für die Beantwortung von Forschungsfragen die Stichprobe der Sonder- und Förderschulen vergrößert. Die Daten aus diesen Erweiterungen der Stichproben gingen jedoch nicht in die internationalen Vergleiche ein und sind auch nicht Teil dieses Berichtsbandes.

Tabelle 1.4 Zusammensetzung der Stichprobe in den an PISA 2012 teilnehmenden Schulen

Allgemeinbildende Schulen (ohne Sonder- und Förderschulen)	Sonder- und Förderschulen	Berufliche Schulen
211	12	7
<i>Internationaler Test PISA 2012</i>		
Max. 25 Fünfzehnjährige + zwei vollständige 9. Klassen	Max. 25 Fünfzehnjährige	Max. 25 Fünfzehnjährige
<i>Ländervergleich Bildungsstandards des IQB</i>		
zwei vollständige 9. Klassen		

Die Repräsentativität der Stichprobe wird in PISA durch ein genau vorgeschriebenes mehrstufiges Ziehungsverfahren gesichert. Technisch ausgedrückt wird eine mehrfach geschichtete (stratifizierte) Wahrscheinlichkeitsstichprobe gezogen. In diesem Verfahren müssen einige Besonderheiten der Situation in Deutschland (wie z. B. die Gliederung in Bundesländer oder Schularten) berücksichtigt werden. Mit Hilfe der statistischen Landesämter wird zunächst eine vollständige Liste aller Schulen in Deutschland erstellt, in denen sich potentiell Schülerinnen und Schüler im Alter von fünfzehn Jahren finden lassen. Das Alter „fünfzehn Jahre“ ist durch den Geburtszeitraum vom 1. Januar 1996 bis zum 31. Dezember 1996 definiert. Diese Liste umfasst die Grundgesamtheit, aus der die Schulen zufallsbasiert gezogen werden. Die Grundgesamtheit wird unterteilt in die 16 Länder und anhand der aktuellen Anzahl fünfzehnjähriger Schülerinnen und Schüler sowie der jeweils vorhandenen allgemeinbildenden Schularten wird bestimmt, wie viele Schulen beziehungsweise Fünfzehnjährige je Bundesland gezogen werden müssen. Diese Unterteilung wird Stratifizierung genannt, jedes Bundesland ist dabei ein Stratum (Schicht). Neben den 16 Ländern erhält die Grundgesamtheit zwei weitere Strata, die beruflichen Schulen sowie Sonder- und Förderschulen. Damit wird sichergestellt, dass einerseits alle Länder sowie auch alle Schularten in dem Anteil gezogen werden, der den Anforderungen der Repräsentativität genügt. Für jede Schule werden zwei Ersatzschulen bestimmt. So kann im Falle einer unvermeidbaren Nichtteilnahme einer Schule (z. B. weil die Schule zwischenzeitlich geschlossen wurde) dennoch das Zufallsprinzip der Ziehung aufrechterhalten werden.

Innerhalb der ausgewählten Schulen wird dann wiederum eine Zufallsstichprobe von 25 fünfzehnjährigen Schülerinnen und Schülern gezogen. Im Rahmen der Stichprobenerweiterung auf Jahrgangsebene wurden pro Schule zwei 9. Klassen per Zufall gezogen.

Im ersten Auswahlschritt wurden in Deutschland 247 Schulen gezogen, von denen schlussendlich 230 an PISA 2012 teilnahmen (vgl. auch Kapitel 10, Technische Grundlagen). Für jede Schule wurden zwei Ersatzschulen gezogen, die in die Stichprobe nachrückten, sofern eine zuerst gezogene Schule ausfiel. Bei den 17 ursprünglich gezogenen

Tabelle 1.5 Zielpopulation und Ausschöpfungsgrad bei PISA 2012 in Deutschland

15-Jährige in Schulausbildung ^a (PISA-Population)	Ausschlüsse auf Schulebene		Ausschlüsse in den Schulen		Erreichte Population	
	absolut ^b	in Prozent der Zielpopulation	absolut	in Prozent der Zielpopulation	absolut	in Prozent der Zielpopulation
798 136	10934	1.37	1357	0.17	756 907	94.8%

Anmerkung: ^a Quelle: Statistisches Bundesamt, Fachserie 11. ^b Quelle: Sampling Frame.

Schulen, die nicht an PISA 2012 teilnahmen, handelt es sich um solche, bei denen entweder gemäß der internationalen PISA-Standards keine Ersetzung vorgesehen war oder bei denen keine Ersatzschulen mehr zur Verfügung standen. Hierzu gehören 11 Fälle von Berufsschulen, an denen es keine fünfzehnjährigen Schülerinnen und Schüler gab; in zwei Fällen handelte es sich um *private* Schulen, die nicht zur Teilnahme an PISA verpflichtet sind und selbst bereits zweite Ersatzschule waren. Zwei weitere private Schulen waren zwischen der Erstellung der vollständigen Schulliste und der Stichprobenziehung geschlossen worden. Eine private Schule sagte ihre Teilnahme so kurzfristig vor Testbeginn ab, dass keine Ersetzung mehr möglich war, und eine dänische Schule, an der lediglich vier Fünfzehnjährige zu erwarten gewesen wären, wurde nach Absprache mit dem internationalen Konsortium ohne Ersatz ausgeschlossen, da der zeitliche und organisatorische Aufwand für die Übersetzung sämtlicher deutscher Testmaterialien ins Dänische – verglichen mit der geringen Schüleranzahl – zu groß gewesen wäre. Damit verblieben letztlich 230 Schulen, an denen die PISA-Studie durchgeführt wurde. Die gewichtete Teilnahmequote auf Schulebene beträgt somit 98.3 Prozent. Diese Zahl bezieht sich auf die ursprünglich gezogenen Schulen und zeigt auf, wie viele dieser Schulen nach eventuellen Ersetzungen tatsächlich an PISA 2012 teilgenommen haben. Bei den Schülerinnen und Schülern beläuft sich die gewichtete Teilnahmequote auf 93.2 Prozent. Beide Werte erfüllen die von der OECD gesetzten Mindeststandards.

In beruflichen Schulen und in Förderschulen wurde jeweils eine Vollerhebung aller fünfzehnjährigen Schülerinnen und Schüler durchgeführt. In den allgemeinbildenden Schulen, die von mehr als 25 Fünfzehnjährigen besucht wurden, wurde eine Zufallsstichprobe gezogen.

Zum Zeitpunkt der Stichprobenziehung und auch der Datenerhebung in PISA 2012 befand sich die Sekundarschullandschaft in mehreren Bundesländern Deutschlands im Umbruch. Mit Ausnahme der Gymnasien wurden verschiedentlich Sekundarschulen in neue oder teilweise neue Schularten überführt, womit auch eine neue Bezeichnung erfolgte. Beispielsweise waren in Bayern zum Zeitpunkt der Vorbereitung der Stichprobenziehung (Ende 2011) einige Schulen noch als Hauptschulen gelistet, die zum Zeitpunkt der Datenerhebung im Frühjahr 2012 dann bereits nominell als Mittelschulen galten.

In vier Bundesländern waren in PISA 2012 noch nominelle Hauptschulen in der Stichprobe (Baden-Württemberg, Hessen, Niedersachsen und Nordrhein-Westfalen) vertreten, während es in den übrigen Ländern entweder bereits keine sogenannten Hauptschulen mehr gab oder aber deren Anteil zu gering war, um für die deutschlandweite PISA-Stichprobe in Frage zu kommen. In Bayern, wo die ehemaligen Hauptschulen zum Zeitpunkt der PISA-Erhebung bereits zu Mittelschulen entwickelt worden waren, gehörten diese zum Zeitpunkt der Vorbereitung für die Stichprobenziehung (Listung der Schulen, vgl. Kapitel 10) zur Gruppe der Hauptschulen und sind daher in diesem Bericht als solche ausgewiesen. Neben den Gymnasien existieren derzeit also zahlreiche Schulen mit mehreren Bildungsgängen, die entweder in kooperativer oder integrierter Form geführt werden.

Innerhalb der Sekundarschularten werden zum Teil alle drei Bildungsgänge (Hauptschule, Realschule, Gymnasium) angeboten (Integrierte Gesamtschulen). Zum Teil werden zwei Bildungsgänge (Hauptschule, Realschule) angeboten; diese Schulen werden in diesem Band zusammenfassend als „Schulen mit mehreren Bildungsgängen“ (MBG) bezeichnet. In mehreren Bundesländern und Schularten ist für die Stichprobe der Fünfzehnjährigen nicht identifizierbar, ob und welchem Bildungsgang der einzelne Schüler und die einzelne Schülerin zugeordnet ist. Bezüglich der Bildungsgänge innerhalb der nicht gymnasialen, allgemeinbildenden Schulen ist auf der Basis der für Deutschland (nicht aber für einzelne Bundesländer) repräsentativen Stichprobe kein differenzierteres Bild möglich, ohne grobe Fehler in Kauf zu nehmen.

Deshalb wird in den Auswertungen in diesem Bericht auf die klar unterscheidbaren Schularten „Hauptschule“, „Realschule“, „Schule mit mehreren Bildungsgängen“, „Integrierte Gesamtschule“, „Gymnasium“, „Sonder-/Förderschule“ sowie „berufliche Schule“ rekuriert. Da in den Sonder-/Förderschulen wie in den beruflichen Schulen nur bestimmte Teilstichproben (z. B. Kriterium „selbstständige Testbeteiligung möglich“, „bereits mit 15 Jahren in beruflicher Schule“) einbezogen werden konnten, kann von den Stichproben nicht auf die Leistungen der Grundgesamtheit der Förder- oder beruflichen Schulen geschlossen werden. Die Stichproben aus diesen Schularten werden benötigt, um das Gesamtbild der Leistungen der schulpflichtigen Fünfzehnjährigen in Deutschland abzusichern, da in PISA Aussagen über die Grundgesamtheit der Fünfzehnjährigen gemacht werden sollen, die in verschiedenen Bildungssystemen weltweit eine Schule besuchen. Daher sind die Förder- und beruflichen Schulen Teil der deutschen Gesamtstichprobe und gehen in die Analysen für die *gesamte* Stichprobe in Deutschland ein, nicht jedoch in die nach Schularten getrennten Berechnungen. Der Anteil der Förder- und beruflichen Schulen an der Gesamtheit der Schulen, die in Deutschland von Fünfzehnjährigen besucht werden, ist zu gering, als dass repräsentative Aussagen über diese beiden Schularten möglich werden. Sie werden aber benötigt, um die Sekundarschullandschaft in Deutschland als Ganzes repräsentativ in Bezug auf die von Fünfzehnjährigen besuchten Schularten abzubilden. Tabelle 1.6 zeigt die Verteilung der getesteten Schülerinnen und Schüler auf die sieben Schularten, die in Deutschland in der Gesamtstichprobe unterschieden werden.

Tabelle 1.6 Untersuchungsbeteiligung der Fünfzehnjährigen nach Schulart und Klassenstufe

Klassenstufe	Haupt- schule	Schule mit mehreren Bildungs- gängen	Inte- grierte Gesamt- schule	Real- schule	Gym- nasium	Berufs- schule	Sonder- und Förder- schule	Summe
insgesamt gezogen	611	826	524	1286	1896	116	165	5424
7	17	6	1	4	0		4	
8	131	104	40	125	54		43	
9	327	433	264	641	855		86	
10 und 11	75	200	167	426	893		6	
keine Zuordnung						99		
insgesamt getestet	550	743	472	1196	1802	99	139	5001
Ausschöpfung in %	90.0	90.0	90.0	93.0	95.0	85.3	84.2	92.2

Anmerkung: Die Tabelle berichtet die realisierten Stichproben ungewichtet. Die gewichtete Ausschöpfung beträgt 93.2 Prozent.

Da Fünfzehnjährige innerhalb von großen Schulen mit einer vergleichsweise hohen Zahl von Gleichaltrigen eine geringere Wahrscheinlichkeit haben, in die Zufallsstichprobe zu geraten, als Jugendliche in kleineren Schulen, muss vor der Auswertung der Daten ein Ausgleich dieser unterschiedlichen Ziehungswahrscheinlichkeiten vorgenommen werden (vgl. auch Kapitel 10). Dieses Verfahren heißt Gewichtung. Analog zur Ziehung der Stichprobe erfolgt die Berechnung der Gewichte in mehreren Stufen. Zunächst erhält jede Schule ein Gewicht, das sie umgekehrt proportional zu ihrer Ziehungswahrscheinlichkeit höher oder niedriger gewichtet. Gibt es in einem Land etwa sehr viele Gymnasien, so hat jedes einzelne Gymnasium nur eine vergleichsweise geringe Chance, zufällig in die Stichprobe gezogen zu werden. Eine Mittelschule im selben Land kann etwa eine wesentlich höhere Ziehungswahrscheinlichkeit besitzen, wenn es nur wenige Mittelschulen gibt. Diese unterschiedliche Wahrscheinlichkeit wird durch die Gewichtung ausgeglichen (vgl. auch Kapitel 10, Technische Grundlagen). Die zweite Stufe der Gewichtung befindet sich innerhalb der Schulen. In einer gezogenen Schule werden aus allen Schülerinnen und Schülern, die die Kriterien für eine Teilnahme an PISA erfüllen, zufällig 25 Jugendliche gezogen. Auch hier gilt: In großen Schulen hat eine einzelne Schülerin oder ein einzelner Schüler eine geringere Wahrscheinlichkeit für die Testteilnahme gezogen zu werden als an kleinen Schulen. Auch diese unterschiedlichen Wahrscheinlichkeiten werden mit Hilfe eines Schülersgewichts ausgeglichen. In Verbindung mit dem Schulgewicht und der Berücksichtigung von Schülerinnen und Schülern, die in der PISA-Stichprobe gelandet sind, jedoch am Testtag nicht anwesend waren, wird ein sogenanntes Schülersgesamtgewicht errechnet. Dieses Schülersgesamtgewicht ermöglicht

es schließlich, dass jeder an PISA teilnehmende Staat repräsentativ abgebildet wird und zudem unabhängig von seiner Größe, Bevölkerungsdichte und seinem Schulsystem in gleichem Maße in die Auswertungen eingeht.

Ausschlussgründe und Ausschöpfung

In Deutschland sind alle staatlichen Schulen, die für die Stichprobe gezogen wurden, zur Teilnahme an PISA verpflichtet. Bei einer Vergleichsstudie wie PISA muss darüber hinaus sichergestellt werden, dass alle ausgewählten Schülerinnen und Schüler auch in der Lage sind, die Testung zu bewältigen. Diese sogenannte *Testfähigkeit* wird im Vorfeld der PISA-Untersuchung bei der Leistung der Schülerinnen und Schüler durch den Schulkoordinator beziehungsweise die Schulkoordinatorin individuell eingeschätzt und notiert. Daher kann es sowohl auf der Schul- als auch auf der Schülerebene zu Ausschlüssen von PISA kommen. Aufgrund der notwendigen Fairness der Testungen gibt es in PISA klare Regeln, wann Schulen beziehungsweise Schülerinnen und Schüler die Voraussetzungen der Testfähigkeit nicht erfüllen: Wer körperlich, kognitiv oder emotional nicht in der Lage ist, die PISA-Aufgaben selbstständig zu bearbeiten, oder wer weniger als ein Jahr in der Testsprache unterrichtet wurde, nimmt nicht am Test teil. Grundsätzlich als testfähig gelten Jugendliche, wenn sie mindestens ein Jahr in Deutschland gelebt haben. Ganze Schulen können dann ausgeschlossen werden, wenn für die Mehrzahl ihrer Schülerinnen und Schüler eines oder mehrere der genannten Kriterien zutreffen. Auf der Schulebene werden in Deutschland daher nur Schulen für Schülerinnen und Schüler mit geistiger, körperlicher oder mehrfacher Behinderung sowie für Kranke ausgeschlossen. Auf der Schülerebene muss insbesondere bei Jugendlichen mit sonderpädagogischem Förderbedarf oder mit mangelnden Deutschkenntnissen die erwähnte *Testfähigkeit* individuell festgestellt werden. Um auch in dieser Hinsicht internationale faire Testbedingungen zu haben, werden in PISA Kriterien für den Ausschluss von Schülerinnen und Schülern exakt definiert. Durch diese festgelegten Ausschlussgründe soll einerseits eine unzumutbare Überforderung der Schülerinnen und Schüler vermieden werden, andererseits ein unzulässiger Ausschluss von Jugendlichen, deren Teilnahme aus unterschiedlichen Gründen unerwünscht sein kann. Insgesamt darf die Ausschlussquote auf Schul- und Schülerebene nicht mehr als jeweils 2,5 Prozent betragen.

1.3.2 Teilnahmeverpflichtung

Auf Beschluss der KMK ist die Teilnahme an PISA in Deutschland generell verpflichtend. Das bedeutet, dass jede Schule, die in die Stichprobe gezogen wird, sich grundsätzlich an der PISA-Erhebung beteiligen muss. Auch jede gezogene Schülerin und jeder gezogene Schüler ist zur Bearbeitung der Tests verpflichtet. Unterschiede zwischen den 16 deutschen Bundesländern bestanden bei PISA 2012 lediglich in Bezug auf die Verpflichtung zur Bearbeitung des Schülerfragebogens. In fünf Bundesländern war auch das

Ausfüllen des Schülerfragebogens obligatorisch, nämlich in Brandenburg, Hessen, Niedersachsen, Sachsen-Anhalt und Thüringen. Zehn Bundesländer entschieden sich für eine freiwillige Beantwortung der Fragen durch die Schülerinnen und Schüler. In diesen Ländern wurde vorab eine schriftliche Genehmigung der Erziehungsberechtigten (sogenannte *Elterngenehmigung*) erbeten, ohne deren Vorliegen die betreffenden Schülerinnen und Schüler keinen Fragebogen erhalten durften. Die zehn Bundesländer mit freiwilliger Bearbeitung des Schülerfragebogens waren Baden-Württemberg, Bayern, Bremen, Hamburg, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, Sachsen und Schleswig-Holstein. In Berlin galt eine Teilverpflichtung, das heißt, Teile des Fragebogens waren zwingend auszufüllen, andere Teile auf freiwilliger Basis. Die Beteiligungsquoten am Schülerfragebogen sind im Vergleich der Bundesländer entsprechend unterschiedlich. In den fünf Ländern, die die Bearbeitung des Schülerfragebogens als obligatorisch festgelegt hatten, füllten zwischen 89 und 95 Prozent der Schülerinnen und Schüler den Fragebogen aus. In den zehn Ländern, die eine Elterngenehmigung verlangten, liegt die Beteiligung zwischen 66 und 85 Prozent. In Berlin wurde ein Rücklauf von 87 Prozent erzielt. Die Beteiligung am Schülerfragebogen lag in Deutschland insgesamt bei 82 Prozent.

1.3.3 Internationale Vergleichbarkeit der Ergebnisse

Eine der größten Herausforderungen einer internationalen Vergleichsstudie ist die Vergleichbarkeit und Fairness der Messinstrumente über die Grenzen der Teilnehmerstaaten hinweg. Dieser Anspruch der Vergleichbarkeit durchzieht die Arbeit an PISA von der Vorbereitung der Testaufgaben über die Übersetzung in die jeweilige Landessprache bis hin zur Testdurchführung sowie Kodierung (Auswertung) der Lösungen. Eine systematische Bevor- oder Benachteiligung von Schülerinnen und Schülern, etwa aufgrund einer besonders hohen oder geringen Vertrautheit mit Aufgabenformaten oder Aufgabeninhalten muss ausgeschlossen werden. Der Prozess der Aufgabenentwicklung beginnt deshalb auf der Grundlage einer Rahmenkonzeption, in der für die Testbereiche relevante Anforderungen aus einer internationalen Perspektive geklärt, begründet und strukturiert werden. In den folgenden Abschnitten werden zentrale Aspekte der Aufgabenentwicklung und des Testdesigns beschrieben.

Aufgabenentwicklung

Die in PISA 2012 eingesetzten Aufgaben in den beiden Nebendomänen Lesen und Naturwissenschaften waren bereits in früheren PISA-Erhebungsrunden entwickelt und eingesetzt worden. Zu diesen Aufgaben liegen einschlägige Erfahrungen und psychometrische Kennwerte vor, die eine hohe Qualität der Items garantieren. Auch im Schwerpunktbereich Mathematik kamen Aufgaben zum Einsatz, die bereits von früheren PISA-Kohorten bearbeitet und die als Link-Units geheim gehalten worden waren. Außerdem

wurden zahlreiche neue Aufgaben entwickelt, um den Aufgabenpool aufzufüllen und alle Inhalte, Kontexte und Prozesse der aktuellen Rahmenkonzeption für die Mathematik (OECD, 2013) angemessen umzusetzen.

Testsicherheit und Datenschutz

In groß angelegten internationalen Vergleichsstudien (oft *Large-Scale-Assessments* genannt, *LSA*) wie PISA ist die Sicherheit der Testaufgaben von größter Bedeutung. Ein ganzes Maßnahmenpaket stellt sicher, dass keine der Testaufgaben ungewollt bekannt werden.

In Deutschland wurden die Prozeduren der Datenerhebung und -verarbeitung auf der Ebene der Länder den gesetzlichen Vorgaben entsprechend gestaltet. Das nationale Projektmanagement arbeitete hier eng mit den Datenschutzbeauftragten aller Länder zusammen. Die Anonymität der gesammelten Informationen aus den Testheften und Fragebögen wurde unter Anwendung aufwändiger Verfahren gesichert. Sowohl Testhefte als auch Fragebögen erhielten eine Identifikationsnummer, wobei jede dieser Nummern eindeutig einem Schüler oder einer Schülerin zugeordnet werden konnte. Diese Identifikationsnummern wurden einmalig kurz vor Beginn der Testsitzung mit den Namen der Schülerinnen und Schüler zusammengeführt, als jeder Schülerin und jedem Schüler ein bestimmtes Testheft sowie ein Fragebogen ausgeteilt wurden. Danach wurden die Identifikationsnummern wieder von den Namen der teilnehmenden Jugendlichen getrennt. Die Namensliste hat zu keinem Zeitpunkt die Schule verlassen. Die Verarbeitung der Daten erfolgt damit vollständig pseudonomisiert, also ausschließlich auf Basis von Identifikationsnummern, sogenannten *IDs* (und nicht von Namen oder Schulen). Jedes Testheft und jeder Fragebogen erhält neben der internationalen, während des Samplings erzeugten Identifikationsnummern, auch eine nationale ID. Anhand dieser nationalen Identifikationsnummer können die Testhefte und Fragebögen lückenlos verfolgt werden, vom Druck über den Versand an die Testleiter, die Testung, den Rückversand ans IEA *Data Processing and Research Center (DPC)*, die dortige Logistik, Erfassung und Verarbeitung bis hin zum Archiv. Dort werden die Unterlagen zehn Jahre gelagert und anschließend per Aktenvernichtung vernichtet.

Antwortformate und Auswertung der Lösungen

Die in PISA 2012 eingesetzten Items lassen sich anhand der jeweils verlangten Antwortformate in drei Kategorien unterteilen: Mehrfachwahl (*Multiple Choice*), Kurzantwort (*Short Answer*) und offenes Antwortformat (*Extended Response*). Bei etwa der Hälfte der Items mussten die Schülerinnen und Schüler aus mehreren angegebenen Antwortmöglichkeiten eine oder mehrere Lösungen auswählen, die sie für korrekt hielten. Die andere Hälfte der Items erforderte eine selbst formulierte Antwort, die entweder aus einem einzelnen Wort (Kurzantwort, numerische Antworten) oder aus mehreren Sätzen beziehungsweise einer Skizze oder Zeichnung bestehen konnte (offenes Antwortformat). Die

Auswertung der Multiple-Choice-Antworten erfolgte elektronisch und automatisiert, während die offenen Antworten von qualifizierten und intensiv geschulten Kodiererrinnen und Kodierern klassifiziert und bewertet wurden. Dabei kamen standardisierte Auswertungsvorschriften zum Einsatz, deren Beherrschung Voraussetzung für eine Zulassung zur Kodierung war. Die Kodiererrinnen und Kodierer mussten jeweils entscheiden, ob die Antworten der Jugendlichen „vollständig richtig“, „teilweise richtig“ oder „falsch“ waren und daraufhin entsprechende Punktzahlen vergeben. Die überwiegende Mehrheit der Items war dichotom, es wurden also lediglich „richtige“ von „falschen“ Antworten unterschieden. Auch dieser Vorgang erfolgt international zuverlässig und nach gleichen Maßstäben. Zum Zweck der Qualitätssicherung wurden aus jedem Teilnehmerstaat Antworten aus einer Auswahl von Testheften von vier Kodiererrinnen und Kodierern unabhängig bewertet und die Übereinstimmung dieser Auswertungen international überprüft. Item für Item wurde vom internationalen Konsortium bei zu geringer Übereinstimmung mit den betreffenden Staaten Kontakt aufgenommen und geklärt, worauf die zu großen Diskrepanzen beruhen. Dies war in Deutschland bei keinem Item der Fall.

Übersetzung und Anpassung

Damit alle an PISA teilnehmenden Schülerinnen und Schüler die Testaufgaben in ihrer Unterrichtssprache bearbeiten konnten, wurden die Tests und Fragebögen nach abgeschlossener Aufgabenentwicklung in den PISA-Staaten übersetzt. Die nationalen Projektmanager erhielten dazu in PISA 2012 zwei ursprüngliche Versionen, von denen eine auf Englisch und eine auf Französisch formuliert war. Die englische Version wurde von zwei unabhängigen Übersetzungsteams in die jeweilige Landes- und Testsprache übertragen und dann von einer dritten Übersetzerin oder einem Übersetzer unter Abgleich mit der französischen Version in eine Schlussversion zusammengeführt. Diese Schlussversion wurde in Deutschland von nationalen Expertinnen und Experten aus der jeweiligen Fachdidaktik überprüft und begutachtet. Besonderes Augenmerk galt der inhaltlichen Korrektheit, der Angemessenheit des sprachlichen Niveaus für die Altersgruppe der Fünfzehnjährigen und der Verwendung von Fachbegriffen, die dem schulischen Sprachgebrauch entsprachen, mit dem die Schülerinnen und Schüler vertraut sind. Die Übersetzung der deutschsprachigen Aufgaben und Fragebögen erfolgte in enger Abstimmung der Projektmanager aus Österreich, der deutschsprachigen Schweiz, Luxemburg und Deutschland. Deutschland übernahm dabei primär die Übersetzung der neu entwickelten Mathematikaufgaben. Die von den nationalen Projektmanagern fertig gestellten Aufgaben und Fragebögen wurden abschließend von Übersetzungsexpertinnen und -experten des internationalen Konsortiums genauestens kontrolliert und freigegeben. Die Übersetzung erfolgte bereits im Format der später eingesetzten Testhefte, das heißt, Abbildungen und Tabellen waren in den Dokumenten ebenso enthalten wie Texte und Umbrüche. So konnte sichergestellt werden, dass alle Schülerinnen und Schüler an derselben Stelle im Text in ihrer jeweiligen Sprache umblättern mussten und stets dieselben Informationen vor sich liegen hatten.

Testheftdesign

Ein weiterer wesentlicher Aspekt der internationalen Vergleichbarkeit von Schulleistungsstudien wie PISA betrifft die Zusammenstellung der Aufgaben in den Testheften. Eine Aufgabe (*Unit*) enthält mehrere Fragen (*Items*), die unterschiedlich schwierig sind und unterschiedliche Wissensfacetten ansprechen. Da die in den theoretischen Rahmenkonzeptionen vorgegebenen Inhalte über die Aufgaben weitgehend abgedeckt werden müssen, übersteigt die Aufgabenzahl die verfügbare individuelle Testzeit um ein Mehrfaches. Aus diesen Gründen wird bei PISA das oben skizzierte *Multi-Matrix-Design* verwendet, das Aufgabencluster systematisch rotiert. Bei PISA 2012 wurden Testhefte mit variierten Clustern eingesetzt.

Die hohen Standards, die für die Vorbereitung und Durchführung von PISA gefordert werden, sind international für eine hohe Qualität und solide Vergleichbarkeit der Ergebnisse unabdingbar und müssen daher weltweit eingehalten werden. Daher wurden in PISA 2012 in allen Phasen des Projekts umfassende Maßnahmen zur Qualitätssicherung ergriffen. Die erste solche Maßnahme ist die Vorbereitung und Ziehung einer *repräsentativen Stichprobe*, die unverzerrt ein Abbild der gesamten Zielpopulation in einem Staat liefert (vgl. auch Kapitel 10, Technische Grundlagen). Die Verfahren der Vorbereitung und Ziehung der Stichprobe werden international streng überwacht und zum Teil auch durch Vertreter des internationalen Konsortiums für PISA 2012 durchgeführt. Weitere Vorschriften und Kontrollen betreffen die Prozesse der Aufgabenentwicklung, die Übersetzung der Testaufgaben und Fragebögen und die Durchführung der Datenerhebung. Eine wichtige Rolle spielt hier die Kompetenz und Zuverlässigkeit der Testleiterinnen und Testleiter. Die Rekrutierung und Schulung übernahm in Deutschland der Vertragspartner IEA *Data Processing and Research Center* in Hamburg. Detaillierte Handbücher (Manuale) beschreiben die Anforderungen der Testung; ausgearbeitete Testleiterskripte müssen in der Testsitzung genau abgearbeitet werden, um in allen Schulen einen standardisierten Ablauf der PISA-Erhebung zu gewährleisten. Darüber hinaus wurde ein umfangreiches *Qualitätsmonitoring* implementiert. Mehrere unabhängige, geschulte Beobachter waren am PISA-Testtag unangekündigt in teilnehmenden Schulen anwesend, um sich zu vergewissern, dass die Testsitzung vorschriftsgemäß ablief und die Standardisierung nicht beeinträchtigt war. Neben internationalen Beobachtern wurden in Deutschland weitere speziell trainierte Personen für die Qualitätssicherung engagiert. Das *Institut zur Qualitätsentwicklung im Bildungswesen* (IQB in Berlin) wurde gebeten, auf der Basis der Beobachtungsdokumente einen Bericht zu verfassen und die ordnungsgemäße Durchführung zu überwachen. Es wurden keine Auffälligkeiten oder Abweichungen vom internationalen Vorgehen verzeichnet.

1.3.4 Skalierung und Auswertung der Daten

Das Multi-Matrix-Design von PISA 2012 ist die in internationalen Bildungsvergleichsstudien gängige Methode, um ein möglichst breites inhaltliches Kompetenzspektrum abzudecken und zugleich die Testbelastung für die Schülerinnen und Schüler in einem zumutbaren Rahmen zu halten. Jeder an PISA teilnehmende Jugendliche hatte lediglich eine Teilmenge aller Aufgaben zu bearbeiten. Um die Ergebnisse aus den unterschiedlichen Testheften vergleichen zu können, müssen die Leistungen auf einer gemeinsamen Skala abgebildet werden. Anhand von Verallgemeinerungen des Rasch-Modells (Rasch, 1960) werden Modelle aus der *Item Response Theory* verwendet, um die Verteilung der von den Schülern erreichten Leistungen abzubilden (vgl. Kapitel 10, Technische Grundlagen). In PISA 2012 wurden wie in den vergangenen Zyklen auch drei Kompetenzbereiche erhoben, die im Rahmen des Rasch-Modells als einzelne Dimensionen beschrieben werden. Entsprechend werden für die Skalierung mehrdimensionale Modelle zur Analyse latenter Zusammenhänge verwendet (Adams, Wilson & Wang, 1997).

Die Ergebnisse, die mit Hilfe von PISA 2012 gewonnen werden, werden insbesondere aus zwei Perspektiven betrachtet: *staatenübergreifend* bezogen auf die 34 OECD-Staaten (sowie weitere Partnerstaaten) und *national* mit besonderem Augenmerk auf einen einzelnen Staat. Die Daten der an PISA 2012 teilnehmenden Jugendlichen werden seitens des internationalen Konsortiums im Sinne des staatenübergreifenden Vergleichs gemeinsam skaliert.

Die Wahl der Skalierungsverfahren ist stark von den angestrebten Aussagen abhängig, die auf der Basis der gewonnenen Daten gemacht werden sollen. Die Auswertung der Daten aus PISA sieht Aussagen über Grundgesamtheiten (Populationen) oder definierte Subpopulationen vor; Aussagen über individuelle Jugendliche werden hingegen nicht gemacht. Dementsprechend werden in PISA Skalierungsverfahren angewendet, die für die Analysen von Kompetenzverteilungen in Teilpopulationen optimal geeignet sind. Zentral für Studien wie PISA ist dabei eine verlässliche Schätzung von Verteilungsparametern sowie die Analyse von Zusammenhängen zwischen Kompetenzen und Merkmalen der Schülerinnen und Schüler (*Hintergrundmerkmale*). Solche Zusammenhänge werden mit Hilfe von *latenten Regressionsanalysen* direkt geschätzt (Mislevy, 1984). Zu den relevanten Hintergrundvariablen zählen unter anderem das Geschlecht, Angaben zu den Elternhäusern und zur besuchten Schule sowie zum Unterricht und alle weiteren Variablen, deren Zusammenhang mit den Kompetenzen der Jugendlichen untersucht wird. Geschätzt werden dabei Verteilungsparameter (*Marginal-Maximum-Likelihood*, MML), welche durch sogenannte *Plausible Values* in Datensätze auf Individualebene überführt werden. Auf der Ebene der Schüler werden also jeweils mehrere Datensätze erzeugt, die simultan in die Analysen eingehen. So ist eine weitere statistische Auswertung mit gängigen Verfahren und Software möglich (Mislevy, Beaton, Kaplan & Sheehan, 1992; OECD, 2012b; Prenzel et al., 2007). Ein entsprechendes Modell, welches alle beschriebenen Modellierungen integriert, wurde von ACER konstruiert und mit der Software

ConQuest 3.0.1 (Adams, Wu & Wilson, 2012) implementiert. Alle Skalierungen für PISA 2012 wurden mit dieser Software durchgeführt.

1.4 Darstellung der Ergebnisse aus PISA

Im Folgenden werden die häufigsten und wichtigsten Arten der Ergebnisdarstellung in diesem Berichtsband erläutert. Darstellungen von PISA-Ergebnissen in Medien erwecken häufig den Eindruck, als ginge es bei dieser internationalen Vergleichsstudie nur darum, Rangplätze und somit Gewinner und Verlierer zu berichten. Dem ist jedoch nicht so. Pauschal kann man wohl davon ausgehen, dass Staaten an der Spitze dieser veröffentlichten Rangordnungen etwas weniger Probleme im Bildungssystem haben als Staaten am unteren Ende. Aus der Sicht verantwortlicher Akteure in einem Bildungssystem interessieren in erster Linie Informationen über Probleme oder Trendinformationen, die Rückmeldung über mögliche Wirkungen von Maßnahmen geben. Viele Informationen müssen an anderen Vergleichsdaten relativiert und statistisch kontrolliert werden, um die Befunde sinnvoll interpretieren zu können. Bei genauerer Betrachtung liefern die typischerweise bei PISA berichteten Tabellen sehr viel mehr Information als nur die Angaben von Mittelwerten, aus denen Rangordnungen gebildet werden. Das beginnt bei Kennwerten für die Streuung in Verteilungen und schließt statistische Maßzahlen ein, die für die Interpretation von Unterschieden nützlich sind. Zu erläutern ist auch der Begriff der Kompetenzstufe. Mit der Unterscheidung von Kompetenzstufen für die Domänen wird illustriert, was Schülerinnen und Schüler auf unterschiedlichen Abschnitten der Punkteskala tatsächlich können. Diese Betrachtung hilft dann auch Bezüge zu den Anforderungen von Lehrplänen oder Standards herzustellen.

Mittelwerte und Streuungsmaße (Standardabweichungen)

Mittelwerte sagen etwas über das im Durchschnitt erreichte Niveau – etwa in einem Mathematiktest – aus. Je nach Verteilung können die Leistungsdaten aber unterschiedlich um diesen Mittelwert streuen. Für die Interpretation von Bildungsergebnissen ist es wichtig zu wissen, wie homogen (oder heterogen) eine Verteilung von Leistungspunkten ist. Statistische Kennzahlen für den Mittelwert und die Streuung (z. B. die Standardabweichung) sind für die Interpretation der Ergebnisse grundlegend. Zur Erleichterung von Vergleichen zwischen den OECD- und Partnerstaaten sowie für Vergleiche über die Zeit werden die Mittelwerte und Standardabweichungen normiert. Konkret werden diese Kennzahlen für die erste Erhebung einer Hauptdomäne statistisch so transformiert, dass für die Gruppe der OECD-Staaten der *Mittelwert* (M) bei 500 Punkten liegt und die *Standardabweichung* (SD) bei 100 Punkten. Diese Normierung ist dann auch der Bezugspunkt für nachfolgende Erhebungen: Wenn sich die Leistungen der OECD-Staaten insgesamt verschlechtern, sinkt der Mittelwert unter 500. Ähnlich können Änderungen der Standardabweichungen eine Erhöhung oder Reduzierung der Leistungsstreuung

anzeigen. Die Ergebnisse der Partnerstaaten werden den OECD-Kennwerten zugeordnet; sie verändern also nicht die Mittelwerte und Standardabweichungen der OECD-Staaten.

Unter Annahme einer Normalverteilung der Kompetenzen liegen etwa zwei Drittel (68.2 Prozent) der Schülerinnen und Schüler aus den teilnehmenden OECD-Staaten im Bereich von ± 1 (einer) Standardabweichung um den Mittelwert (400 bis 600 Punkte). Das bedeutet, dass die Kompetenzwerte von 15.9 Prozent der Jugendlichen mehr als eine Standardabweichung unter dem Durchschnitt liegen und entsprechend die Werte von 15.9 Prozent den Mittelwert deutlich – um mindestens eine Standardabweichung – übertreffen. Innerhalb von zwei Standardabweichungen um den Mittelwert (300 bis 700 Punkte) liegen 95.4 Prozent der Kompetenzwerte. Fast alle Schülerinnen und Schüler werden zum Beispiel im Bereich Mathematik also einem Wert zwischen 300 und 700 Punkten zugeordnet, lediglich 2.3 Prozent liegen bei jeweils weniger als 300 oder mehr als 700 Punkten. Diese Einteilung der erzielten Kompetenzwerte in einen Mittelwert und Standardabweichung entspricht einer Standardisierung der Daten, die auch für den internationalen Vergleich der Schülerinnen und Schüler hilfreich ist.

Perzentile und Perzentilbänder

Neben Mittelwert und Standardabweichung werden in vielen Abbildungen für den internationalen Vergleich in PISA Angaben über *Perzentile* gemacht. Ein *Perzentilwert* ist eine Art Schnittpunkt, der einen in Prozent angegebenen Bereich einer Verteilung vom restlichen Bereich trennt. Am 95. Perzentil etwa wird derjenige Kompetenzwert berichtet, ab dem die oberen 5 Prozent der Verteilung beginnen. Beispielhaft für den Bereich Mathematik können Perzentilangaben dazu verwendet werden, Aussagen darüber zu machen, wie viele Schülerinnen und Schüler aus der Stichprobe einen ähnlichen hohen Wert erzielt haben beziehungsweise wie groß der Anteil von Jugendlichen ist, die über diesem Leistungswert liegen (bei einem Perzentilwert von 75 sind das 25 Prozent). Perzentilwerte erlauben auch eine Aussage darüber, auf welchem Leistungsniveau zum Beispiel die besten 5 Prozent unterschiedlicher Staaten beginnen. Graphisch werden Perzentile im vorliegenden Bericht auch als *Perzentilbänder* dargestellt (vgl. Abbildung 1.1). Diese Bänder veranschaulichen zugleich die Breite einer Leistungsverteilung und werden deshalb in den Kapiteln im Staatenvergleich abgebildet.

Kompetenzstufen

Die in PISA verwendeten Kompetenzskalen werden auch in sogenannte Kompetenzstufen unterteilt. Diese Kompetenzstufen entsprechen auf statistischer Ebene einem Wertebereich mit einem definierten Minimum und Maximum von Kompetenzpunkten und auf Fachdomänenebene einer entsprechend zugewiesenen Beschreibung der Kompetenzen, die Jugendliche typischerweise auf den einzelnen Kompetenzstufen beherrschen. Dazu gehört auch die Definition der kognitiven Anforderungen, die benötigt werden, um die der entsprechenden Kompetenzstufe zugeordneten Aufgaben lösen zu können.

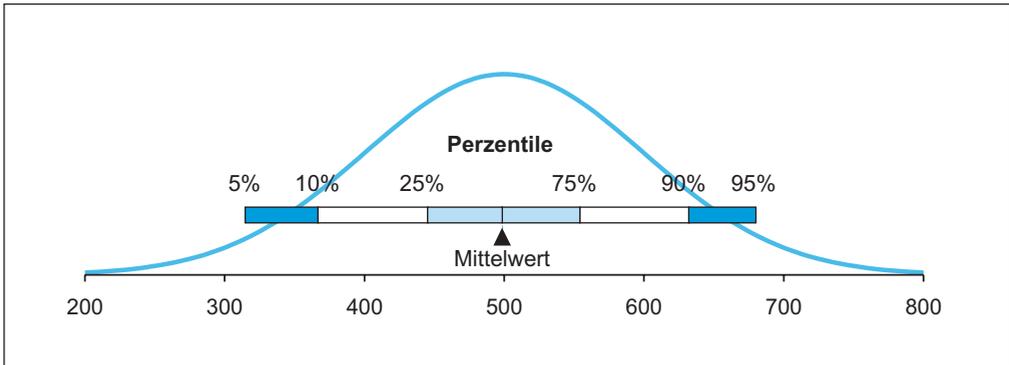


Abbildung 1.1: Perzentilband und Normalverteilungskurve

Formal bedeutet dies, dass Schülerinnen und Schüler mit einer für eine Stufe mittleren Kompetenz eine für diese Stufe mittelschwierige Aufgabe mit einer Wahrscheinlichkeit von 62 Prozent lösen können (Prenzel et al., 2007). Aufgaben, deren Schwierigkeiten einer höheren Kompetenzstufe entsprechen, werden von der gleichen Schülerin oder dem gleichen Schüler mit einer deutlich geringeren Wahrscheinlichkeit gelöst. Für Aufgaben aus niedrigeren Kompetenzstufen gilt eine entsprechend höhere Lösungswahrscheinlichkeit.

Kompetenzstufen in PISA sind demnach eine inhaltliche Interpretation der Tests und eine Form der Übersetzung von Punkten in anschauliche Kompetenzbeschreibungen der Schülerinnen und Schüler. Von besonderem Interesse sind hier diejenigen Jugendlichen, deren Kompetenzen auf sehr schlechte Chancen für ein nachfolgendes Lernen innerhalb und außerhalb der Schule schließen lassen. Schülerinnen und Schüler, deren Kompetenz sich auf oder unterhalb der ersten Kompetenzstufe befindet, werden vermutlich große Schwierigkeiten beim Übergang von der Schule in ein eigenständiges Erwerbsleben (etwa im Rahmen einer Berufsausbildung) haben. Auf der anderen Seite lassen Schülerinnen und Schüler auf der obersten Kompetenzstufe ein besonderes Potential für Spitzenleistungen in diesen Bereichen erkennen. An den Anteilen dieser obersten Kompetenzstufen lässt sich ablesen, inwieweit es gelingt, in einem Bildungssystem Nachwuchs mit Spitzenleistungen zu fördern.

Standardfehler und Signifikanztests

Studien wie PISA haben zum Ziel, von in Stichproben gemessenen Kennwerten auf entsprechende Kennwerte in der Grundgesamtheit zu schließen. Entsprechende statistische *Schätzungen* sind immer mit einer gewissen Unsicherheit und Messungenauigkeit verbunden. Ein aussagekräftiges Maß für diese Unsicherheit ist der sogenannte *Standardfehler*. Je kleiner der Standardfehler ist, desto genauer ist ein aus der Stichprobe geschätzter Populationskennwert. Ähnlich wie bei der oben beschriebenen Standardabweichung dient der Standardfehler zur Definition eines Wertebereichs, der für die Ein-

ordnung und Interpretation der Kennwerte hilfreich ist. Der Standardfehler gibt an, in welchem Wertebereich der „wahre Wert“ des Populationskennwerts mit einer bestimmten *Irrtumswahrscheinlichkeit* liegt. Per Konvention von meistens maximal 5 Prozent Irrtumswahrscheinlichkeit entspricht die Spannweite des Wertebereichs ± 1.96 Standardfehlern. Damit erhält man ein Maß dafür, inwieweit eine andere Stichprobe aus derselben Population zu anderen Kennwerten geführt hätte. Insofern können die angegebenen Standardfehler auch dazu verwendet werden, anhand der geschätzten Populationswerte zu prüfen, ob sich die Mittelwerte zweier Gruppen statistisch signifikant voneinander unterscheiden. Nur dann kann tatsächlich von Unterschieden zwischen Gruppen gesprochen werden, die nicht auf die Ungenauigkeit der Studie (etwa bei der Stichprobenziehung oder der Messung von Kompetenzen und Einstellungen) zurückzuführen sind. Auf mögliche Quellen von Ungenauigkeiten (Stichprobenfehler, Messfehler), die im Rahmen groß angelegter Studien wie PISA berücksichtigt werden müssen, wird in Kapitel 10 (Technische Grundlagen) eingegangen.

Praktische Relevanz und Effektstärken

Um beurteilen zu können, ob die Unterschiede in den Mittelwerten von zwei oder mehr Stichproben statistisch aussagekräftig sind, werden sogenannte Signifikanztests durchgeführt. Mit einer vorher gesetzten Irrtumswahrscheinlichkeit (z. B. 5 oder 1 Prozent) wird geprüft, ob der Unterschied zwischen den Stichprobenkennwerten auch in der Grundgesamtheit besteht. Bei einem Signifikanzniveau von 5 Prozent wird man in fünf von hundert Fällen fälschlicherweise einen Unterschied behaupten, bei einem Signifikanzniveau von 1 Prozent reduziert sich die Irrtumswahrscheinlichkeit weiter auf einen Fall von hundert usw.

Bei üblichen Signifikanztests spielt (neben der Größe des Unterschieds in den Kennwerten) auch die Größe der Stichprobe eine wichtige Rolle. Bei einer sehr großen Stichprobe kann ein Unterschied statistisch signifikant werden, der bei einer kleinen Stichprobe als nicht signifikant berechnet worden wäre. Die statistische Signifikanz eines Unterschieds sagt also wenig über die Größenordnung und Relevanz des Unterschieds aus. *Large-Scale-Assessments* wie PISA zeichnen sich unter anderem durch sehr große Stichprobenumfänge aus. Deshalb können hier dem Betrag nach sehr kleine Unterschiede statistisch signifikant werden. Im vorliegenden Berichtsband wird daher bei Mittelwertvergleichen auch das Maß der Effektstärke angegeben. Die Effektstärke ist ein Maß für Unterschiede, das an der Streuung der Werte relativiert ist. Sie gibt an, ob ein Gruppenunterschied beispielsweise weniger bedeutsam ist, weil die Streuung des Merkmals jeweils in der Gruppe ohnehin relativ hoch ist beziehungsweise ob ein solcher Unterschied weniger bedeutsam ist, weil die Streuung des Merkmals innerhalb der Gruppen vergleichsweise gering ausfällt. In PISA könnte ein solches Merkmal etwa die mathematische Kompetenz bei Mädchen und Jungen sein. Für Mittelwertsunterschiede wird die Effektstärke als Quotient der Differenz der Mittelwerte und der Standardabweichung berechnet. Effektstärken werden gern in Anlehnung an Cohen (1988) so inter-

pretiert, dass die Maßzahl d von 0.2 als „kleiner“, von 0.5 als „mittlerer“ und von 0.8 als „großer“ Effekt bezeichnet wird. Speziell im Kontext von Vergleichsstudien ist eine Übersetzung solcher Effektstärken in Entwicklungszeiten wie Schuljahre anschaulicher. Verschiedene frühere Analysen in Deutschland, insbesondere eine PISA-Messwiederholung aus den Jahren 2003/2004 (Prenzel et al., 2006), können als Referenzpunkt dienen, in deren Rahmen Kompetenzunterschiede innerhalb eines Schuljahres gemessen wurden. Als grobe Faustregel entsprechen Unterschiede in der Größenordnung von etwa 25 bis 30 Punkten einem durchschnittlichen Entwicklungsabstand oder Leistungszuwachs von einem Schuljahr. Aufgrund der in PISA angewandten Metrik entsprechen 25 bis 30 Punkte 0.25 beziehungsweise 0.30 Standardabweichungen, wie dies durch Cohen's d ausgedrückt wird.

1.5 Interpretation der Befunde

Das Hauptaugenmerk des vorliegenden ersten Berichts zu PISA 2012 in Deutschland liegt auf der Beschreibung bedeutsamer Bildungsergebnisse. Ein zentrales Anliegen von PISA ist es darüber hinaus auch, Zusammenhänge zwischen Kompetenzmaßen und weiteren theoretisch wie praktisch relevanten Merkmalen zu identifizieren und darzustellen. PISA ermöglicht eine Beschreibung vieler Merkmale, die aus theoretischer Sicht als Voraussetzungen oder Bedingungen für den Erwerb von Bildung und Kompetenzen gelten können. Solche Befunde über das Zusammenspiel verschiedener Merkmale mit Kompetenzen können im Sinne eines Bildungsmonitorings diskutiert und in einer Bildungslandschaft verortet werden. Wie bereits angesprochen können Aussagen über Zusammenhänge in PISA nicht kausal interpretiert werden. Die Prüfung der kausalen Relevanz von Einflussfaktoren setzt eine Reihe von Bedingungen voraus (z. B. mehrere Messzeitpunkte, systematische Variation von Bedingungen, Kontrolle weiterer Einflüsse), die bei PISA normalerweise nicht erfüllt sind. Problematisch sind deshalb Aussagen über Ursachen für festgestellte Unterschiede im Vergleich von Bildungsergebnissen oder Merkmalen von Lernumgebungen. Die hier berichteten statistischen Auswertungen beziehen sich ausschließlich auf gegenseitige Verbindungen von Kompetenzen und bestimmten Hintergrundmerkmalen von Schülern. Das Einhergehen hoher oder niedriger Kompetenzen bei den Jugendlichen mit einzelnen ihrer Merkmale ist zwar eine notwendige, nicht aber eine hinreichende Bedingung für kausale Relevanz.

1.6 Nationale und internationale Organisation der PISA-Studie

PISA ist das Ergebnis der Zusammenarbeit einer Vielzahl von Instituten und Organisationen, sowohl auf nationaler als auch auf internationaler Ebene. Die politische Steuerung obliegt dem sogenannten PISA *Governing Board* (PGB), in welchem alle Teilnehmerstaaten mit mindestens einer Vertreterin oder einem Vertreter repräsentiert sind. Für Deutschland sind dies in PISA 2012 MR'in Elfriede Ohrnberger (Bayerisches Staatsministerium für Unterricht und Kultus), Dr. Annemarie Klemm (Freie Hansestadt Bremen, Referat 20 Qualitätsentwicklung und Standardsicherung) sowie Dr. Susanne von Below (Bundesministerium für Bildung und Forschung). Die Koordination auf internationaler Ebene hat das Sekretariat der OECD in Paris inne, verantwortlich ist dort Andreas Schleicher. Unter der Federführung von ACER und dort unter der Leitung des Projektleiters Raymond J. Adams wurde ein internationales Konsortium mit der wissenschaftlichen Durchführung von PISA beauftragt. In diesem Konsortium kooperieren die Institute ACER, aSpE, cApStAn, DIPE, ETS, ILS, IPN, NIER, CRP/EMACS und WESTAT. Die theoretischen Rahmenkonzeptionen wurden von international bestellten Experten Gruppen zu den Fachdomänen (Lesen, Mathematik, Naturwissenschaften) sowie zur Entwicklung der Fragebögen und zur Sicherung der technischen Standards ausgearbeitet.

Auf nationaler Ebene sind in allen 65 PISA-Teilnehmerstaaten sogenannte *National Project Managers* (NPM) für die sachgemäße Vorbereitung und Durchführung der PISA-Studie verantwortlich. In Deutschland wurde das nationale Projektmanagement nach Ausschreibung durch die KMK an das Zentrum für Internationale Vergleichsstudien (ZIB) vergeben. Das ZIB wurde im Jahr 2010 vom Bundesministerium für Bildung und Forschung (BMBF) und der Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland (KMK) als An-Institut der Technischen Universität München (TUM) gegründet. Als gemeinnütziger Verein bündelt das ZIB die Expertise dreier Forschungseinrichtungen im Bereich der empirischen Bildungsforschung: die TUM School of Education (München), das Deutsche Institut für Internationale Pädagogische Forschung (DIPE, Frankfurt) und das Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN, Kiel) haben sich unter dem Dach des ZIB zusammengeschlossen. Mit Prof. Dr. Manfred Prenzel hat der Dekan der TUM School of Education an der TUM und zugleich Vorstandsvorsitzender des ZIB, die Funktion des nationalen Projektmanagers übernommen. Das ZIB stimmt seine Arbeit mit der Amtschefskommission „Qualitätssicherung in Schulen“ ab (Vorsitz: Ministerialdirektor Dr. Peter Müller, Bayern, und Staatssekretär Burkhard Jungkamp, Brandenburg).

Der PISA-Arbeitsgruppe unter Leitung von Manfred Prenzel an der TUM School of Education gehören und gehörten an: Elisabeth Blaser, Jörg-Henrik Heine, Stefan Hofherr, Dr. Katharina Hohn, Julia Mang, Jessica Mattern, Dr. Katharina Müller, Dr. Johanna Ray, Dr. Christine Sälzer, Dr. Anja Schiepe-Tiska, Stefanie Schmidtner, Fabian Zehner sowie zahlreiche studentische Hilfskräfte. Ihre Expertise zur Hauptdomäne Mathematik brachten ein: Prof. Dr. Kristina Reiss, Prof. Dr. Anke Lindmeier (TU

München), Prof. Dr. Stefan Ufer und Prof. Dr. Aiso Heinze (IPN Kiel). Darüber hinaus besteht eine enge Zusammenarbeit der PISA-Arbeitsgruppe mit dem IQB in Berlin (Prof. Dr. Petra Stanat, Prof. Dr. Hans Anand Pant, Dr. Ulrich Schroeders), dem IPN in Kiel (Dr. Silke Rönnebeck, Dr. Katrin Schöps) und dem DIPF in Frankfurt (Dr. Nina Jude, Dorothea Mildner, Dr. Jan Hochweber).

Die Vorbereitung und Durchführung der Datenerhebung und -erfassung lag wie in früheren PISA-Erhebungsrounds beim IEA DPC in Hamburg. Die Projektkoordination dort hatte Heiko Sibbers inne. Zuständig waren darüber hinaus Anja Waschk, Jennifer Oberlein, Tobias Haring (Feldarbeit), Dr. Lars Borchert, Jens Gomolka (internationales Datenmanagement), Guido Martin, Daniela Specks, Steffi Kadenbach (Kodierung), Wolfram Jarchow und Regina Borchhardt (nationales Datenmanagement).

Durch das Engagement aller beteiligten Schulen konnte mit PISA 2012 bereits der fünfte Durchgang des Projekts erfolgreich durchgeführt werden. Die Bereitschaft der Schulen, diese Vergleichsstudie durch ihre Teilnahme zu unterstützen und aus den Ergebnissen zu lernen, ist für das Gelingen eines großen Unternehmens wie PISA unerlässlich. Das nationale Projektmanagement dankt allen Schülerinnen und Schülern, ihren Eltern und Lehrern und ihren Schulleitungen herzlich für ihre Mitarbeit!

Literatur

- Adams, R. J., Wilson, M. & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21 (1), 1–23.
- Adams, R. J., Wu, M. & Wilson, M. (2012). ACER ConQuest [Computer software]: Australian Council for Educational Research (ACER). Zugriff am 23.09.2013. Verfügbar unter <http://www.acer.edu.au/conquest/overview2>
- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.
- Bromme, R., Prenzel, M. & Jäger, M. (2011). *Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. Eine Analyse von Anforderungen an die Darstellung, Interpretation und Rezeption empirischer Befunde. Gutachten für das BMBF*. Münster: Universität Münster.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education*, 30 (3), 319–330.
- Klieme, E. & Vieluf, S. (2013). Schulische Bildung im internationalen Vergleich. Ein Rahmenmodell für Kontextanalysen in PISA. In N. Jude & E. Klieme (Hrsg.), *PISA 2009 – Impulse für die Schul- und Unterrichtsforschung* (S. 229–246). Weinheim: Beltz.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. München: Wolters Kluwer.

- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2010). *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Neuwied: Carl Link.
- Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, 49 (3), 359–381.
- Mislevy, R. J., Beaton, A. E., Kaplan, B. & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29, 133–161.
- OECD. (2006). *Assessing scientific, reading and mathematical literacy. A framework for PISA 2006*. Paris: OECD.
- OECD. (2009). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris: OECD.
- OECD. (2012a). *Education at a glance 2012. OECD Indicators*. Paris: OECD.
- OECD. (2012b). *PISA 2009 technical report*. Paris: OECD.
- OECD. (2013). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. et al. (Hrsg.). (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M. et al. (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.
- Prenzel, M. & Lankes, E.-M. (2013). Was können Schülerinnen und Schüler über ihren Unterricht sagen? Ein Blick in die Schülerfragebogen von internationalen Vergleichsstudien. In N. McElvany & H. G. Holtappels (Hrsg.), *Empirische Bildungsforschung. Theorien, Methoden, Befunde und Perspektiven* (S. 93–107). Münster: Waxmann.
- Rasch, G. W. (1960). *Probabilistic models for some intelligence and attainment tests* (Studies in mathematical psychology). Chicago: The University of Chicago Press.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Berlin: Hans Huber.
- Seidel, T. & Prenzel, M. (2008). Large scale assessment. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of competencies in educational contexts* (S. 279–304). Göttingen: Hogrefe & Huber.
- Tenorth, H.-E. (2004). Bildungsstandards und Kerncurriculum. Systematischer Kontext, bildungstheoretische Probleme. *Zeitschrift für Pädagogik*, 50 (5), 650–661.
- Tenorth, H.-E. (2005). „Grundbildung“, „Basiskompetenzen“ und allgemeine Bildung. In A. Kauf, K. Liebers & A. Prengel (Hrsg.), *Länderübergreifende Curricula für die Grundschule* (S. 93–107). Bad Heilbrunn: Klinkhardt.

2 Zwischen Grundlagenwissen und Anwendungsbezug: Mathematische Kompetenz im internationalen Vergleich

Christine Sälzer, Kristina Reiss, Anja Schiepe-Tiska,
Manfred Prenzel und Aiso Heinze

2.1 Einleitung

Mathematik ist eine der ältesten Wissenschaften und gilt seit der Antike als wichtiger Bestandteil der Kultur. Teilbereiche der Mathematik, wie Arithmetik und Geometrie, waren damals bereits in Bildungskonzepten wie den „sieben freien Künsten“ (*septem artes liberales*, Gleis, 2006) verankert. So prägt die Mathematik seit langer Zeit auch schulische Curricula. Für die Bildung der breiten Bevölkerung hat die Mathematik spätestens seit dem Mittelalter über das bürgerliche Rechnen an alltäglicher Bedeutung gewonnen. Insbesondere ein Name wie Adam Ries, der sich über ein entsprechendes Lehrbuch um das mathematische Wissen von Handwerkern und Kaufleuten verdient gemacht hat, steht dabei auch für eine frühe Orientierung an alltäglichen Anwendungen der Mathematik. Grundlagen und Anwendungen der Mathematik sind heute weltweit ein wichtiger Bestandteil des Curriculums; eine Schullaufbahn ohne mathematische Bildung ist praktisch nicht vorstellbar. Jahrzehntlang war man in Deutschland der Ansicht, der Mathematikunterricht sei im Vergleich zu anderen Staaten als anspruchsvoll und erfolgreich zu bewerten, ohne dass es dafür belastbare empirische Evidenz gab. Es war die *Third International Mathematics and Science Study* (TIMSS), die eine gesicherte Standortbestimmung erlaubte (Baumert et al., 1997). Die erwartungswidrig mittelmäßigen Leistungen von Schülerinnen und Schülern in Deutschland lösten einen Schock aus, der durch die Ergebnisse der ersten PISA-Studie noch einmal verstärkt wurde. Hier lagen die Mathematikleistungen signifikant unter dem Durchschnitt der OECD-Staaten (Klieme, Neubrand & Lüdtke, 2001). Hinzu kamen Befunde über sehr hohe Anteile von Jugendlichen (fast ein Viertel der Alterskohorte), deren mathematische Kompetenz kaum über das Niveau der Grundschule hinausreichte, sowie Belege für erhebliche Disparitäten in Bezug auf Region, Zuwanderungshintergrund und Geschlecht (Neubrand & Klieme, 2002). Die Ergebnisse zeichneten das Bild eines wenig effizienten und sozial ungerechten Bildungssystems in Deutschland, das im Vergleich der OECD-Länder kaum mehr konkurrenzfähig war.

Erste Gründe für die schwachen Leistungen in der Mathematik waren bereits in einer Problemanalyse in Folge der wenig zufriedenstellenden TIMSS-Ergebnisse herausgearbeitet worden (BLK Projektgruppe Innovationen im Bildungswesen, 1997). Auf der

Basis zusätzlicher Befunde aus der TIMSS-Video-Studie sowie aktueller Erkenntnisse der mathematikdidaktischen und allgemeinen Lehr-Lern-Forschung wurden unter anderem Belege für eine höchst problematische Aufgabenqualität, unzureichende Förderung des Grundwissens, mangelnde Abstimmung über die Jahrgangsstufen und fehlende Qualitätssicherung gefunden (vgl. z. B. die Beiträge in Klieme & Baumert, 2001).

Die Problemanalyse nach TIMSS beschränkte sich naturgemäß vor allem auf den mathematisch-naturwissenschaftlichen Unterricht. Mit den PISA-Ergebnissen rückten jedoch auch allgemeine Aspekte in den Vordergrund. So wurde durch die bei PISA festgestellten schwachen Ergebnisse im Lesen offensichtlich, dass es nicht nur um begrenzte Schwachstellen im Bereich von Mathematik und Naturwissenschaften ging, sondern um generelle Probleme im deutschen Schulsystem. Als ein wesentlicher Faktor wurde die fehlende Kontrolle von Bildungsergebnissen bei der Steuerung von Bildungsprozessen identifiziert – und dies auf allen Ebenen, vom Klassenzimmer über die Schule bis zur Schulaufsicht und Bildungspolitik.

In der Folge wurde daher die Idee aufgegriffen, dass Schule in Deutschland neben einer curriculumorientierten, inputgeleiteten Steuerung auch von den Wirkungen und Ergebnissen von Bildungsprozessen her gedacht werden muss (Outputsteuerung). Eine viel beachtete Expertise von Klieme et al. (2003) machte deutlich, dass länderübergreifende Bildungsstandards hier ein probates Mittel sein könnten. Bildungsstandards legen fest, welche Ergebnisse Schule und Unterricht am Ende eines festen Zeitabschnitts zeigen sollen. Werden sie auf nationaler Ebene definiert, so könnten sie auch geeignet sein, bestehende Disparitäten zu verringern. Die Einführung von Bildungsstandards (vgl. u. a. KMK, 2003, 2004) kann daher unter anderem als eine Reaktion auf das wenig zufriedenstellende Abschneiden deutscher Jugendlicher bei PISA 2000 gesehen werden. Inzwischen existieren neben den für die PISA-Kohorte der Fünfzehnjährigen relevanten Bildungsstandards für die Sekundarstufe I (Hauptschulabschluss und Mittlerer Schulabschluss) im Fach Mathematik auch bundesweite Bildungsstandards für die Primarstufe sowie für die Sekundarstufe II (Allgemeine Hochschulreife).

Die Ergebnisse der ersten PISA-Studie beziehungsweise der folgenden Studien haben die in den „Konstanzer Beschlüssen“ der Kultusministerkonferenz von 1997 festgehaltene Entscheidung, regelmäßig an internationalen und nationalen *Large-Scale-Assessments* teilzunehmen, noch bekräftigt. Diese groß angelegten Schulleistungsstudien werden seither systematisch genutzt, um Erkenntnisse über den Stand unseres Bildungssystems zu gewinnen, wie sich besonders auch in langfristigen Projekten wie VERA oder dem Ländervergleich für die Bildungsstandards zeigt.

Als eine Konsequenz, die direkt der Verbesserung von Unterricht dienen sollte und konnte, sei das 1998 gestartete Modellversuchsprogramm zur „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“ (als SINUS bekannt) erwähnt. Es hatte zum Ziel, evidenzbasierte Wege zur Qualitätsentwicklung und Professionalisierung im Bereich des Mathematikunterrichts zu entwickeln, zu erproben und zu verbreiten (z. B. Prenzel, Friedrich & Stadler, 2009). Die erfolgreiche Arbeit wurde von den

Sekundarschulen auf die Grundschule übertragen, lief dort koordiniert bis 2013 und wird noch immer auf der Ebene einzelner Bundesländer weitergeführt.

PISA hat wie kaum eine andere Studie in Deutschland ein starkes öffentliches Interesse auf sich gezogen und bildungspolitischen Themen zu viel Aufmerksamkeit verholfen. Es ist nicht erstaunlich, dass Forderungen nach tiefgreifenden Veränderungen des Bildungssystems die Folge waren. Mit den Bildungsstandards ist ein wichtiger Schritt in Richtung Qualitätssicherung an Schulen gemacht worden.

In PISA 2003 war mathematische Kompetenz die Hauptdomäne und wurde in diesem Jahr zum ersten Mal grundlegend und unter Berücksichtigung von Teilkompetenzen erfasst und beschrieben. Seit diesem Zeitpunkt ist im Hinblick auf Schule und Unterricht Vieles geschehen, wovon man sich Verbesserungen erhoffen darf. So wurden Bildungsstandards in Mathematik für den Mittleren Schulabschluss verabschiedet und bundesweit eingeführt. Mit SINUS wurde ein länderübergreifendes Programm zur Verbesserung des mathematisch-naturwissenschaftlichen Unterrichts geschaffen, das erfolgreich unter einer großen und sehr aktiven Beteiligung von Lehrerinnen und Lehrern durchgeführt werden konnte. Vergleichsarbeiten werden inzwischen in den meisten Bundesländern flächendeckend geschrieben und erlauben eine Einordnung schulischer Leistungen. Schließlich kann man feststellen, dass seit 2003 nicht nur das Problembewusstsein, sondern auch die öffentliche Aufmerksamkeit für Bildung eine deutliche Steigerung erfahren hat. Die Entwicklung der mathematischen Kompetenz von Schülerinnen und Schülern in Deutschland seit der zweiten PISA-Studie sollte also besonders interessant und aufschlussreich sein. Auch wenn Bildungsprozesse sich nicht ohne Weiteres in einfache Ursache-Wirkungs-Zusammenhänge einpassen lassen, kann man vor diesem Hintergrund konkrete Fragen formulieren, deren Beantwortung eine vorsichtige Bilanz erlauben wird.

- Wie hat sich die mathematische Kompetenz von Fünfzehnjährigen in Deutschland im internationalen Vergleich entwickelt?
- Welche Streuung gibt es in den Leistungen?
- Wie haben sich die Anteile leistungsschwächerer beziehungsweise leistungsstärkerer Schülerinnen und Schüler entwickelt?
- Hat sich der Anteil der Fünfzehnjährigen erhöht, die zumindest ein akzeptables Niveau mathematischer Kompetenz zeigen?
- Gibt es weiterhin deutliche Unterschiede zwischen Jungen und Mädchen?
- Welche Entwicklungen kann man in Bezug auf die unterschiedlichen Schularten erkennen?

2.2 Mathematische Kompetenz in PISA

Im Vordergrund der PISA-Konzeption steht die mathematische Grundbildung; um sie messen zu können, muss sie in Aufgaben umgesetzt werden. Diese Aufgaben sind in *Kontexten* angesiedelt, haben einen *mathematischen Inhalt*, konzentrieren sich auf *mathematische Prozesse* und verlangen zur Lösung *fundamentale mathematische Fähigkeiten*.

Mathematische Kompetenz hat viele Facetten, die in einer einzigen Studie nicht alle Berücksichtigung finden können. Auch PISA misst nicht *mathematisches Wissen* oder *mathematische Kompetenz* im Allgemeinen, sondern beschränkt sich ganz bewusst auf einen Ausschnitt mit Relevanz für alle Jugendlichen – unabhängig von ihrem Bildungsgang. Bereits in der ersten PISA-Studie im Jahr 2000 wurde dabei zwischen der Beherrschung mathematischer Verfahren, der Kenntnis von Fakten beziehungsweise der schematischen Anwendung von Lösungsstrategien und der Fähigkeit der Jugendlichen, ihr mathematisches Wissen funktional, flexibel und reflektierend für die Lösung kontextbezogener Problemstellungen einzusetzen, unterschieden (OECD, 1999). Letzteres ist für PISA von zentraler Bedeutung, denn hier wird die wichtige Rolle von Mathematik gerade auch für das spätere Leben der Fünfzehnjährigen betont. Mathematik ist ein unverzichtbares Kultur- und Bildungsgut, sodass ein Test prädiktiv für die zukünftige Auseinandersetzung mit Mathematik und für das Lernen von Mathematik sein sollte. Es geht folglich um eine Testkonzeption, die mathematische Grundbildung so messen kann, dass ein differenziertes Bild über die Ausprägung fundamentaler mathematischer Fähigkeiten entsteht.

Seit Beginn der PISA-Erhebungen sieht sich die Studie dem Grundbildungskonzept verpflichtet, das in Kapitel 1 ausführlich beschrieben wurde. Was konkret unter *Mathematical Literacy* oder *mathematischer Grundbildung* verstanden wird, ist in der theoretischen Rahmenkonzeption für PISA 2012 festgehalten (OECD, 2013). Im Vordergrund steht dabei die Fähigkeit, mathematisches Wissen und mathematische Prozesse funktional auf Problemstellungen anzuwenden, die sowohl innermathematisch als auch in anderen Kontexten situiert sein können. *Mathematische Kompetenz* in diesem Sinne eröffnet jungen Menschen eine Möglichkeit, Herausforderungen im persönlichen, beruflichen und gesellschaftlichen Leben, aber auch im wissenschaftlichen Bereich zu verstehen, zu bearbeiten und somit zu meistern (OECD, 2013). Vor dem Hintergrund des Grundbildungskonzepts geht es in PISA auch darum, anhand der festgestellten Kompetenzen Bildungsergebnisse mit Blick auf die weitere Entwicklung der Schülerinnen und Schüler einzuordnen (OECD, 2013; Winter, 1995).

Mathematische Grundbildung (*Mathematical Literacy*), so wie sie in PISA 2012 zu Grunde gelegt wird, berücksichtigt unterschiedliche Komponenten. Die Definition unterscheidet drei zentrale Elemente, nämlich *Inhalte*, *Prozesse* und *Kontexte* (OECD, 2013). Konkret sind dabei mathematische Inhalte und Prozesse in Kontexten situiert, in denen die Fünfzehnjährigen ihre fachliche Kompetenz zeigen können. Die OECD definiert *Mathematical Literacy* entsprechend als „die Fähigkeit einer Person, Mathematik

in zahlreichen Kontexten anzuwenden, zu interpretieren und Formeln zu verwenden. Dazu gehört mathematisches Schlussfolgern ebenso wie die Anwendung mathematischer Konzepte, Vorgehensweisen, Fakten und Werkzeuge, um Phänomene zu beschreiben, zu erklären und vorherzusagen. Mathematische Grundbildung hilft Personen, die Rolle zu erkennen und zu verstehen, die Mathematik in der Welt spielt, fundierte mathematische Urteile abzugeben und Mathematik in einer Weise zu verwenden, die den Anforderungen des Lebens dieser Person als konstruktivem, engagiertem und reflektiertem Bürger entspricht“ (OECD, 2013, S. 25). Das Konzept der mathematischen Grundbildung fokussiert somit nicht einseitig Anwendungen von Mathematik im Alltag, sondern beinhaltet das Verstehen mathematischer Denk- und Arbeitsweisen und damit den bereits genannten Blick auf Mathematik als ein Kultur- und Bildungsgut.

Gegenüber PISA 2003 (OECD, 2004), 2006 und 2009 wurde diese Definition bei PISA 2012 erkennbar präzisiert, indem die Aspekte der Anwendung, Interpretation und der Verwendung von Formeln explizit aufgenommen wurden. Möglichkeiten und Grenzen von PISA werden auf diese Weise bereits in der Definition berücksichtigt, sie ist deutlicher und anschaulicher geworden. Insbesondere umfasst sie nun neben den genannten kognitiven Elementen auch Einstellungen (*Attitudes*, vgl. Kapitel 3) gegenüber der Mathematik, so dass sowohl kognitive als auch affektive Zielsetzungen berücksichtigt werden (OECD, 2013).

2.2.1 Elemente der Rahmenkonzeption: Inhalte, Prozesse und Kontexte

Zur Strukturierung der vielfältigen Facetten mathematischer Grundbildung dienen in PISA die bereits genannten Elemente *Inhalte*, *Prozesse* und *Kontexte*. Diese werden im Folgenden beschrieben.

Ebenso wie in PISA 2003 und den folgenden Zyklen sind auch in PISA 2012 vier Inhaltsbereiche mathematischer Kompetenz ausgewiesen:

- Veränderung und Beziehungen
- Raum und Form
- Quantität
- Unsicherheit und Daten

Damit ist das Strukturierungselement der Inhaltsbereiche gegenüber PISA 2003 quasi unverändert und folglich eine solide Basis für die Abbildung von Trends. Einzig beim vierten Inhaltsbereich, *Unsicherheit und Daten*, gibt es eine kleine Änderung in der Benennung. Er wurde in PISA 2003 noch mit „Unsicherheit“ bezeichnet. Die Ergänzung um den Begriff der Daten dient einer genaueren Beschreibung der abgedeckten Aspekte und impliziert keine grundlegenden inhaltlichen Veränderungen (OECD, 2013). Er deckt sich damit in der Bezeichnung auch mit der entsprechenden Leitidee in den Bildungsstandards für den Mittleren Schulabschluss der Kultusministerkonferenz (KMK,

2003). Im Vergleich zu den Standards fällt auf, dass der dort definierte Inhaltsbereich „Messen“ in PISA nicht explizit ausgewiesen ist. Ein Grund hierfür ist, dass er bei PISA auch als Aspekt der anderen Domänen (z. B. Naturwissenschaften) gesehen wird und zudem innerhalb mathematischer Inhaltsbereiche vorkommt. So ist Messen beispielsweise ein integraler Bestandteil des Bereichs *Raum und Form*. Darüber hinaus werden Messfehler im Rahmen von PISA explizit dem Bereich *Unsicherheit und Daten* zugeordnet.

Die vier Inhaltsbereiche werden in der theoretischen Rahmenkonzeption zu PISA 2012 mit drei mathematischen *Prozessen* verbunden, die grundlegende mathematische Aktivitäten konkretisieren (OECD, 2013):

- Situationen mathematisch formulieren;
- mathematische Konzepte, Fakten, Prozeduren und Schlussfolgerungen anwenden;
- mathematische Ergebnisse interpretieren, anwenden und bewerten.

Eine Situation mathematisch zu *formulieren*, bedeutet die Übertragung eines Problems in mathematische Strukturen und Repräsentationen (OECD, 2013). Dazu gehört beispielsweise das Erkennen von Regelmäßigkeiten, Beziehungen und Mustern in alltäglichen Situationen. Genauso kann es aber auch darum gehen, ein Alltagsproblem so in seinen grundlegenden Strukturen zu sehen, dass es mathematisch formuliert und somit analysiert werden kann. Die *Anwendung* mathematischer Konzepte, Fakten, Prozeduren und Schlussfolgerungen umfasst Arbeitsweisen zum erfolgreichen Bearbeiten mathematischer Fragestellungen. Das kann das Lösen einer Gleichung, das Ziehen logischer Schlussfolgerungen oder das Entnehmen mathematischer Informationen aus Tabellen und Abbildungen sein (OECD, 2013, S. 29). Mathematische Ergebnisse zu *interpretieren*, anzuwenden und zu bewerten, umfasst insbesondere die Reflexion über mathematische Lösungen, Ergebnisse und Schlussfolgerungen. Hier wird auf das Formulieren von Erklärungen und Argumenten innerhalb eines Problemkontexts abgezielt. Im Detail sind damit Tätigkeiten gemeint wie die Bewertung einer Lösung im Rahmen eines realen Problems oder die Benennung der Grenzen eines Modells, das zur Lösung eines Problems herangezogen wird (S. 29f.).

Das dritte Strukturierungselement in Bezug auf die Mathematik sind die *Kontexte*, in denen mathematische Probleme verortet werden. Kontexte (PISA 2003: Situationen) sind nicht als zufällige Einkleidungen zu sehen, sondern bestimmen wesentlich die Auswahl adäquater mathematischer Strategien. Sie beinhalten Elemente der Lebenswelt von Schülerinnen und Schülern, schaffen also eine Anbindung an ihren Alltag und dienen insbesondere dazu, verschiedene individuelle Interessen der Jugendlichen anzusprechen (OECD, 2013) und mathematische Anforderungen in einen relevanten Rahmen zu rücken. Deshalb wird mit vier Kontextkategorien eine möglichst breite Palette von Interessen abgedeckt: Unterschieden werden *persönliche*, *berufliche*, *gesellschaftliche* und *wissenschaftliche* Kontexte. Persönliche Kontexte betreffen das unmittelbare Umfeld der Jugendlichen, etwa ihre Person, ihre Familie, ihren Freundeskreis oder ihre Klasse.

Typische Themen für Aufgaben aus dem persönlichen Kontext handeln von Gesundheit, Sport und Spiel, Nahrungszubereitung und Einkaufen oder auch Finanzplanung. Berufliche Kontexte sind der Arbeitswelt zugeordnet und werden in Aufgaben etwa zu Messungen und Entscheidungsfindungen, zu Buchhaltung oder Architektur übertragen. Inhalte des dritten Kontexts, der Gesellschaft, konzentrieren sich auf soziale Gefüge als Teil des Lebensumfelds von Individuen. Diese können lokal, national oder auch global angelegt sein und Aspekte wie Wahlsysteme, öffentliche Verkehrsmittel, demographische Entwicklung oder Werbung beinhalten. Wissenschaftliche Kontexte beziehen die Anwendung von Mathematik beispielsweise auf naturwissenschaftliche und technologische Themen wie etwa Wetter und Klima, Medizin, Raumfahrt oder Genetik.

Schlussendlich dienen die Kontexte im Rahmen der Untersuchung bei PISA dazu, mathematische Kompetenz auf bestimmte lebensnahe Gegebenheiten zu beziehen, um sie für die Jugendlichen relevant zu machen.

Inhalte und Kontexte haben auch frühere Untersuchungen bestimmt. Im Vergleich zur theoretischen Rahmenkonzeption von PISA 2003 ist nun mit der expliziten Ausweisung der Prozesse ein Ordnungsfaktor geschaffen, der die Nutzung von Mathematik vermutlich leichter nachvollziehbar und präziser beschreibbar gemacht hat. Allerdings gibt es auch hier die von anderen Bereichen bereits bekannte Schwierigkeit, die Prozesse wirklich trennscharf in Aufgabenstellungen umzusetzen.

2.2.2 Fundamentale mathematische Fähigkeiten

Ein Kernelement in PISA 2003 waren acht mathematische Teilkompetenzen, womit mehr oder minder typische Arbeitsweisen gemeint waren, die beim Lösen von Aufgaben benötigt werden. Beispiele sind Argumentieren, Kommunizieren, Modellieren oder auch der Umgang mit symbolischen, formalen und technischen Elementen (Niss, 2003; Niss & Højgaard, 2011). In PISA 2012 wurden sie leicht modifiziert und auf sieben sogenannte *fundamentale mathematische Fähigkeiten* („capabilities“) verdichtet (OECD, 2013).

Die fundamentalen mathematischen Fähigkeiten sind von zentraler Bedeutung für PISA. Grundsätzlich wird davon ausgegangen, dass ein Individuum über sie verfügt oder sie lernend entwickeln kann. Entsprechend sollte es jedem Menschen prinzipiell möglich sein, sich auf mathematische Weise mit seiner Umwelt auseinanderzusetzen und geeignete Probleme zu lösen. Je weiter die mathematische Grundbildung bei einem Menschen entwickelt ist, umso besser sollte dieser Mensch in der Lage sein, die sieben fundamentalen mathematischen Fähigkeiten beim Lösen eines Problems einzusetzen (Turner & Adams, 2012). Das problemgerechte Nutzen dieser fundamentalen Fähigkeiten, das mit Items zunehmender Schwierigkeitsstufen herausgefordert wird, dient später zur Differenzierung hierarchischer Kompetenzstufen (vgl. Abschnitt 2.3.2). Es sind die folgenden sieben grundlegenden mathematischen Fähigkeiten, die in der Rahmenkonzeption von PISA explizit formuliert sind:

- kommunizieren
- mathematisieren
- repräsentieren
- argumentieren
- Problemlösestrategien entwickeln
- mit Mathematik symbolisch, formal und technisch umgehen
- mathematische Hilfsmittel verwenden

In der Rahmenkonzeption werden diese Fähigkeiten im Einzelnen charakterisiert (OECD, 2013). *Kommunizieren* ist danach eine fundamentale mathematische Fähigkeit, die das Erkennen und Verstehen einer Problemsituation voraussetzt. Sie beinhaltet das Festhalten und Beschreiben von Zwischenschritten auf dem Weg zu einer Lösung und das Erklären und Begründen nach dem Erreichen der Lösung. Unter *Mathematisieren* wird verstanden, dass die Fünfzehnjährigen in der Lage sind, ein Realproblem in eine geeignete mathematische Form zu übersetzen. Dazu gehört unter anderem das Strukturieren einer Situation, das Formulieren von Annahmen oder das Aufstellen und Überprüfen eines Modells. Mathematisieren ist damit ein Teilaspekt des mathematischen Modellierens. *Repräsentieren* ist durch den Umgang mit Abbildungen, Tabellen, Diagrammen, aber auch mit Formeln und Gleichungen bestimmt. Damit ist diese grundlegende Fähigkeit eng mit den beiden zuvor beschriebenen Fähigkeiten (Kommunizieren und Mathematisieren) verknüpft, da die Jugendlichen auch hier zwischen Realsituationen und mathematischen Darstellungsformen vermitteln und diesen Prozess für andere Personen nachvollziehbar machen müssen. *Argumentieren* als vierte fundamentale Fähigkeit bedeutet, logische Denkprozesse durchzuführen, Elemente von Problemstellungen miteinander zu verbinden und daraus Schlussfolgerungen zu ziehen. Wie beim Kommunizieren sind auch hier Begründungen gefordert. Die Fähigkeit, *Problemlösestrategien zu entwickeln*, meint ein planvolles Vorgehen bei der Bearbeitung von Problemstellungen. Insbesondere geht es dabei um die laufende Überprüfung des Problemlöseprozesses und gegebenenfalls eine Anpassung der Lösungsstrategie. *Symbolisch, formal und technisch mit Mathematik umzugehen* umfasst den regelkonformen Umgang mit symbolischen Ausdrücken in einem mathematischen Kontext sowie die Verwendung formaler Konstrukte, welche mathematischen Definitionen, Regeln und formalen Systemen folgen. Die siebte fundamentale Fähigkeit, *mathematische Hilfsmittel verwenden*, zielt schließlich auf den Einsatz von Werkzeugen ab. Das können beispielsweise Messinstrumente, der Taschenrechner oder auch computergestützte Hilfsmittel sein. Die Reflexion über Möglichkeiten und Grenzen solcher Werkzeuge ist ebenfalls Teil dieser grundlegenden mathematischen Fähigkeit.

Anhand der inhaltlichen Beschreibung der fundamentalen mathematischen Fähigkeiten wird deutlich, dass sie oftmals miteinander verzahnt sind und sich deshalb nicht trennscharf voneinander abgrenzen lassen (vgl. auch Rahmenkonzeption PISA 2000). Gerade das Kommunizieren ist Teil mehrerer anderer Fähigkeiten beziehungsweise wird benötigt, um die anderen Fähigkeiten überhaupt erst sichtbar zu machen. Dennoch wird

Tabelle 2.1: Zusammenhang zwischen mathematischen Prozessen und fundamentalen mathematischen Fähigkeiten

	Situationen mathematisch formulieren	Mathematische Konzepte, Fakten, Prozeduren und Schlussfolgerungen anwenden	Mathematische Ergebnisse interpretieren, anwenden und bewerten
<i>Kommunizieren</i>	Angaben, Fragen, Aufgaben, Objekte, Bilder oder Animationen (bei computerbasierter Erhebung) lesen, verstehen und zusammenfügen, um ein mentales Modell der Situation zu erstellen	Eine Lösung verständlich darstellen, den Lösungsweg aufzeigen und/oder mathematische Zwischenergebnisse zusammenfassen und präsentieren	Erklärungen und Argumente im Kontext des Problems darstellen und kommunizieren
<i>Mathematisieren</i>	Mathematische Variablen und Strukturen eines Problems in einer Realsituation identifizieren und Annahmen bilden, damit diese genutzt werden können	Ein Verständnis des Kontexts anwenden, um den mathematischen Problemlöseprozess zu lenken oder zu beschleunigen, zum Beispiel auf einer Genauigkeitsebene arbeiten, die dem Kontext entspricht	Den Umfang und die Grenzen einer mathematischen Lösung verstehen, die eine Konsequenz des angewandten mathematischen Modells sind
<i>Repräsentieren</i>	Eine mathematische Repräsentation einer realen Information erstellen	Bei der Interaktion mit einem Problem eine Vielzahl an Repräsentationen zusammenfügen, in Beziehung setzen und anwenden	Mathematische Ergebnisse in einer Vielzahl an Formaten in Relation zu einer Situation oder einer Verwendung interpretieren; zwei oder mehr Repräsentationen in Relation zu einer Situation vergleichen und bewerten
<i>Argumentieren</i>	Eine Begründung für die erkannte oder ausgedachte Repräsentation einer Realsituation darlegen, verteidigen oder liefern	Eine Begründung für die Vorgänge und Abläufe, die für die Festlegung eines mathematischen Ergebnisses oder einer Lösung verwendet wurden, darlegen, verteidigen oder liefern; Daten verknüpfen, um eine mathematische Lösung zu erreichen, Verallgemeinerungen zu bilden oder einen mehrstufigen Beweis zu liefern	Mathematische Lösungen überlegen und Erklärungen und Beweise schaffen, die eine mathematische Lösung für ein kontextualisiertes Problem unterstützen, widerlegen oder relativieren
<i>Problemlösestrategien entwickeln</i>	Einen Plan oder eine Strategie auswählen oder entwickeln, um kontextualisierte Probleme mathematisch neu auszurichten	Wirksame und anhaltende Kontrollmechanismen durch ein mehrstufiges Verfahren, das zu einer mathematischen Lösung, Schlussfolgerung oder Verallgemeinerung führt, aktivieren	Eine Vorgehensweise entwickeln und implementieren, um eine mathematische Lösung zu einem kontextualisierten Problem zu interpretieren, bewerten oder überprüfen
<i>Mit Mathematik symbolisch, formal und technisch umgehen</i>	Entsprechende Variablen, Symbole, Diagramme und standardisierte Modelle verwenden, um eine Realsituation durch symbolische/formale Sprache zu repräsentieren	Formale Konstrukte, welche auf Definitionen, Regeln und formalen Systemen sowie auf verwendeten Algorithmen basieren, verstehen und nutzen	Die Beziehung zwischen Problemkontext und der Repräsentation der mathematischen Lösung verstehen. Dieses Verständnis anwenden, um die Lösung im Kontext zu interpretieren und die Durchführbarkeit und mögliche Einschränkungen der Lösung abzuschätzen
<i>Mathematische Hilfsmittel verwenden</i>	Mathematische Hilfsmittel verwenden, um mathematische Strukturen wiederzuerkennen oder mathematische Beziehungen darzustellen	Unterschiedliche Hilfsmittel kennen, die bei der Realisierung von Vorgängen und Abläufen zur Festlegung mathematischer Lösungen unterstützen können, und imstande sein, sie angemessen einzusetzen	Mathematische Hilfsmittel verwenden, um die Annehmbarkeit einer mathematischen Lösung und jegliche Grenzen und Beschränkungen dieser Lösung bezogen auf den Problemkontext festzustellen

in der Rahmenkonzeption der Versuch unternommen, wesentliche Charakteristika herauszustellen, indem sie konkret mit den Prozessen (formulieren, anwenden und interpretieren) verbunden werden. Tabelle 2.1 veranschaulicht diese Verknüpfung.

Diese rasterhafte Darstellung spiegelt wesentliche Elemente für die Entwicklung von Aufgaben wider, so wie sie in PISA genutzt werden. Sie ist aber keinesfalls dafür gedacht, lückenlos alle Elemente mathematischer Kompetenz zusammenzutragen. Dies wird besonders deutlich in der Darstellung der wesentlichen Elemente der Rahmenkonzeption von PISA 2012 (OECD, 2013). Graphisch werden diese durch den sogenannten mathematischen Modellierungskreislauf (Abbildung 2.1) veranschaulicht.

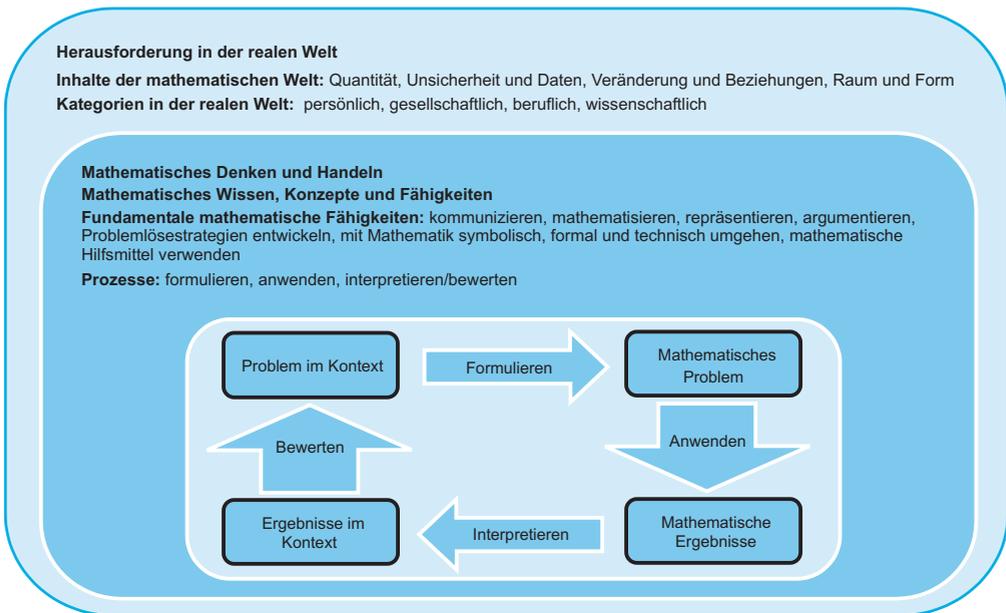


Abbildung 2.1: Mathematischer Modellierungskreislauf (OECD, 2013, S. 26)

Der äußere Kasten verdeutlicht, dass sich mathematische Grundbildung im Sinne von PISA bei der Lösung eines Problems zeigt, das in der realen Welt entsteht. Es ist dort einem bestimmten Kontext und in der Mathematik einem Inhaltsbereich zugeordnet. Der mittlere Kasten zeigt, wie mathematisches Denken und Handeln über die fundamentalen Fähigkeiten zur Problemlösung genutzt werden kann. Der innere Kasten beschreibt schließlich den Prozess, den ein Problemlöser im Hinblick auf eine Lösung durchlaufen wird (vgl. auch de Lange, 1987; Blum, 2002).

Mathematische Kompetenz, so wie sie in PISA verstanden wird, stellt damit einen Bezug zwischen Herausforderungen in der realen Welt und mathematischem Denken und Handeln her. Konkret bedeutet dies, Situationen in eine mathematische Sprache zu übertragen (Blum, Drücke-Noe, Hartung & Köller, 2010) und Probleme mit Hilfe mathematischer Ansätze erfolgreich zu lösen (OECD, 2013).

Mit der theoretischen Rahmenkonzeption für Mathematik in PISA 2012 liegt ein Basisdokument vor, das die Konzeption von PISA 2003 aufnimmt und dabei präzisiert. Die überarbeitete Rahmenkonzeption liefert ein besseres Verständnis für die Grundlage der in PISA untersuchten mathematischen Grundbildung und der damit verbundenen mathematischen Kompetenz. Was unter mathematischer Kompetenz zu verstehen ist, ist aber im Kern genauso unverändert geblieben wie die Definition der Kompetenzstufen.

Einschub: Die Bildungsstandards für den Mittleren Schulabschluss (KMK, 2003)

PISA hat die deutschen Bildungsstandards im Fach Mathematik etwa für den Mittleren Schulabschluss nicht unerheblich beeinflusst. Es gibt Gemeinsamkeiten, aber es gibt auch ganz wesentliche Unterschiede zwischen den Bildungsstandards und den Grundlagen der PISA-Testkonstruktion. So orientieren sich die Bildungsstandards für Mathematik an einer Differenzierung in inhaltsbezogene und allgemeine mathematische Kompetenzen, die drei Anforderungsbereichen zugeordnet werden. Die inhaltsbezogenen Kompetenzen werden in Form von fünf Leitideen beschrieben, die fachliche Bereiche widerspiegeln (*Zahl, Messen, Raum und Form, Funktionaler Zusammenhang, Daten und Zufall*; vgl. auch KMK, 2003). Die allgemeinen Kompetenzen beschreiben Formen des mathematischen Arbeitens (*argumentieren, problemlösen, modellieren, verwenden von Darstellungen, Umgang mit formalen, symbolischen und technischen Elementen, kommunizieren*; vgl. auch KMK, 2003). Schließlich gibt es unterschiedliche Anforderungen bei der Bearbeitung von Aufgaben, nämlich *Reproduzieren, Zusammenhänge Herstellen, Verallgemeinern und Reflektieren*. Damit sind mathematische Inhalte und Prozesse der Ausgangspunkt für kompetenzorientiertes Arbeiten in der Mathematik. Die Inhalte entsprechen in etwa denen, die in PISA 2012 verwendet werden, und was dort „fundamentale mathematische Fähigkeiten“ sind, ist im Wesentlichen mit den „allgemeinen mathematischen Kompetenzen“ vergleichbar. Allerdings werden in den Bildungsstandards weder die Prozesse (*formulieren, anwenden, interpretieren*) noch die Kontexte (*persönlich, beruflich, gesellschaftlich, wissenschaftlich*) explizit erwähnt.

2.3 Der Mathematiktest in PISA 2012 und seine Auswertung

In PISA 2012 wird *Mathematical Literacy* in ihrer Ausdifferenzierung über Inhalte, Prozesse und Kontexte durch eine Vielzahl von Testaufgaben abgebildet. Bei den insgesamt 110 Aufgaben für den Bereich Mathematik handelt es sich einerseits um 36 sogenannte Link-Items (vgl. Kapitel 10), also Aufgaben, die bereits in früheren PISA-Erhebungsrounden zum Zuge gekommen waren, und andererseits 74 neu entwickelte Aufgaben. Dabei werden sowohl die vier Inhaltsbereiche und drei Prozesse als auch die vier Kontextkategorien durch Items abgedeckt.

Tabelle 2.2: Anteile der Aufgaben nach Inhalten, Prozessen und Kontexten an der maximal möglichen Punktzahl für Mathematik

Kategorie	Anteil an der maximal möglichen Punktzahl (Prozent)
<i>Inhalte</i>	
Veränderung und Beziehungen	ca. 25
Raum und Form	ca. 25
Quantität	ca. 25
Unsicherheit und Daten	ca. 25
GESAMT	100
<i>Prozesse</i>	
Situationen mathematisch formulieren	ca. 25
Mathematische Konzepte, Fakten, Prozeduren und Überlegungen anwenden	ca. 50
Mathematische Ergebnisse interpretieren, anwenden und bewerten	ca. 25
GESAMT	100
<i>Kontexte</i>	
Persönlich	ca. 25
Beruflich	ca. 25
Gesellschaftlich	ca. 25
Wissenschaftlich	ca. 25
GESAMT	100

In der vorgegebenen Testzeit können die Schülerinnen und Schüler jeweils nur eine Auswahl von Testaufgaben bearbeiten. PISA arbeitet daher mit einem sogenannten Rotationsdesign, bei dem diese Auswahl nach einem systematischen Schlüssel erfolgt. Der „Pool“ an Mathematikaufgaben in PISA ist dabei so angelegt, dass sich die Aufgaben angemessen auf die unterschiedlichen Inhalte, Prozesse und Kontexte verteilen und so die Rahmenkonzeption für mathematische Grundbildung widerspiegeln. Tabelle 2.2 zeigt die Zusammensetzung der zu erreichenden Punktzahlen nach Inhalten, Prozessen und Kontexten.

Das Spektrum an Inhalten, Prozessen und Kontexten, welche durch die Mathematikaufgaben in PISA 2012 abgedeckt werden, wird durch eine gezielte Auswahl der Schwierigkeit von Aufgaben ergänzt. Aus psychometrischer Sicht ist ein *Large-Scale-Assessment*, das die Kompetenz einer bestimmten Alterskohorte messen soll, dann am effektivsten, wenn die Verteilung der Schwierigkeiten der Testaufgaben möglichst genau den Kompetenzen der Studienteilnehmer entspricht (OECD, 2013). Das bedeutet insbesondere, dass die einfachsten und die schwierigsten Aufgaben ein Spektrum eingrenzen, das der Kompetenzspanne der beteiligten Jugendlichen gerecht wird. Auf dieser Basis sind auch die Kompetenzstufen (vgl. Abschnitt 2.3.2) angelegt, die eine ansteigende Ausprägung der fundamentalen mathematischen Fähigkeiten abbilden.

2.3.1 Aufgabenformate

Für die Mathematik gab es in PISA 2012 drei Arten von Aufgabenformaten, nämlich offene Aufgaben, geschlossene Aufgaben und Multiple-Choice-Aufgaben (OECD, 2013). Bei *offenen Aufgaben* sind die Schülerinnen und Schüler aufgefordert, selbst Antworten zu formulieren, die relativ ausführlich sein können. Ein einzelnes Wort oder eine Zahl genügen hier nicht für eine vollständig richtige Lösung; vielmehr sollen die Jugendlichen beispielsweise ihren Lösungsweg darstellen oder begründen, wie sie zu ihrer Antwort gekommen sind. Solche Aufgaben müssen von entsprechend geschulten Expertinnen und Experten kodiert werden, das heißt, die Entscheidung darüber, ob eine Antwort richtig, falsch oder teilweise richtig ist, muss Fall für Fall auf der Basis einer soliden mathematischen Fachexpertise getroffen werden. Im Vergleich dazu sind *geschlossene Aufgaben* stärker auf bestimmte kurze Antworten der Schülerinnen und Schüler hin strukturiert, sodass die Entscheidung, ob eine Antwort richtig oder falsch ist, hier wesentlich schneller getroffen werden kann und teilweise automatisiert durchgeführt wird. Eine typische Antwort auf eine geschlossene Frage wäre etwa ein Zahlenwert oder ein einzelnes Wort. Das dritte in PISA eingesetzte Aufgabenformat sind sogenannte *Multiple-Choice-Fragen*, in denen aus mehreren vorformulierten Antworten jeweils eine oder mehrere korrekte Aussagen ausgewählt werden müssen. Jedes der drei Aufgabenformate ist im PISA-2012-Mathematiktest etwa bei einem Drittel der Aufgaben eingesetzt worden (OECD, 2013).

2.3.2 Skalierung und Kompetenzstufen

Der Schwierigkeitsgrad einer Mathematikaufgabe ist in der Regel keine feste Größe, sondern kann von unterschiedlichsten Faktoren wie etwa dem Vorwissen und den Problemlösefähigkeiten des Bearbeiters oder dem gewählten Kontext beeinflusst werden. Allerdings gibt es einzelne Merkmale von Aufgaben, die allgemein als schwierigkeitsgenerierend angesehen werden. So unterscheiden die Bildungsstandards die Anforderungsbereiche *Reproduzieren*, *Zusammenhänge Herstellen* sowie *Verallgemeinern und Reflektieren*, weil es im Allgemeinen einfacher sein wird, eine Lösung zu reproduzieren als sie zu begründen. Doch lässt sich daraus keine Regel zur Bestimmung eines Schwierigkeitsgrades ableiten. Jede Lehrkraft weiß, dass man sich bei solchen Zuordnungen durchaus täuschen kann und die Bildungsforschung konnte dies bestätigen (Baumert et al., 1997). Entsprechend ist die Stufung von Anforderungen im Prinzip nur auf einer empirischen Grundlage möglich, die konkrete Lösungsraten einbezieht, um den Schwierigkeitsgrad einer Mathematikaufgabe zu bestimmen. In PISA wird dies, wie in umfangreichen Schulleistungstudien üblich, über eine Skalierung der Aufgaben nach der *Item Response Theory* (IRT) gehandhabt. Genauer ist in Kapitel 10 beschrieben.

Konkret wird man eine Aufgabe als eher leicht ansehen, wenn sie von sehr vielen Schülerinnen und Schülern richtig gelöst wird. Diese empirische Bedingung für die Aufgabenschwierigkeit soll zu den mathematischen Kompetenzanforderungen der Aufgaben passen, so wie sie in der theoretischen Rahmenkonzeption jeder PISA-Erhebungsrunde adressiert werden. Eine Person, die hohe mathematische Fähigkeiten hat, sollte Mathematikaufgaben mit einer höheren Wahrscheinlichkeit lösen als es Personen mit geringeren Fähigkeiten gelingen wird. Dieser im Grunde einfache Gedanke ist die Basis der Messung mathematischer Grundbildung in PISA auf einer eindimensionalen Kompetenzskala. Personen und Aufgaben, oder genauer Personenfähigkeit und Aufgabenschwierigkeiten, werden so auf einer gemeinsamen Skala abgebildet, dass Personen auf einem bestimmten Fähigkeitsniveau Aufgaben auf genau diesem Schwierigkeitsniveau mit einer gewissen, angemessen hohen Wahrscheinlichkeit lösen.

Betrachtet man nun die Aufgaben, die in etwa einem ähnlichen Fähigkeitsniveau zuzuordnen sind, so kann der Zusammenhang zwischen diesem Niveau und den Aufgabenmerkmalen untersucht werden. Dieser Zusammenhang dient als Grundlage für die Beschreibung von Kompetenzstufen der mathematischen Grundbildung. Bereits bei PISA 2003, als die Stufen der mathematischen Kompetenz international definiert wurden, hat sich dies als sinnvoll erwiesen. Es wurden dabei sechs Kompetenzstufen ausgewiesen und verbal beschrieben (OECD, 2004; Blum et al., 2004). Auf der Basis der theoretischen Rahmenkonzeption der OECD für die Mathematik in PISA 2012 und damit gegenüber PISA 2003 unverändert, beinhalten die sechs Kompetenzstufen der Mathematik die in Tabelle 2.3 beschriebenen Anforderungen.

Bereits in PISA 2000 und entsprechend auch in allen Folgestudien wurden Fünfzehnjährige, die in ihrer Mathematikleistung nicht über die Kompetenzstufe I hinaus kamen, als „Risikogruppe“ angesehen (Baumert et al., 2001). Diese Einschätzung wird auch in der vorliegenden Untersuchung beibehalten. Diese Jugendlichen verfügen allenfalls über ein elementares mathematisches Wissen und sind weit vom Erreichen der Ziele entfernt, die mit dem Mathematikunterricht in der Sekundarstufe verbunden sind (KMK, 2003, 2004). Das niedrige Kompetenzniveau lässt fraglos Schwierigkeiten bei ihrer weiteren schulischen oder beruflichen Bildung erwarten.

2.3.3 Aufgabenbeispiele

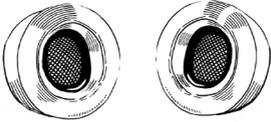
Welche mathematischen Kompetenzen bei der Lösung von PISA-Aufgaben erforderlich sind, lässt sich am besten anhand von konkreten Aufgabenstellungen beschreiben. Die folgenden Beispiele zeigen diese fachlichen Anforderungen. Sie verdeutlichen dabei insbesondere die Anwendungsorientierung des PISA-Mathematiktests und zeigen auch seine Verbindung mit den Bildungsstandards. Auch wenn die PISA-Aufgaben von der Testkonzeption her nicht darauf abzielen, die deutschen Bildungsstandards für den Mittleren Schulabschluss zu messen, so finden sich bei einzelnen Aufgaben gleichwohl

Tabelle 2.3: Stufen mathematischer Kompetenz in PISA 2012

Kompetenzstufe	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
VI > 669 Punkte	Schülerinnen und Schüler auf dieser Stufe können Informationen, die sie aus der Untersuchung und Modellierung komplexer Problemsituationen erhalten, konzeptualisieren, verallgemeinern und auf neue Situationen anwenden. Sie können verschiedene Informationsquellen und Darstellungen miteinander verknüpfen und flexibel zwischen diesen hin und her wechseln. Schülerinnen und Schüler auf dieser Stufe besitzen die Fähigkeit zu anspruchsvollem mathematischem Denken und Argumentieren. Sie können dieses mathematische Verständnis und ihre Beherrschung symbolischer und formaler mathematischer Operationen und Beziehungen nutzen, um Ansätze und Strategien zum Umgang mit neuartigen Problemsituationen zu entwickeln. Schülerinnen und Schüler auf dieser Stufe können ihr Tun und ihre Überlegungen, die zu ihren Erkenntnissen, Interpretationen und Argumentationen geführt haben, präzise beschreiben und kommunizieren, einschließlich der Beurteilung von deren Angemessenheit für die jeweilige Ausgangssituation.
V 607–668 Punkte	Schülerinnen und Schüler auf dieser Stufe können Modelle für komplexe Situationen konzipieren und mit ihnen arbeiten, einschränkende Bedingungen identifizieren und Annahmen spezifizieren. Sie können im Zusammenhang mit diesen Modellen geeignete Strategien für die Lösung komplexer Probleme auswählen, sie miteinander vergleichen und bewerten. Schülerinnen und Schüler auf dieser Stufe können strategisch vorgehen, indem sie sich auf breit gefächerte, gut entwickelte Denk- und Argumentationsfähigkeiten, passende Darstellungen, symbolische und formale Beschreibungen und für diese Situationen relevante Einsichten stützen. Sie sind imstande, über ihr Tun zu reflektieren und ihre Interpretationen und Überlegungen zu formulieren und zu kommunizieren.
IV 545–606 Punkte	Schülerinnen und Schüler auf dieser Stufe können effektiv mit expliziten Modellen komplexer konkreter Situationen arbeiten, auch wenn sie einschränkende Bedingungen enthalten oder die Aufstellung von Annahmen erfordern. Sie können verschiedene Darstellungsformen, darunter auch symbolische, auswählen und zusammenführen, indem sie sie direkt zu Aspekten von Realsituationen in Beziehung setzen. Schülerinnen und Schüler auf dieser Stufe können in diesen Kontexten gut ausgebildete Fertigkeiten anwenden und mit einem gewissen mathematischen Verständnis flexibel argumentieren. Sie können Erklärungen und Begründungen für ihre Interpretationen, Argumentationen und Handlungen geben und sie anderen mitteilen.
III 483–544 Punkte	Schülerinnen und Schüler auf dieser Stufe können klar beschriebene Verfahren durchführen, auch solche, die sequenzielle Entscheidungen erfordern. Sie können einfache Problemlösungsstrategien auswählen und anwenden. Schülerinnen und Schüler auf dieser Stufe können Darstellungen interpretieren und nutzen, die aus verschiedenen Informationsquellen stammen, und hieraus unmittelbare Schlüsse ableiten. Sie können kurze Berichte zu ihren Interpretationen, Ergebnissen und Überlegungen geben.
II 421–482 Punkte	Schülerinnen und Schüler auf dieser Stufe können Situationen in Kontexten interpretieren und erkennen, die nicht mehr als direkte Schlussfolgerungen erfordern. Sie können relevante Informationen einer einzigen Quelle entnehmen und eine einzige Darstellungsform benutzen. Schülerinnen und Schüler auf dieser Stufe können elementare Algorithmen, Formeln, Verfahren oder Regeln anwenden. Sie sind zu direkten Schlussfolgerungen und wörtlichen Interpretationen der Ergebnisse imstande.
I 358–420 Punkte	Schülerinnen und Schüler auf dieser Stufe können auf Fragen zu vertrauten Kontexten antworten, bei denen alle relevanten Informationen gegeben und die Fragen klar definiert sind. Sie können Informationen identifizieren und Routineverfahren gemäß direkten Instruktionen in expliziten Situationen anwenden. Sie können Handlungen ausführen, die klar ersichtlich sind und sich unmittelbar aus den jeweiligen Situationen ergeben.
unter I < 358 Punkte	

gemeinsame Aspekte, die im Folgenden herausgearbeitet werden. Dadurch wird noch einmal deutlich, inwieweit der Kompetenzbegriff in PISA und die curriculare Nähe der Standards zwar von unterschiedlichen Perspektiven her gedacht sind, aber inhaltliche Aspekte teilen (vgl. Hartig & Frey, 2012, die eine sehr hohe Korrelation zwischen Mathematiktests bei PISA und den Bildungsstandards berichten).

MP3-SPIELER

Music City MP3 Specialists		
<p>MP3 player</p>  <div style="background-color: #cccccc; padding: 5px; margin-top: 10px; text-align: center; font-weight: bold;">155 zeds</div>	<p>Headphones</p>  <div style="background-color: #cccccc; padding: 5px; margin-top: 10px; text-align: center; font-weight: bold;">86 zeds</div>	<p>Speakers</p>  <div style="background-color: #cccccc; padding: 5px; margin-top: 10px; text-align: center; font-weight: bold;">79 zeds</div>

FRAGE 1: MP3-SPIELER

Olivia hat die Preise für den MP3-Spieler, die Kopfhörer und die Lautsprecher mit dem Taschenrechner addiert.

Ihr Ergebnis war 248.



Olivias Ergebnis ist falsch. Sie hat einen der folgenden Fehler gemacht. Welchen Fehler hat sie gemacht?

- A Sie hat einen der Preise zweimal dazu addiert.
- B Sie hat vergessen, einen der drei Preise einzugeben.
- C Sie hat bei einem der drei Preise die letzte Ziffer weggelassen.
- D Sie hat einen der Preise subtrahiert, anstatt ihn zu addieren.

Abbildung 2.2: Beispielaufgabe „MP3-Spieler“

Das Multiple-Choice-Item aus der Mathematik-Unit *MP3-Spieler* adressiert ein alltägliches Problem, nämlich den nicht korrekten Gebrauch eines Taschenrechners. Drei Zahlen sollen addiert werden, das Ergebnis stimmt nicht und wird kritisch hinterfragt.

Die Aufgabe ist dem Inhaltsbereich *Quantität* entnommen, bewegt sich in einem persönlichen Kontext und erfordert einfaches Anwenden mathematischen Wissens. Man kann auf unterschiedlichen Wegen zur Lösung kommen, doch bietet sich auf jeden Fall Abschätzen und Überschlagen an. Addiert man beispielsweise die drei Zahlen, dann ist das Ergebnis $320 = 155 + 86 + 79$. Weil 248 kleiner als 320 ist, kann Lösung A nicht stimmen und auch B fällt sofort heraus. Antwort D kann man ausschließen, indem man die möglichen Ergebnisse abschätzt, sodass C die Lösung des Items ist.

Überschlagsrechnungen zur Kontrolle von Ergebnissen sind ein wichtiges Instrument der täglichen Arbeit. In den Bildungsstandards sind sie im Rahmen der Leitidee *Zahl* genannt. Offensichtlich stellen sie für die Fünfzehnjährigen in dieser einfachen Form keine große Hürde dar. Die Aufgabe liegt auf Kompetenzstufe I mit einem Schwierigkeitsindex von 398 Punkten.

DIE RADFAHRERIN HEIKE (H)



Heike hat gerade ein neues Fahrrad bekommen. Es hat einen Tachometer, der am Lenker befestigt ist.

Der Tachometer kann Heike die zurückgelegte Strecke und ihre Durchschnittsgeschwindigkeit für eine Tour anzeigen.

FRAGE 1: HEIKE DIE RADFAHRERIN

Auf einer Tour ist Heike 4 km in den ersten 10 Minuten gefahren und dann 2 km in den nächsten 5 Minuten.

Welche der folgenden Aussagen ist richtig?

- A Heikes Durchschnittsgeschwindigkeit war in den ersten 10 Minuten größer als in den nächsten 5 Minuten.
- B Heikes Durchschnittsgeschwindigkeit war in den ersten 10 Minuten und in den nächsten 5 Minuten die gleiche.
- C Heikes Durchschnittsgeschwindigkeit war in den ersten 10 Minuten niedriger als in den nächsten 5 Minuten.
- D Es ist nicht möglich, anhand der Angaben etwas über Heikes Durchschnittsgeschwindigkeit zu sagen.

FRAGE 2: HEIKE DIE RADFAHRERIN

Heike ist 6 km zum Haus ihrer Tante gefahren. Ihr Tachometer hat für die gesamte Tour durchschnittlich 18 km/h angezeigt.

Welche der folgenden Aussagen ist richtig?

- A Heike hat 20 Minuten gebraucht, um zum Haus ihrer Tante zu kommen.
- B Heike hat 30 Minuten gebraucht, um zum Haus ihrer Tante zu kommen.
- C Heike hat 3 Stunden gebraucht, um zum Haus ihrer Tante zu kommen.
- D Es ist nicht möglich zu sagen, wie lange Heike gebraucht hat, um zum Haus ihrer Tante zu kommen.

FRAGE 3: HEIKE DIE RADFAHRERIN

Heike ist mit ihrem Fahrrad von zuhause zum Fluss gefahren, der 4 km entfernt ist. Dafür hat sie 9 Minuten gebraucht. Nach Hause ist sie auf einer kürzeren Strecke von 3 km gefahren. Dafür hat sie nur 6 Minuten gebraucht.

Wie groß war Heikes Durchschnittsgeschwindigkeit in km/h auf der Tour zum Fluss und wieder zurück?

Durchschnittsgeschwindigkeit für die Tour:km/h

Abbildung 2.3: Beispielaufgabe „Die Radfahrerin Heike“

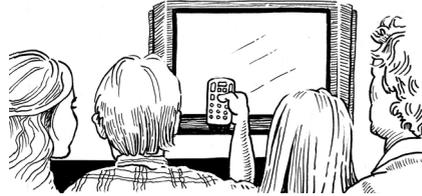
Die Aufgabe (Unit) *Die Radfahrerin Heike* (inklusive aller Items) trifft in besonderer Weise den Erfahrungsbereich fünfzehnjähriger Schülerinnen und Schüler: Es geht um den Tachometer am Fahrrad und Möglichkeiten, damit Strecken und Zeiten beziehungsweise Geschwindigkeiten zu messen. Alle drei Teile dieser Aufgaben gehören zum Inhaltsbereich *Veränderungen und Beziehungen*, zum persönlichen Kontext und zum Prozess des Anwendens. Die ersten beiden Items haben zudem einen starken Bezug zum Prozess des Interpretierens.

Die erste Frage ist einfach zu lösen und lässt sich mit einem Schwierigkeitsindex von 441 empirisch der Kompetenzstufe II zuordnen. Es müssen zwei Angaben zu Zeiten (10 Minuten bzw. 5 Minuten) und der dabei jeweils zurückgelegte Weg (4 km bzw. 2 km) miteinander verglichen werden, aus vier Antwortmöglichkeiten ist die richtige Antwort zu wählen. Weil die Zahlenwerte einfach sind, stellen sie keine besonderen Anforderungen an den Problemlöser dar. Voraussetzung für die Auswahl der korrekten Lösung B ist allerdings, dass das Konzept der Proportionalität in der gegebenen Situation erkannt und angewendet wird. Auch die zweite Frage ist im Multiple-Choice-Verfahren und mit A zu beantworten. Gegeben sind hier eine Strecke und die Durchschnittsgeschwindigkeit auf dieser Strecke, zu berechnen ist die Zeit, die dafür benötigt wird. Sie stellt rechnerisch keine hohen Anforderungen, setzt aber ein Verständnis des Begriffs der Durchschnittsgeschwindigkeit voraus. Empirisch ist sie mit einem Schwierigkeitsindex von 511 auf Kompetenzstufe III anzusiedeln.

Noch einmal deutlich schwieriger ist die dritte Frage zu beantworten. Hier sind zwei Angaben von Entfernung und Zeit zu berücksichtigen und in Beziehung zu setzen. So kann man sich überlegen, dass Helen insgesamt 7 km zurücklegt und dafür 15 Minuten braucht, was einer Durchschnittsgeschwindigkeit von 28 km/h entspricht. Damit ist der rechnerische Aufwand begrenzt, allerdings ist die Mathematisierung der Situation offensichtlich eine Hürde für viele Fünfzehnjährige gewesen. Sie gehört zu den schwierigsten PISA-Aufgaben und wurde mit einem Schwierigkeitsindex von 697 der Kompetenzstufe VI zugeordnet.

KABELFERNSEHEN

Die Tabelle unten zeigt für fünf Länder Daten über die Haushalte, die Fernseher besitzen. Sie zeigt außerdem den Prozentsatz derjenigen Haushalte, die Fernseher besitzen und auch Kabelfernsehen beziehen.



Land	Anzahl der Haushalte, die Fernseher besitzen	Prozentsatz der Haushalte, die Fernseher besitzen, im Vergleich zu allen Haushalten	Prozentsatz der Haushalte, die Kabelfernsehen haben, im Vergleich zu Haushalten, die einen Fernseher besitzen
Japan	48,0 Millionen	99,8 %	51,4 %
Frankreich	24,5 Millionen	97,0 %	15,4 %
Belgien	4,4 Millionen	99,0 %	91,7 %
Schweiz	2,8 Millionen	85,8 %	98,0 %
Norwegen	2,0 Millionen	97,2 %	42,7 %

Quelle: ITU, Welt-Telekommunikation Indikatoren 2004/2005
ITU, Welt-Telekommunikation/ICT Entwicklungsbericht

FRAGE 1: KABELFERNSEHEN

Die Tabelle zeigt, dass in der Schweiz 85,8 % aller Haushalte Fernseher besitzen.

Welche Schätzung kommt der Gesamtanzahl der Haushalte in der Schweiz am nächsten, wenn man von den Angaben in der Tabelle ausgeht?

- A 2,4 Millionen
- B 2,9 Millionen
- C 3,3 Millionen
- D 3,8 Millionen

FRAGE 2: KABELFERNSEHEN

Kevin sieht sich in der Tabelle die Angaben für Frankreich und Norwegen an.

Kevin sagt: „Weil der Prozentsatz aller Haushalte, die Fernseher besitzen, in beiden Ländern fast gleich ist, sind es in Norwegen mehr Haushalte, die Kabelfernsehen haben.“
Erkläre, warum diese Aussage falsch ist. Begründe deine Antwort.

.....
.....

Abbildung 2.4: Beispielaufgabe „Kabelfernsehen“

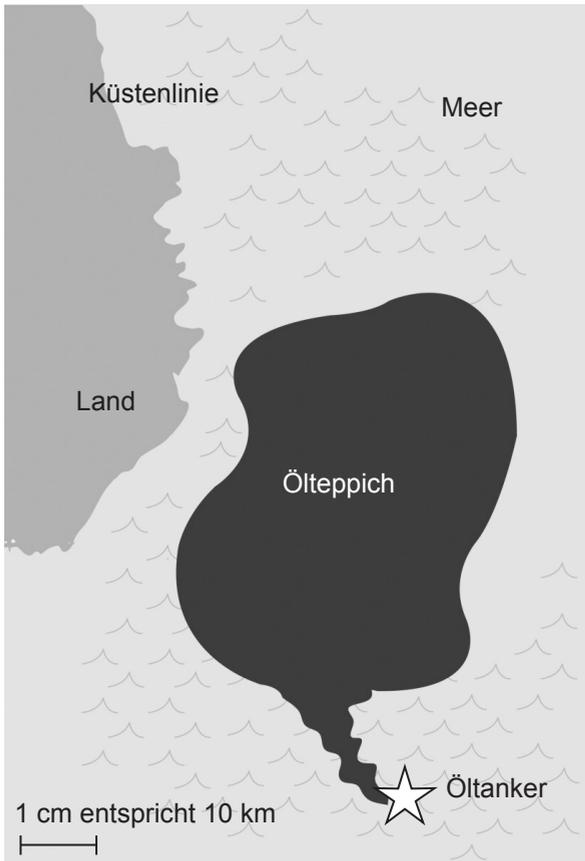
Auch in dieser Aufgabe steht ein den Jugendlichen vertrauter Kontext im Vordergrund. Die verwendeten Begriffe müssen nicht erarbeitet werden, sondern sind aus dem Alltag bekannt. Es geht dann zunächst darum, einer Tabelle Informationen zu entnehmen und diese für geeignete Fragestellungen zu nutzen. Hier müssen Relationen in der Tabelle gesehen und korrekt interpretiert werden. Die eigentliche Herausforderung ist im zweiten Teil gegeben, in dem begründet werden muss, warum eine Aussage fehlerhaft ist. Diese Begründung erfordert, verschiedene Werte richtig zu interpretieren und in ihren Abhängigkeiten zu sehen, also den Unterschied zwischen absoluten und relativen Häufigkeiten zu nutzen. Der erste Teil hat die eindeutige Lösung C, die entsprechend angekreuzt sein muss. Im zweiten Teil ist es notwendig, die Anzahl der Haushalte mit einem Fernsehgerät in die Überlegungen einzubeziehen, um zu einer korrekten Lösung zu kommen.

Es ist der Inhaltsbereich *Unsicherheit und Daten*, der sich in dieser Aufgabe findet. Man kann sie dem gesellschaftlichen Kontext zuordnen und es geht im Wesentlichen um den Prozess des Interpretierens. Dabei ist hier ganz besonders die kritische Reflexion gefordert, die auch in den Bildungsstandards eine wichtige Komponente darstellt.

Der erste Teil der Aufgabe gehört mit einem Schwierigkeitsindex von 537 zur Kompetenzstufe III, der zweite Teil mit einem Schwierigkeitsindex von 577 zur Kompetenzstufe IV. Offensichtlich ist es schwierig, einen Sachverhalt zu erklären und stichhaltig zu begründen.

ÖLTEPPICH

Ein Öltanker ist auf See gegen einen Felsen gefahren und hat dabei ein Loch in seine Öltanks gerissen. Der Tanker war ungefähr 65 km vom Land entfernt. Nach einigen Tagen hatte sich das Öl ausgebreitet, wie es auf der Karte unten dargestellt ist.



FRAGE 1: ÖLTEPPICH

Nutze den Maßstab der Karte und schätze die Fläche des Ölteppichs in Quadratkilometern (km²).

Antwort: km²

Abbildung 2.5: Beispielaufgabe „Ölteppich“

Die Aufgabe „Ölteppich“ geht von einem realen Problem aus, so wie es in den vergangenen Jahren immer wieder durch die Presse ging. Für viele Jugendliche kann damit der generelle Kontext als vertraut angenommen werden, wobei diese Vertrautheit zumeist weniger auf persönlicher als vielmehr auf gesellschaftlicher Ebene zu sehen ist. Wichtig ist nun, dass es für diese Aufgabe weder in der Realität noch in der vorliegenden Version einen eindeutigen Lösungsalgorithmus gibt. Die irreguläre Form des Ölteppichs verlangt letztendlich eine sinnvolle Annäherung an bekannte ebene Figuren, sodass mit geeigneten Formeln weitergearbeitet werden kann. Wichtig ist es, die Skalierung zu beachten. Dann kann man sich etwa ein Rechteck mit den Seiten 40 km und 60 km über den oberen Teil des Ölteppichs gelegt denken und erkennen, dass sich bedeckte und unbedeckte Flächen ungefähr ausgleichen. Die Lösung wäre somit 2400 km^2 . Wenn man den unteren Teil des Ölteppichs hinzunimmt, könnte man auf weitere 100 km^2 und somit auf insgesamt 2500 km^2 kommen. Aber auch ein Überdecken mit zwei Kreisen vom Radius 20 km bietet sich an, was zu einer Größe von $2 \cdot \pi \cdot 20^2 \text{ km}^2$, also ebenfalls zu etwa 2500 km^2 führt. Es kommt somit vor allem darauf an, dass die Jugendlichen einen eigenen Lösungsweg entwickeln und dabei ihr Wissen aus der Geometrie sinnvoll anwenden. Je nachdem, wie sie es machen, wird sich ein anderer Zahlenwert ergeben, wobei Lösungen zwischen 2200 km^2 und 3300 km^2 als korrekt gewertet wurden, wenn der angegebene Lösungsweg einen richtigen Ansatz darstellt.

Die Aufgabe lässt sich inhaltlich dem Bereich *Raum und Form* und einem wissenschaftlichen Kontext zuordnen. Dabei spielt das Anwenden die tragende Rolle. Es geht darum, dass geometrische Strukturen in einer Alltagssituation gesehen werden, zu einem adäquaten Modell dieser Situation führen und im Rahmen des Modells eine Lösung erarbeitet werden kann.

Mit einem Schwierigkeitsindex von 681 ist diese Aufgabe der Kompetenzstufe VI zugeordnet. Dabei liegt die Schwierigkeit vermutlich weniger bei den zu verwendenden Formeln, sondern eher bei der nicht eindeutigen Lösbarkeit. Es ist ein rationales Vorgehen gefordert, für das es aber keine klar formulierte Regel gibt. Aufgaben dieser Art sind nicht untypisch für die Mathematik und finden sich auch in den Bildungsstandards (z. B. Aufgabe „Riesenfass“ in den Bildungsstandards für die Hauptschule; KMK, 2004, S. 16).

2.4 Ergebnisse des internationalen Vergleichs

PISA 2012 mit Mathematik als Hauptdomäne bietet nun nach neun Jahren wieder die Gelegenheit, umfassend der Frage nachzugehen, wie die mathematische Kompetenz der Jugendlichen in Deutschland aus internationaler Perspektive einzuordnen ist und wo Problemlagen und Herausforderungen bestehen. Im Rahmen einer Gesamteinschätzung des Niveaus mathematischer Kompetenz (Abschnitt 2.4.1) betrachten wir – bezogen auf die Gesamtskala Mathematik – auch die Leistungsstreuung und die Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen (Abschnitt 2.4.1 bis 2.4.3). Im Anschluss daran richtet sich der Blick auf die Ergebnisse für die vier Teilskalen, die relative Stärken und Schwächen bezogen auf die vier untersuchten mathematischen Inhaltsbereiche im internationalen Vergleich der OECD-Staaten erkennen lassen (Abschnitt 2.4.4). Abschließend werfen wir in Abschnitt 2.4.5 einen Blick auf Veränderungen in der mathematischen Kompetenz seit PISA 2003.

2.4.1 Ergebnisse des internationalen Vergleichs auf der Gesamtskala

Der OECD-Durchschnitt mathematischer Kompetenz war in PISA 2003 auf 500 normiert worden, die Standardabweichung auf 100 (Prenzel et al., 2004). Da die OECD seit PISA 2003 um vier Staaten (Chile, Estland, Israel, Slowenien) erweitert wurde, beruht der OECD-Durchschnitt in PISA 2012 auf einer anderen Zusammensetzung von Staaten als in PISA 2003. Veränderungen im Gesamtmittelwert für die OECD seit PISA 2003 und auf der damals normierten Skala zeigen deshalb möglicherweise Effekte der neuen Komposition der OECD an, und nicht unbedingt Änderungen in der mathematischen Leistung. Unter Nutzung der nicht veröffentlichten Link-Items, die in allen PISA-Runden eingesetzt werden, können die Leistungen zwischen den Erhebungen auf der Basis einer gemeinsamen Skala (unter einer kriterialen Bezugsnorm) verglichen werden. In PISA 2012 beträgt der OECD-Durchschnitt für die mathematische Kompetenz 494 Punkte ($SE = 0.5$; $SD = 92$; $SE = 0.3$). Die Abnahme gegenüber dem OECD-Durchschnitt in PISA 2003 kann zu einem Teil auf die veränderte Zusammensetzung der OECD-Staaten zurückgeführt werden, denn zwei der neu hinzugekommenen Staaten liegen deutlich unter dem internationalen Mittelwert.

Die wichtigsten Befunde des internationalen Vergleichs der Fünfzehnjährigen werden in Abbildung 2.6 (Mittelwerte, Standardabweichungen und Perzentilbänder der OECD-Staaten) zusammengefasst.

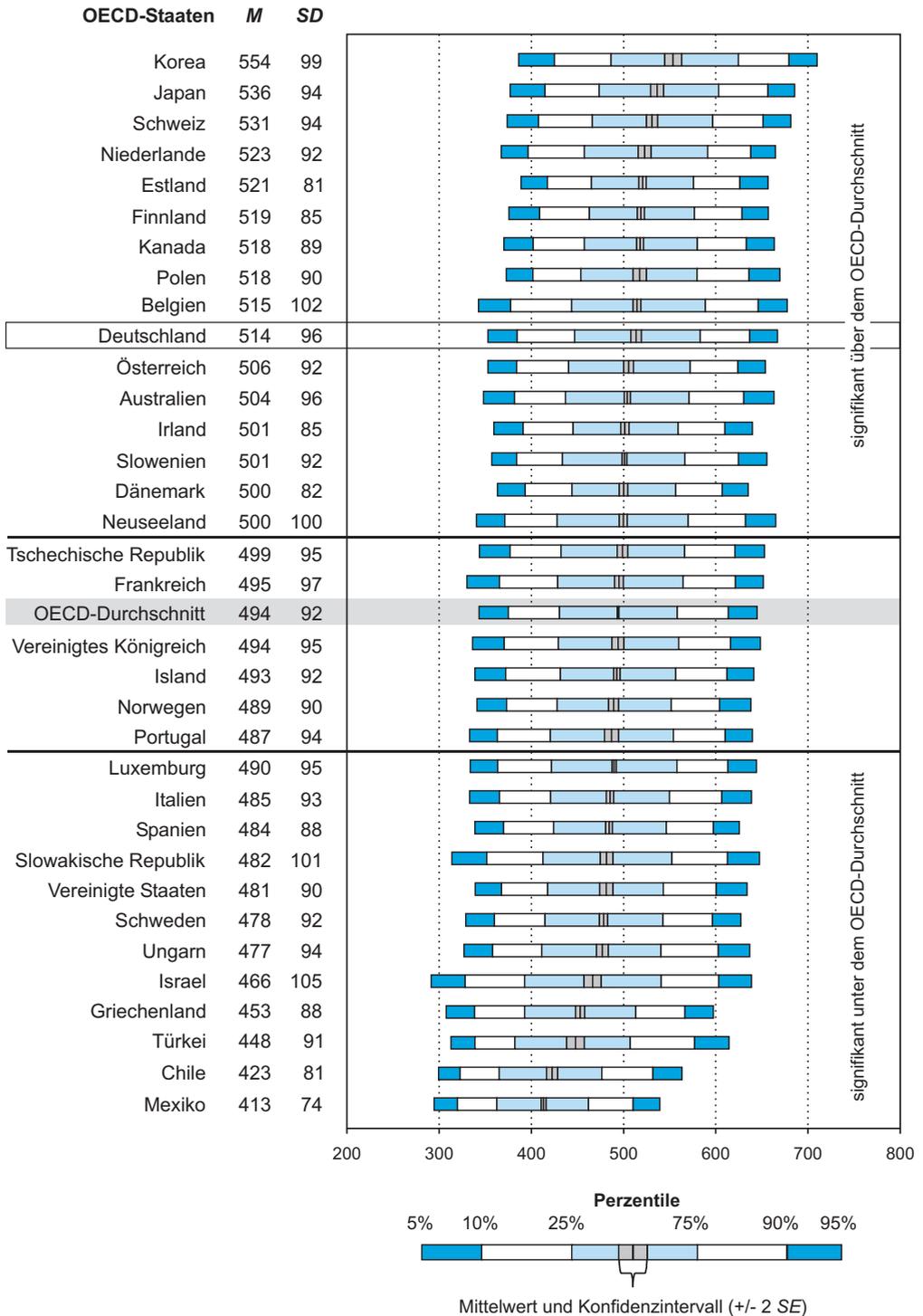


Abbildung 2.6: Perzentilbänder mathematischer Kompetenz in den OECD-Staaten

Die Abbildung 2.6 berichtet die Mittelwerte aller OECD-Staaten sowie sogenannte Perzentilbänder. Während die Mittelwerte über das im Durchschnitt erreichte Niveau mathematischer Kompetenz Auskunft geben, veranschaulichen die Perzentilbänder die Streuung und Verteilung der Leistungen. Bei den abgebildeten Perzentilen handelt es sich um Skalenwerte, unterhalb derer sich jeweils 5, 10, 25, 75, 90 beziehungsweise 95 Prozent aller Schülerinnen und Schüler in den verschiedenen Staaten befinden. Die Perzentilbänder zeigen somit graphisch an, wie groß die Streuung der Kompetenz innerhalb eines Landes ist (vgl. Kapitel 1). Die Kennwerte im Einzelnen können dem Anhang (Tabelle 1) entnommen werden.

Die Gruppe der OECD-Staaten, deren *Mittelwert* statistisch signifikant ($p < .05$) über dem OECD-Durchschnitt von 494 Punkten liegt, umfasst 16 Staaten. Innerhalb dieser Gruppe erstreckt sich das Kompetenzniveau von 500 Punkten (Neuseeland, Dänemark) bis zu 554 Punkten (Korea). Auch Deutschland gehört mit 514 Punkten ($SE = 2.9$; $SD = 96$; $SE = 1.6$) zu den OECD-Staaten mit einer überdurchschnittlichen Leistung in der Mathematik. Deutschland bestätigt damit das Ergebnis aus der letzten Erhebung (PISA 2009; vgl. Frey, Heinze, Mildner, Hochweber & Asseburg, 2010). Nur mehr fünf OECD-Staaten (Korea, Japan, Estland, Niederlande, Schweiz) erreichen ein (statistisch signifikant) höheres Niveau mathematischer Kompetenz als Deutschland. Bezieht man die Mathematikleistungen der OECD-Partnerstaaten in die Betrachtung ein, dann findet man weitere Staaten (bzw. Regionen), die eine im Durchschnitt signifikant höhere mathematische Kompetenz aufweisen als Deutschland, nämlich Shanghai (China), Singapur, Hongkong (China), Chinesisch Taipeh, Macao (China) und Liechtenstein. Definiert man eine absolute Spitzengruppe anhand des Kriteriums, dass der Mittelwert mindestens eine halbe Standardabweichung über dem OECD-Durchschnitt liegt, dann versammeln sich dort ausschließlich Staaten aus dem asiatischen Raum: Shanghai (China, 613 Punkte), Singapur (573 Punkte), Hongkong (China, 561 Punkte), Chinesisch Taipeh (560 Punkte) sowie Korea (554 Punkte). Betrachtet man die Abstände zum OECD-Durchschnitt, aber auch zum Mittelwert für Deutschland, dann repräsentieren diese Staaten ein deutlich anderes und höheres Niveau mathematischer Kompetenz. Bei der nationalen *Follow-up*-Erhebung zu PISA 2003 belief sich der Leistungsabstand zwischen der 9. und 10. Jahrgangsstufe auf knapp dreißig Punkte (Ehmke, Blum, Neubrand, Jordan & Ulfing, 2006). Nimmt man diesen Wert als Anhaltspunkt, um die Mathematikleistungen der Spitzengruppe abzuschätzen, dann beträgt der Vorsprung von Korea gegenüber Deutschland bereits mehr als ein Schuljahr.

Eine Gruppe von sechs OECD-Staaten unterscheidet sich nicht signifikant vom OECD-Durchschnitt. In ihr finden sich europäische Staaten wie Frankreich (495 Punkte), das Vereinigte Königreich (494 Punkte) oder Norwegen (489 Punkte). Die Gruppe der zwölf OECD-Staaten, die signifikant unter dem OECD-Durchschnitt liegt, reicht von Luxemburg (490 Punkte) bis Mexiko (413 Punkte). In dieser Gruppe von Staaten unter dem OECD-Durchschnitt befinden sich auch Italien (485 Punkte), Spanien (484 Punkte), die Vereinigten Staaten (481 Punkte) und Schweden (478 Punkte).

Die *Streubreite* in der Kompetenzverteilung ist ebenfalls in Abbildung 2.6 an den Perzentilbändern zu erkennen. Die Länge eines Perzentilbandes bildet den Abstand zwischen den 5 Prozent leistungsstärksten zu den 5 Prozent leistungsschwächsten Schülerinnen und Schülern ab und veranschaulicht die Bandbreite der in einem Land erzielten Leistungen. Zugleich zeigen die Perzentilbänder über die Staaten hinweg, inwieweit sich die Leistungen der Schülerinnen und Schüler im Vergleich mehrerer Länder überlappen. In grau sind der Mittelwert sowie die beiden Konfidenzintervalle abgetragen (vgl. Kapitel 1). Der Abbildung 2.6 kann entnommen werden, dass die Breite der Leistungsstreuung in allen OECD-Staaten beträchtlich ist und dass es zwischen allen Staaten einen großen Überlappungsbereich gibt. Die Streubreite innerhalb der Staaten unterscheidet sich jedoch deutlich voneinander. Die größte Streubreite unter den OECD-Staaten weist Israel auf, die geringste Mexiko.

Die Streubreite der mathematischen Kompetenz liegt in Deutschland mit einer Standardabweichung von 96 Punkten statistisch im Bereich des OECD-Durchschnitts ($SD = 92$ Punkte). Anhand der Perzentilbänder ist zu erkennen, dass die Staaten oberhalb des OECD-Durchschnitts nicht unbedingt durch einen großen Abstand zwischen den leistungsstärksten und den leistungsschwächsten Schülerinnen und Schülern auffallen. In dieser Gruppe von Staaten ist beispielsweise die Leistungsstreuung in Belgien ($SD = 102$ Punkte) und Korea ($SD = 99$ Punkte) höher als im OECD-Durchschnitt; geringe Streuungen findet man dagegen in Estland ($SD = 81$), Dänemark ($SD = 82$) Finnland ($SD = 85$) und Irland ($SD = 85$). Ein hohes durchschnittliches Leistungsniveau muss also durchaus nicht durch eine hohe Streuung „erkaufte“ werden.

2.4.2 Verteilungen auf die Kompetenzstufen

Mit Hilfe von Kompetenzstufen lässt sich gut veranschaulichen, welche Anforderungen Schülerinnen und Schüler mit bestimmten Punktwerten mit hoher Wahrscheinlichkeit meistern und an welchen sie vermutlich scheitern werden. Mit besonderem Interesse wird bei PISA verfolgt, wie groß der Anteil von Schülerinnen und Schülern auf der untersten Kompetenzstufe (I und darunter) ist, deren Mathematikverständnis allenfalls den Anforderungen der letzten Grundschulklasse entspricht. Für das Bildungs-, Wirtschafts- und Sozialsystem ist ein geringer Anteil von Jugendlichen auf oder unter der ersten Kompetenzstufe ein wichtiger Indikator (vgl. Wößmann & Piopiunik, 2009). Auf der anderen Seite weisen die Anteile auf den höchsten Kompetenzstufen darauf hin, inwieweit es gelingt, Talente zu entwickeln und Nachwuchs für Studiengänge und Berufsfelder zu gewinnen, die hohe Anforderungen an das mathematische Verständnis stellen. Dementsprechend präsentiert die Abbildung 2.7 auf der linken Seite die prozentualen Anteile der Jugendlichen auf der Kompetenzstufe I und darunter, auf der rechten Seite die prozentualen Anteile auf den Kompetenzstufen V und VI.

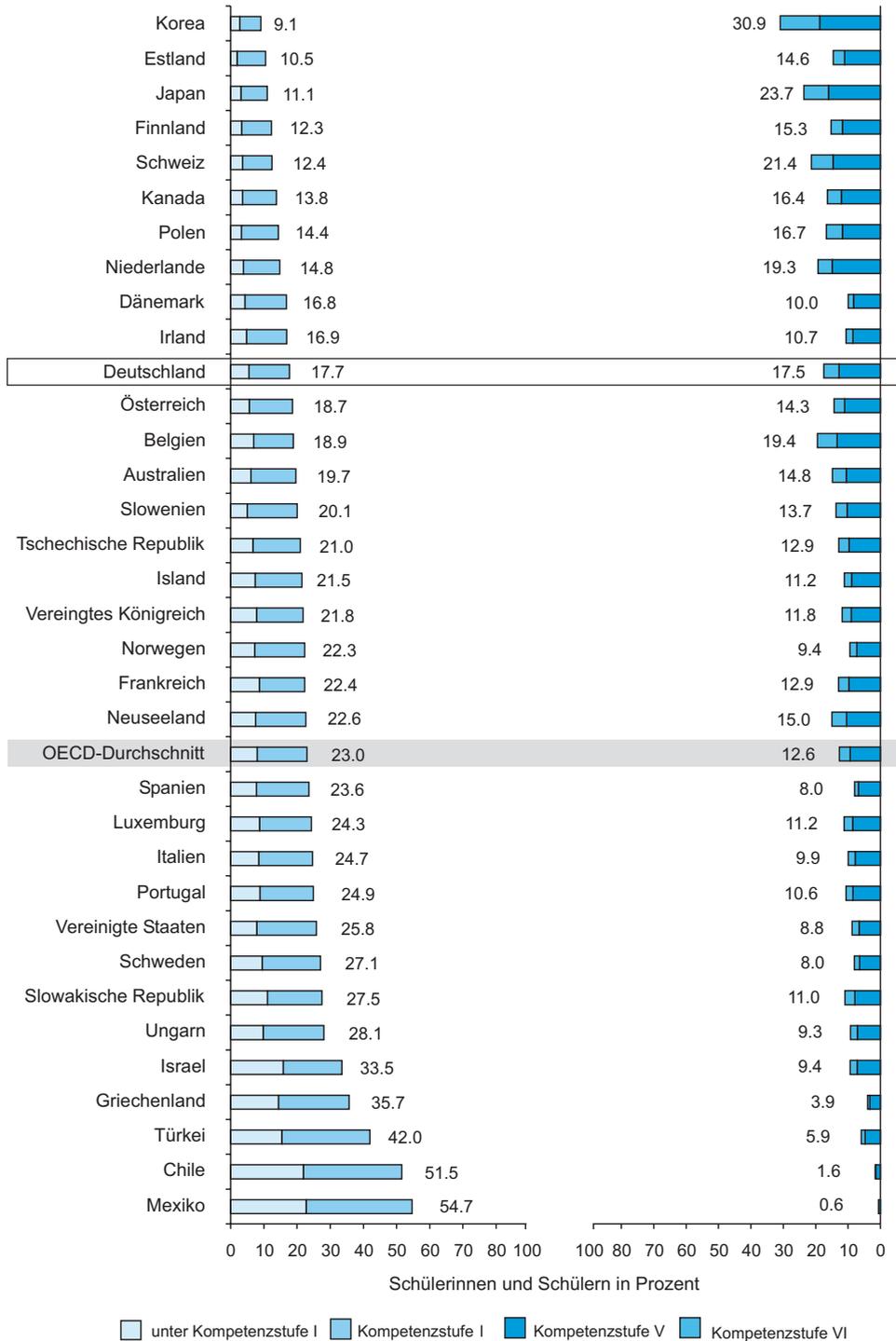


Abbildung 2.7: Prozentuale Anteile von Schülerinnen und Schülern auf Kompetenzstufe I oder darunter beziehungsweise auf Kompetenzstufe V oder VI

Über die OECD-Staaten hinweg werden 23.0 Prozent der Fünfzehnjährigen unterhalb der Kompetenzstufe II eingeordnet; sie befinden sich also auf der Stufe I oder sogar noch darunter. Dieser Anteil fällt in Deutschland mit 17.7 Prozent deutlich geringer aus. Allerdings bedeutet dieser Wert auch, dass jeder sechste Jugendliche in Deutschland die Mindestanforderungen für ein anschlussfähiges mathematisches Verständnis nicht erreicht und erhebliche Probleme haben dürfte, einen Ausbildungsplatz zu finden beziehungsweise eine Ausbildung erfolgreich abzuschließen sowie anspruchsvollere mathematische Anforderungen im Alltag zu bewältigen. Der internationale Vergleich zeigt, dass dieser Anteil durchaus kleiner sein kann. Benchmarks für Deutschland setzen vor allem Nachbarstaaten wie die Niederlande, Polen oder die Schweiz, bei denen der Anteil von Jugendlichen auf und unter der ersten Kompetenzstufe um mehrere Prozentpunkte geringer ist.

Im obersten Leistungsbereich beläuft sich der Anteil von Schülerinnen und Schülern auf den Kompetenzstufen V und VI für die OECD-Staaten durchschnittlich auf 12.6 Prozent. Auch mit Blick auf die Förderung einer Spitzengruppe erweist sich Deutschland mit einem Anteil von 17.5 Prozent als „überdurchschnittlich“ im Vergleich zu den OECD-Staaten. Allerdings zeigt sich hier wiederum, dass nicht nur Staaten wie Korea (30.9 Prozent) und Japan (23.7 Prozent) eine sehr viel erfolgreichere Talentförderung betreiben, sondern auch Staaten in unserer unmittelbaren Nachbarschaft. Die Anteile der Fünfzehnjährigen auf Kompetenzstufe V und VI betragen in der Schweiz 21.4 Prozent, in den Niederlanden 19.3 Prozent, in Belgien 19.4 Prozent. Gerade die Schweiz und die Niederlande können somit als Beispiele für eine erfolgreiche Förderung im unteren Leistungsbereich und zugleich in der Spitze gelten.

2.4.3 Kompetenzunterschiede zwischen Jungen und Mädchen

Mit großer Spannung wird in den internationalen Vergleichsstudien jeweils verfolgt, wie die Kompetenzunterschiede zwischen Mädchen und Jungen speziell in der Mathematik ausfallen. Spannung kommt deshalb auf, weil seit den ersten Vergleichen bei PISA immer wieder Staaten identifiziert werden, in denen keine Geschlechterunterschiede festzustellen sind beziehungsweise diese sogar den überbrachten Stereotypen widersprechen. Ein ähnliches Bild findet man übrigens auch bei den Untersuchungen im Grundschulbereich (z. B. Brehl, Wendt & Bos, 2012). Beispiele für vernachlässigbare Geschlechterdifferenzen in der mathematischen Kompetenz – gerade auch in leistungsstarken Staaten – fordern zu besonderen Anstrengungen heraus, diese Disparitäten weltweit aufzulösen.

Die in Abbildung 2.8 dargestellten Mittelwerte für PISA 2012 zeigen für die Gesamtskala Mathematik einen durchschnittlichen Differenzwert zwischen Mädchen und Jungen in der Größenordnung von 11 Punkten (gerundet) zugunsten der Jungen über alle OECD-Staaten hinweg. Diese Differenz kann statistisch gegen den Zufall abgesichert werden. Der Abbildung zufolge sind die Unterschiede in zehn Staaten statistisch

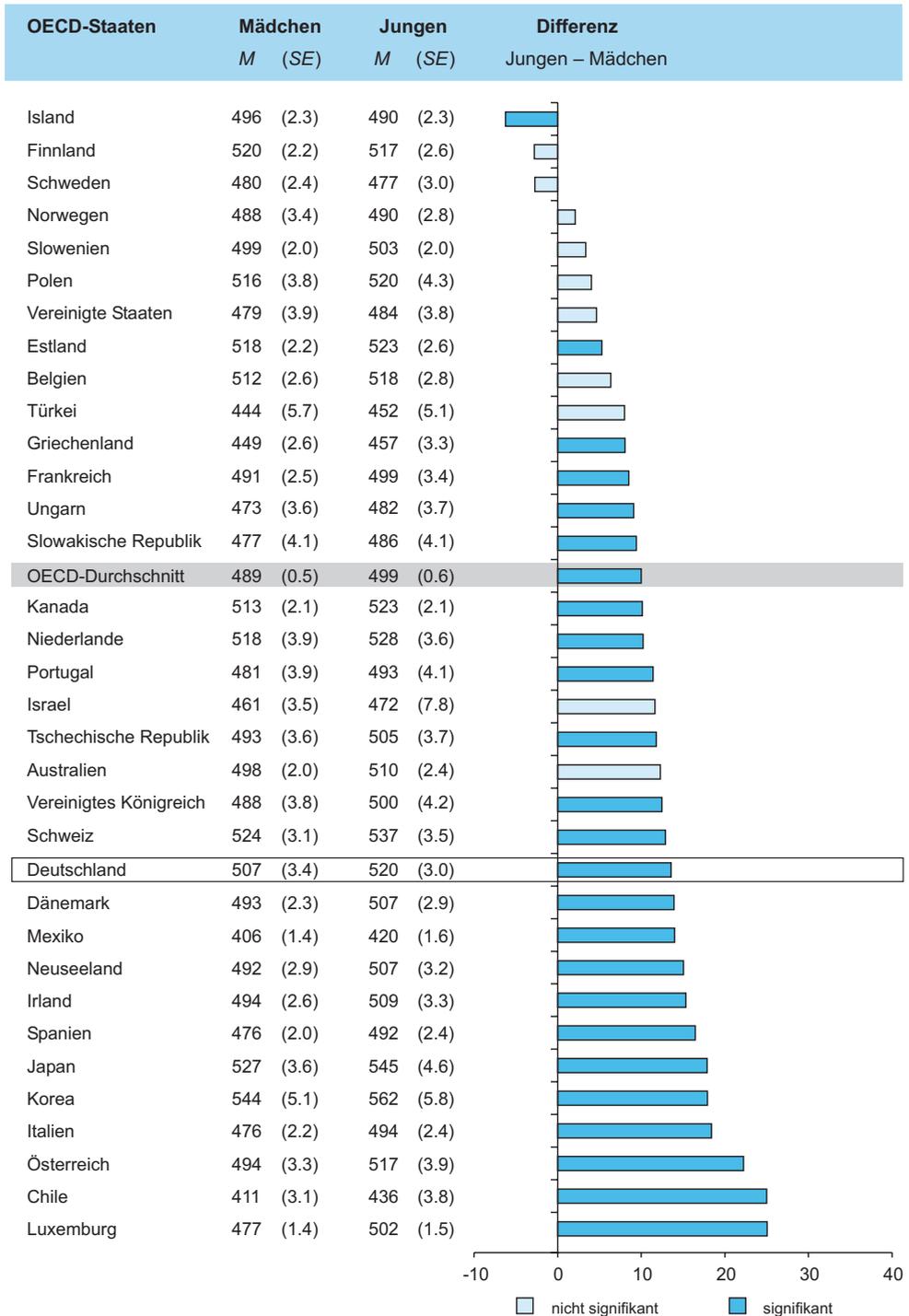


Abbildung 2.8: Mittelwerte mathematischer Kompetenz von Jungen und Mädchen in den OECD-Staaten

nicht signifikant, zum Beispiel in Finnland, Schweden, Norwegen, aber auch in Polen, der Türkei, Australien oder in den Vereinigten Staaten. In Island erreichen die Mädchen statistisch signifikant höhere Werte in der Gesamtskala mathematischer Kompetenz. In Deutschland ist das umgekehrt, hier beträgt die gerundete Differenz 14 Punkte, und der Unterschied ist statistisch abgesichert. Vergleicht man die Verteilungen der mathematischen Kompetenz für die Geschlechter in Deutschland, dann überlappen sich diese weitgehend. Die Mittelwertsdifferenz kommt insbesondere dadurch zustande, dass die Spitzengruppe der Jungen ein höheres Niveau erreicht als die der Mädchen. Den Beispielen einiger nordischer Staaten folgend, könnte es sich auch für den Gesamtmittelwert Deutschlands als sehr förderlich erweisen, wenn weitere und verstärkte Anstrengungen unternommen würden, gezielt die mathematische Kompetenz der Mädchen zu fördern.

2.4.4 Ergebnisse in den vier Inhaltsbereichen mathematischer Kompetenz

Neben der Betrachtung der Gesamtskala Mathematik lohnt es sich, einen differenzierenden Blick in die vier Inhaltsbereiche zu werfen, die in PISA 2012 für die mathematische Kompetenz erfasst worden sind. Mit ihrer Hilfe können gegebenenfalls besondere Stärken und Schwächen unserer Schülerinnen und Schüler identifiziert werden. Vier Subskalen (vgl. Abschnitt 2.2.1) bilden die Kompetenz in den Inhaltsbereichen *Veränderung und Beziehungen*, *Quantität*, *Raum und Form* sowie *Unsicherheit und Daten* ab. Bereits in PISA 2003 waren diese vier Inhaltsbereiche (damals: „übergreifende Ideen“, vgl. Blum et al., 2004) untersucht worden. In PISA 2003 hatten sich für Deutschland zum Teil markante Unterschiede in der Kompetenz der Schülerinnen und Schüler gezeigt, wenn man statt der Skala mathematischer Kompetenz als Ganzes die einzelnen Inhaltsbereiche betrachtete (Blum et al., 2004). Die Leistungen der Schülerinnen und Schüler variierten zwischen 493 Punkten (Inhaltsbereich *Unsicherheit*) und 514 Punkten (Inhaltsbereich *Quantität*). Die Differenz von 21 Punkten zwischen den niedrigsten und den höchsten Kompetenzwerten war beträchtlich und entsprach mehr als einem halben Schuljahr Lernfortschritt. Daher ist im Rahmen von PISA 2012 von besonderem Interesse, ob die Jugendlichen mittlerweile konsistentere Leistungen in den vier Inhaltsbereichen zeigen.

Veränderung und Beziehungen

Zum Inhaltsbereich *Veränderung und Beziehungen* gehören Aufgaben, in denen es um funktionale und relationale Beziehungen zwischen mathematischen Objekten geht. Im schulischen Kontext kommt das Themengebiet der Algebra nahe an diesen Inhaltsbereich heran. Abbildung 2.9 zeigt die Mittelwerte und Perzentilbänder für die OECD-Staaten. Der OECD-Durchschnitt für den Inhaltsbereich *Veränderung und Beziehungen* liegt mit 493 Punkten etwas unter dem Mittel der Gesamtskala. Der Mittelwert für

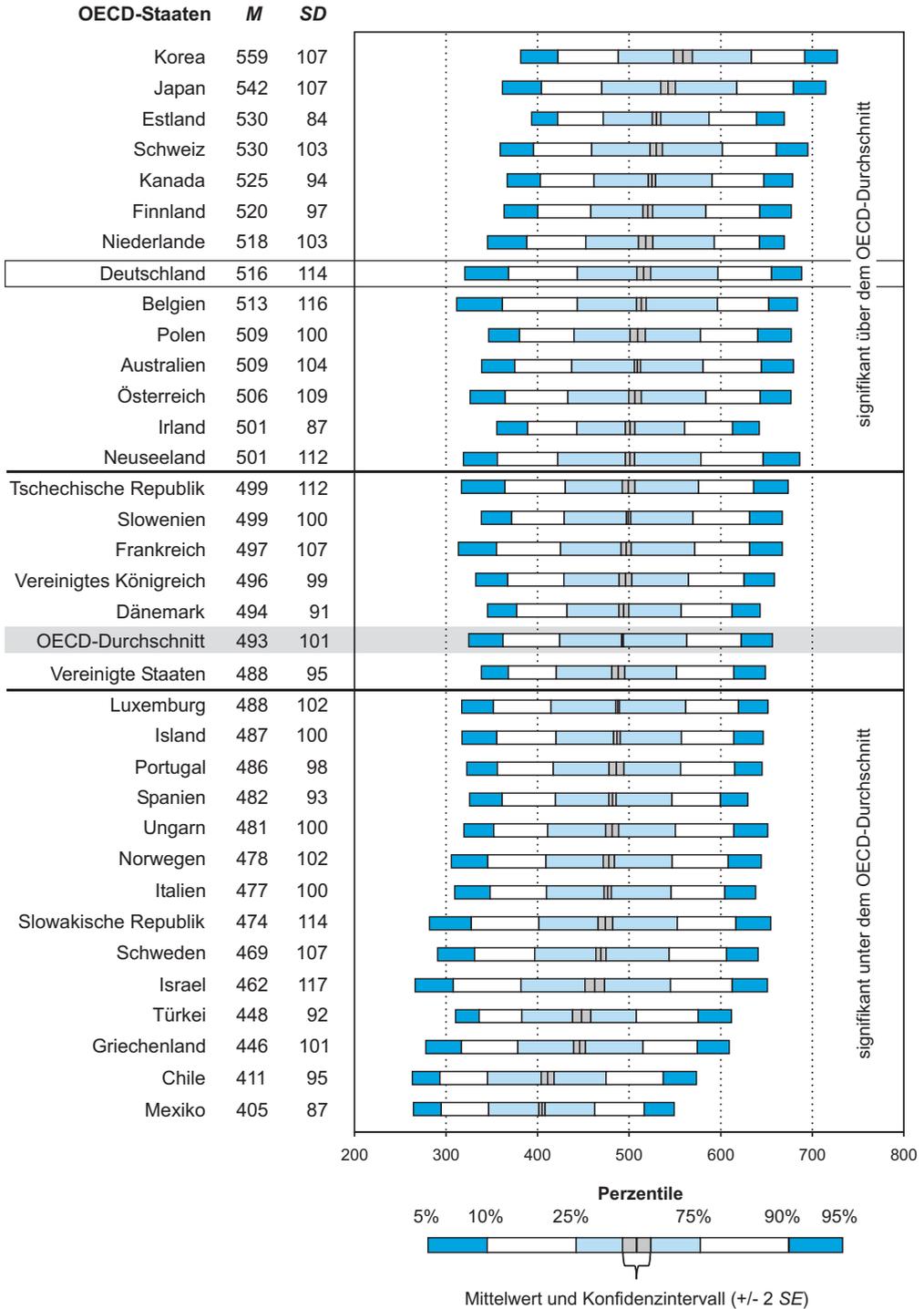


Abbildung 2.9: Perzentilbänder für den Inhaltsbereich *Veränderung und Beziehungen* in den OECD-Staaten

Deutschland von 516 Punkten befindet sich signifikant über dem OECD-Durchschnitt und nahe beim deutschen Gesamtdurchschnitt in Mathematik (514 Punkte). Die Jungen schneiden in Deutschland mit 521 Punkten in diesem Bereich um 11 Punkte besser ab als die Mädchen (510 Punkte).

Spitzenwerte mit 559 und 542 Punkten erzielen Korea und Japan; auch ansonsten gleicht die Verteilung der Staaten weitgehend der Gesamtskala.

Quantität

Die Verwendung von Zahlen, um Situationen zu strukturieren oder zu beschreiben, steht im Mittelpunkt des Inhaltsbereiches *Quantität*. In Abbildung 2.10 sind die zugehörigen Mittelwerte und Perzentilbänder für die OECD-Staaten dargestellt. Der OECD-Durchschnitt liegt bei 495 Punkten und entspricht damit dem Durchschnitt der Gesamtskala Mathematik. Die Schülerinnen und Schüler in Deutschland erreichen eine durchschnittliche Kompetenz von 517 Punkten und liegen damit signifikant über dem OECD-Durchschnitt. *Quantität* ist damit ein Inhaltsbereich, in dem die Schülerinnen und Schüler aus Deutschland relative Stärken aufweisen. Auch in diesem Bereich erreicht Korea mit 537 Punkten den höchsten Wert im OECD-Vergleich; gefolgt von den Niederlanden (532 Punkte) und der Schweiz (531 Punkte). Japan, welches auf der Gesamtskala Mathematik den zweithöchsten durchschnittlichen Wert der OECD-Länder erzielt, erreicht im Bereich *Quantität* aus allen vier Inhaltsbereichen den geringsten Wert (518 Punkte). Ebenso wie beim Inhaltsbereich *Veränderung und Beziehungen* erzielen Chile (421 Punkte) und Mexiko (414 Punkte) die mit Abstand niedrigsten Werte auf der Kompetenzskala *Quantität*.

Raum und Form

Das Gebiet der Geometrie wird im Inhaltsbereich *Raum und Form* abgedeckt, für den Abbildung 2.11 die Perzentilbänder und Mittelwerte der OECD-Staaten zeigt. Deutschland erreicht hier mit 507 Punkten seinen niedrigsten Mittelwert im Vergleich zu den anderen Inhaltsbereichen. *Raum und Form* beziehungsweise die Geometrie dürfen damit als relativer Schwachpunkt der mathematischen Kompetenz der Schülerinnen und Schüler in Deutschland betrachtet werden. Dennoch liegt der Mittelwert für Deutschland immer noch signifikant über dem OECD-Durchschnitt von 490 Punkten. Korea erreicht in diesem Teilbereich mit Abstand die höchsten Kompetenzwerte (573 Punkte), auch Japan (558 Punkte) und die Schweiz (544 Punkte) sind sehr erfolgreich. Alle drei Staaten haben im Inhaltsbereich *Raum und Form* ihren jeweils höchsten Kompetenzwert aus allen vier Inhaltsbereichen.

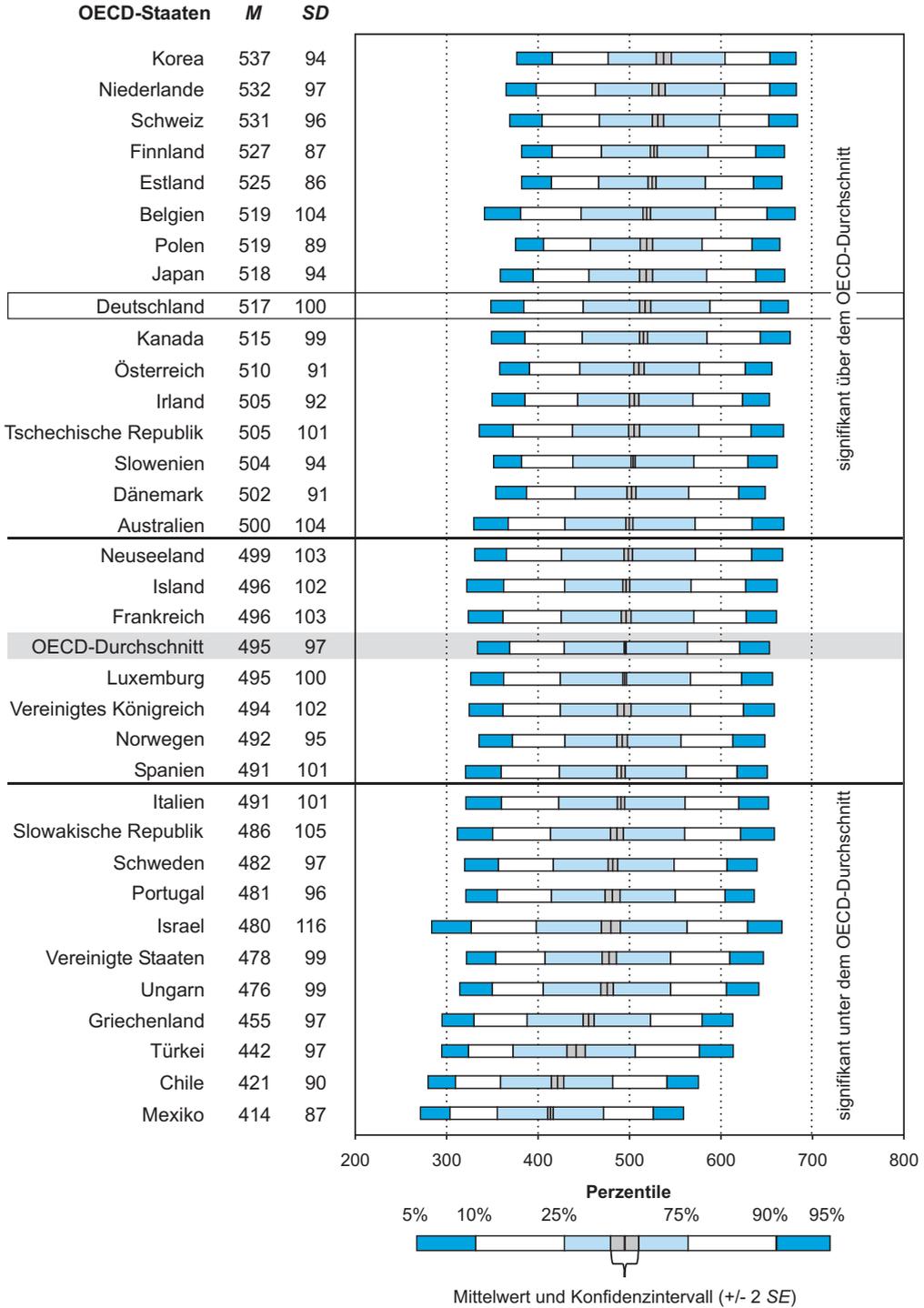


Abbildung 2.10: Perzentilbänder für den Inhaltsbereich *Quantität* in den OECD-Staaten

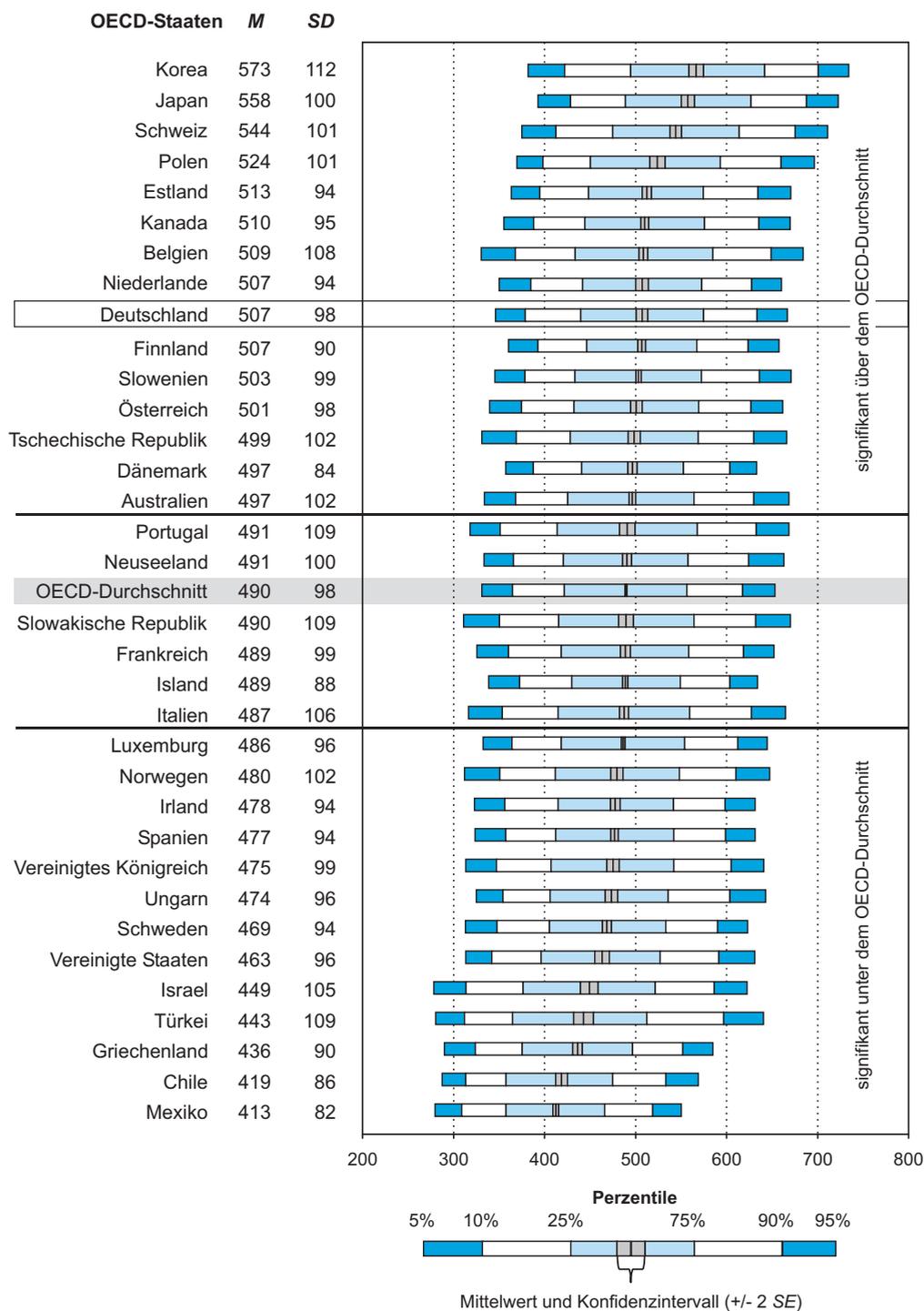


Abbildung 2.11: Perzentilbänder für den Inhaltsbereich *Raum und Form* in den OECD-Staaten

Unsicherheit und Daten

Im Inhaltsbereich *Unsicherheit und Daten* geht es um statistische Daten und Zufallsphänomene. Abbildung 2.12 gibt die Mittelwerte und Perzentilbänder für die OECD-Staaten wieder. In diesem Bereich erzielten bei PISA 2003 die Schülerinnen und Schüler in Deutschland die niedrigsten Kompetenzwerte (Blum et al., 2004). In PISA 2012 schneiden die Jugendlichen in Deutschland mit 509 Punkten zwar im Vergleich zu *Veränderung und Beziehungen* (516 Punkte) und *Quantität* (517 Punkte) etwas schlechter ab, allerdings ist die Distanz zwischen stärkeren und schwächeren Inhaltsbereichen in Deutschland seit 2003 deutlich geringer geworden. Auch im Inhaltsbereich *Unsicherheit und Daten* liegt der Mittelwert für Deutschland signifikant über dem OECD-Durchschnitt. Korea erzielt auch im Inhaltsbereich *Unsicherheit und Daten* den Spitzenwert von 538 Punkten, die Niederlande (532 Punkte) und Japan (528 Punkte) schneiden ähnlich gut ab. Chile (430 Punkte) und Mexiko (413 Punkte) sind, wie in den übrigen drei Bereichen, die beiden Staaten mit den niedrigsten Kompetenzwerten.

Zusammenfassung der vier Inhaltsbereiche

Insgesamt zeigt sich bei einer detaillierten Betrachtung der vier Inhaltsbereiche mathematischer Kompetenz, dass der OECD-Staat mit der durchschnittlich höchsten mathematischen Kompetenz, nämlich Korea, auch in allen vier Inhaltsbereichen die besten Werte im Vergleich der OECD-Staaten erzielt. Gleichzeitig ist Korea jedoch auch das Land, in dem die erreichten Werte in den vier Inhaltsbereichen nach Japan (39 Punkte) am weitesten auseinander liegen (36 Punkte). Die Differenz zwischen *Raum und Form* als stärkstem Bereich sowie *Quantität* als schwächstem Bereich ist in den beiden Staaten, die insgesamt die höchste durchschnittliche mathematische Kompetenz aufweisen, also substanziell. Auch in den Niederlanden, die in allen Inhaltsbereichen über dem OECD-Mittel liegen, treten große Unterschiede zwischen den Inhaltsbereichen hervor (25 Punkte). Zwischen *Unsicherheit und Daten* (507 Punkte) und *Raum und Form* (532 Punkte) liegt eine bemerkenswerte Distanz, die relative Stärken und Schwächen in den Schulen der Niederlande erkennen lässt.

Staaten mit besonders geringen Differenzen zwischen den Inhaltsbereichen finden sich vorwiegend rund um den OECD-Durchschnitt. Deutschland ist einer derjenigen Staaten, die insgesamt signifikant über dem OECD-Durchschnitt liegen, und eine relativ geringe Diskrepanz zwischen den Inhaltsbereichen zeigen. Die Leistungen der Schülerinnen und Schüler in Deutschland sind über die Inhaltsbereiche hinweg konsistent über dem OECD-Durchschnitt. Dennoch weisen die Befunde auf relative Stärken in den Bereichen *Veränderung und Beziehungen* und *Quantität* hin sowie auf relative Schwächen in den beiden anderen Bereichen *Unsicherheit und Daten* und *Raum und Form*. Auch hier zeichnet sich somit ein Ansatzpunkt für eine weitere Verbesserung mathematischer Kompetenz in Deutschland ab.

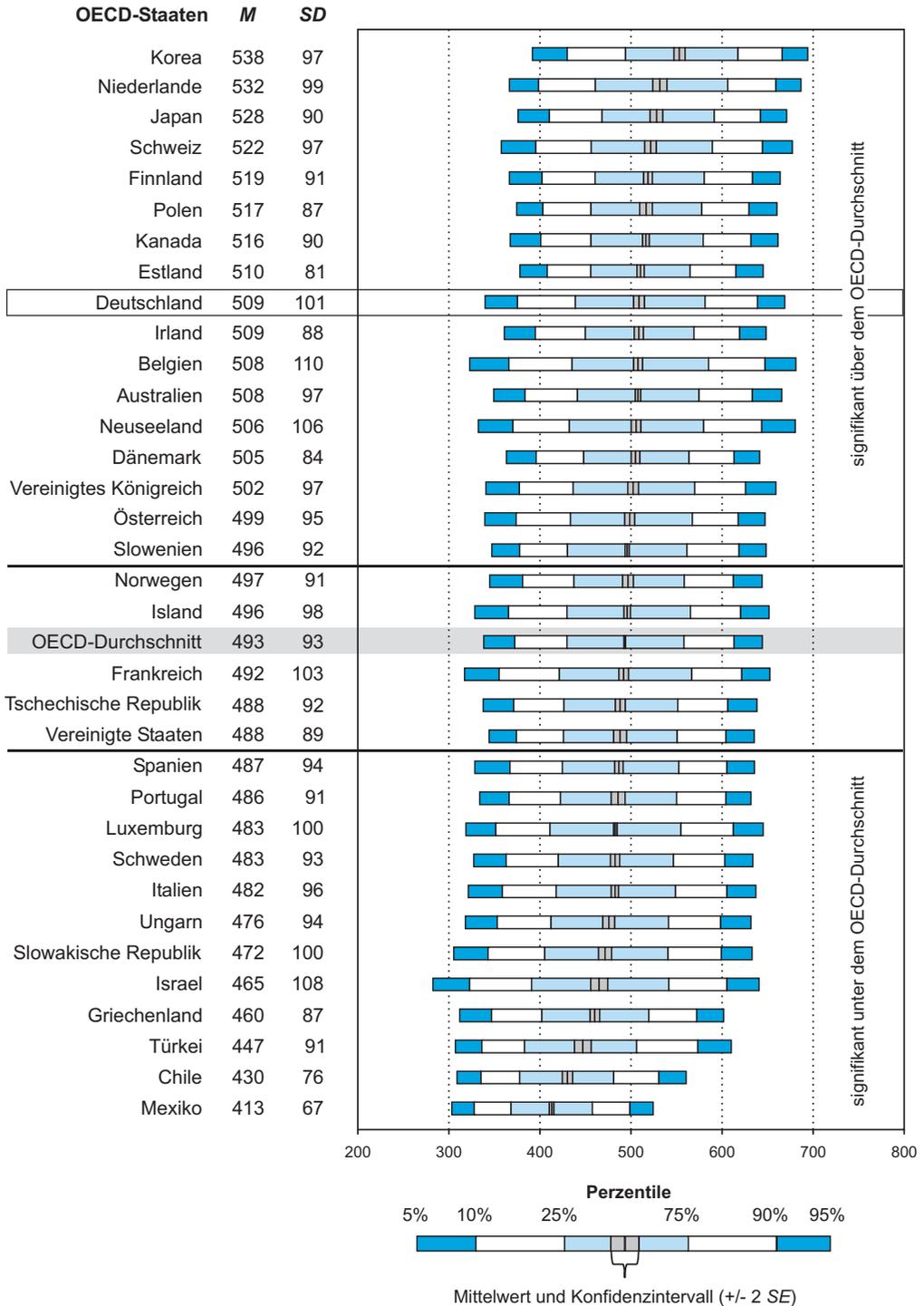


Abbildung 2.12: Perzentilbänder für den Inhaltsbereich *Unsicherheit und Daten* in den OECD-Staaten

2.4.5 Veränderungen seit PISA 2003

Vor dem Hintergrund der mittelmäßigen Ergebnisse der ersten PISA-Runde ist das Interesse an Informationen über Entwicklungen und Trends in Deutschland besonders stark ausgeprägt. Gerade im Bereich der Mathematik wurden mehrere Initiativen zur Sicherung und Weiterentwicklung der Unterrichtsqualität unternommen. PISA 2012 bietet nun die Möglichkeit, die Veränderungen in der mathematischen Kompetenz seit PISA 2003, als die mathematische Kompetenz schon einmal die Hauptdomäne war, im internationalen Vergleich zu betrachten.

Die Abbildung 2.13 berichtet die Mittelwerte der OECD-Staaten, die sowohl 2003 als auch 2012 an PISA teilgenommen haben, sowie jeweils die Differenz zwischen beiden Messzeitpunkten. Die Staaten sind dabei nach der absoluten Größe der Differenz angeordnet.

Für dreizehn Staaten sind signifikante Abnahmen der mathematischen Kompetenz seit PISA 2003 zu verzeichnen. Bei neun Staaten lassen sich die Veränderungen nicht statistisch absichern, dementsprechend ist ihr Niveau mathematischer Kompetenz seit 2003 in etwa gleich geblieben. In sieben Staaten sind signifikante Steigerungen der Mathematikleistung nachzuweisen. Bereits diese Übersicht weist auf einen beträchtlichen und interessanten Wandel in den Unterrichtsergebnissen der OECD-Staaten hin. Es ist also keineswegs selbstverständlich, gute Ergebnisse über einen Zeitraum von neun Jahren zu halten. Offensichtlich sind aber sowohl eine deutliche Steigerung als auch eine Abnahme der Mathematikleistung in einem solchen Zeitraum möglich.

In der Gruppe der Staaten, für die Abnahmen zu berichten sind, findet man mehrere Länder, die in den ersten PISA-Runden durch besonders gute Mathematikleistungen auffielen. Stark verschlechtert hat sich die mathematische Kompetenz in Schweden (31 Punkte), Finnland (26 Punkte) und Neuseeland (23 Punkte). Der durchschnittliche Rückgang in diesen Staaten erstreckt sich bis zu einem Leistungsabstand von einem Schuljahr. Auch für viele Staaten in der Nachbarschaft Deutschlands sind signifikante Abnahmen der mathematischen Kompetenz zu beobachten (z. B. die Tschechische Republik, die Slowakische Republik, Frankreich, Niederlande, Belgien, Dänemark).

Auf der anderen Seite konnte sich eine kleinere Gruppe von Staaten im Verlauf der vergangenen neun Jahre deutlich in ihrer mathematischen Kompetenz verbessern. Die größten Zuwächse sind hier für Mexiko und Polen zu verzeichnen, nämlich jeweils 28 Punkte, was eine Verbesserung der durchschnittlichen Mathematikleistungen in der Größenordnung eines Schuljahres bedeutet. Bemerkenswert sind aber auch die Fortschritte in der Türkei, Portugal und Italien, die von 25 bis 19 Punkten reichen. Statistisch signifikant sind außerdem die Zuwächse in Korea (12 Punkte) und in Deutschland (11 Punkte). Deutschland hat sich mit diesem Zuwachs signifikant über den OECD-Durchschnitt entwickelt. Berücksichtigt man zudem, dass Deutschland in der Mathematik bereits von PISA 2000 nach PISA 2003 deutliche Verbesserungen erzielen konnte (Blum et al., 2004), dann bedeuten die Veränderungen der vergangenen Jahre nicht nur eine Stabilisierung, sondern eine weitere Verbesserung der mathematischen Kompetenz.

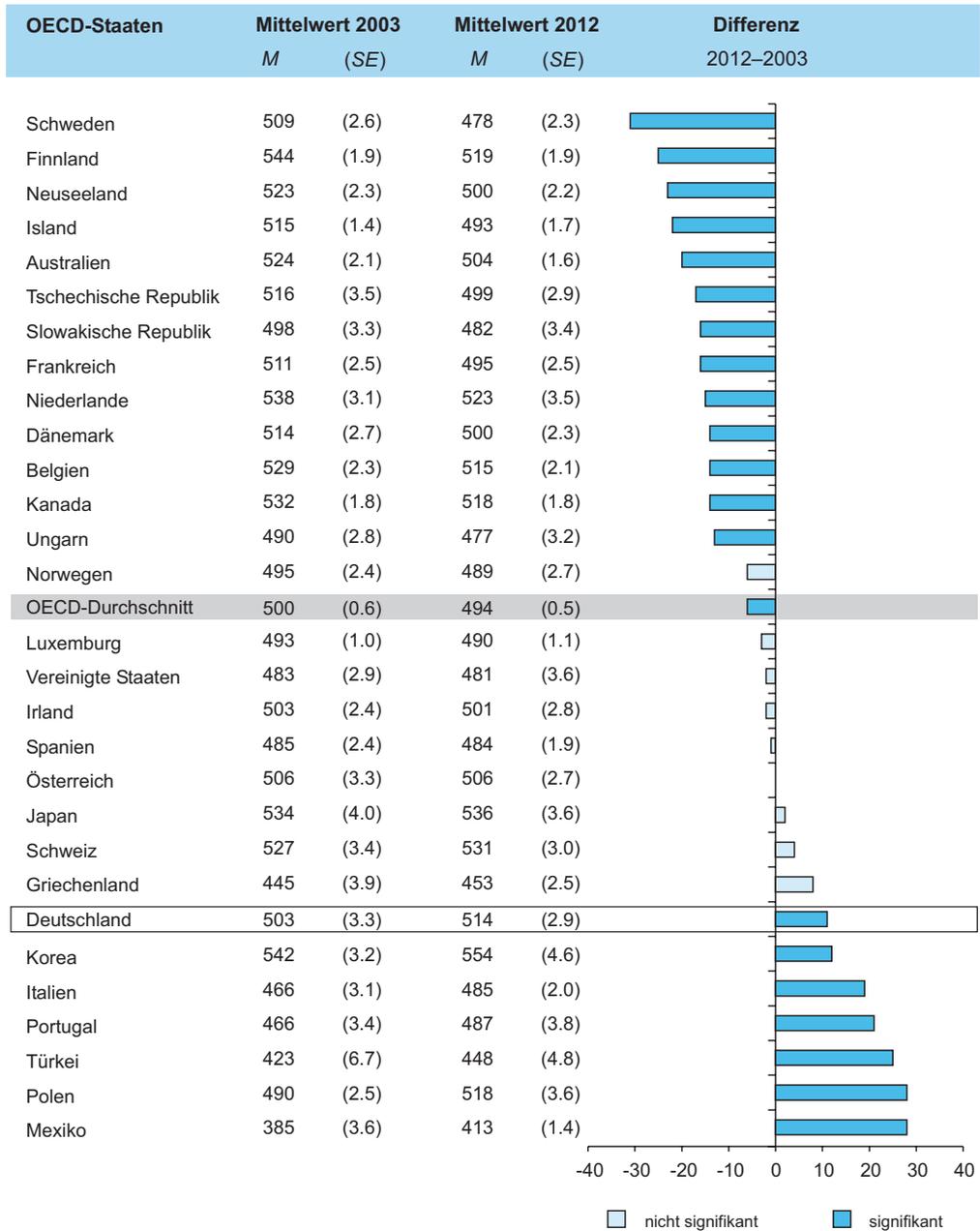


Abbildung 2.13: Mittelwerte mathematischer Kompetenz in OECD-Staaten in PISA 2003 und PISA 2012

Es hilft bei der Bewertung dieses Fortschrittes, wenn man sich vor Augen führt, dass sich viele Staaten in der Nachbarschaft Deutschlands in diesem Zeitraum verschlechtert haben. Selbst in Staaten, die für Deutschland aufgrund ihrer Schulstruktur immer wieder als vorbildlich dargestellt wurden (z. B. Kahl, 2003), haben sich die Bildungsergebnisse im Mathematikbereich in den vergangenen neun Jahren deutlich verschlechtert. Der Mythos nordischer Bildungssysteme (vgl. Kobarg & Prenzel, 2009) ist dabei ein weiteres Stück verblasst. Die Fortschritte, die hier für Deutschland berichtet werden können, dürften deshalb eher auf die Anstrengungen zur Qualitätssicherung und Unterrichtsentwicklung als auf das Bildungssystem an sich zurückzuführen sein.

2.5 Ergebnisse für Deutschland

2.5.1 Kompetenzen in verschiedenen Schularten

In diesem nationalen PISA-Bericht werden die Ergebnisse für fünf Schularten gesondert ausgewiesen, die sich allerdings in der Bildungsbeteiligung unterscheiden: Hauptschulen, Realschulen, Gymnasien, Integrierte Gesamtschulen sowie Schulen mit mehreren Bildungsgängen (vgl. Kapitel 1). Nicht eigens dargestellt werden die Ergebnisse der Stichproben fünfzehnjähriger Schülerinnen und Schüler an Förderschulen und Berufsschulen, denn es handelt sich hier nicht um repräsentative Stichproben für diese Schularten. Allerdings müssen die an diesen Schulen gemessenen Testleistungen in der Gesamtstichprobe der Fünfzehnjährigen berücksichtigt werden. Aufgrund größerer Veränderungen im Zuschnitt der Schularten in den Ländern können Veränderungen in den Mathematikleistungen seit den ersten PISA-Erhebungen mit Ausnahme des Gymnasiums nicht sinnvoll für die einzelnen Schularten ausgewiesen werden.

Die Abbildung 2.14 stellt die bei PISA 2012 gemessenen Mittelwerte auf der Gesamtskala Mathematik für die fünf genannten Schularten dar. Die Perzentilbänder veranschaulichen die Streuung innerhalb der Schularten und lassen Überlappungen der Leistungsverteilungen erkennen. In der Abbildung ist außerdem die Bildungsbeteiligung für die jeweilige Schulart angegeben (Prozent des Altersjahrgangs).

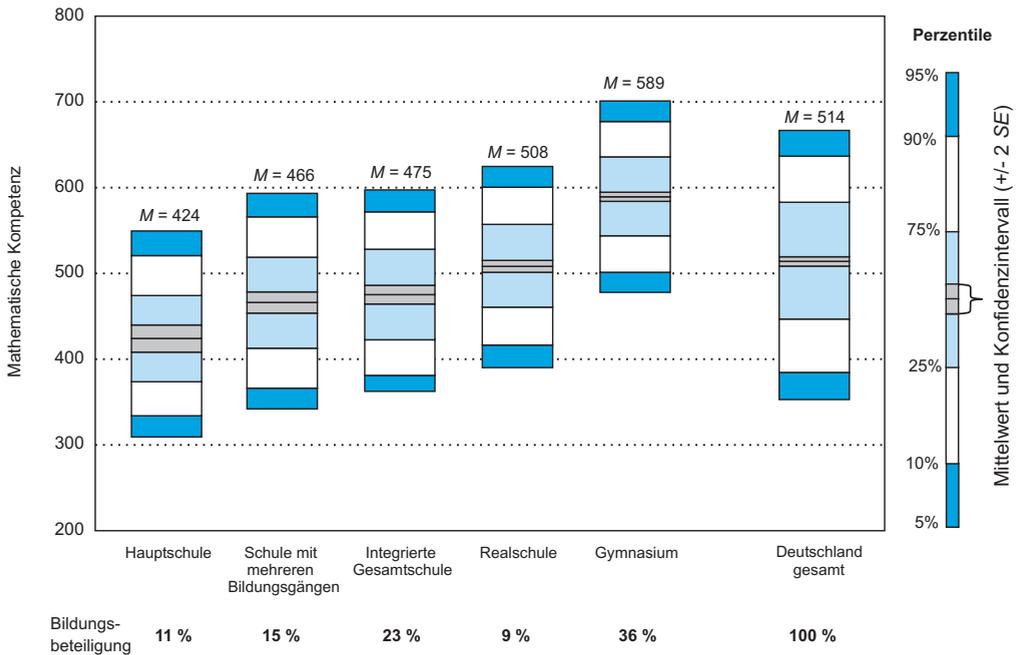


Abbildung 2.14: Perzentilbänder mathematischer Kompetenz in Deutschland nach Schulart

Wie zu erwarten ist, übertreffen die Gymnasien mit einem Mittelwert von 589 Punkten signifikant den Durchschnitt für Deutschland (514 Punkte) in der mathematischen Kompetenz. Der Mittelwert der Realschulen (508 Punkte) bewegt sich im Bereich des deutschen Durchschnitts. Die Mittelwerte für die Hauptschulen (424 Punkte), die Schulen mit mehreren Bildungsgängen (466 Punkte) und die Integrierten Gesamtschulen (475 Punkte) hingegen liegen signifikant unter dem Mittelwert für Deutschland. Obwohl die Mittelwertsunterschiede sehr große Leistungsabstände zwischen den Schularten belegen, deuten die Perzentilbänder doch auf bemerkenswerte Überlappungen in den Verteilungen. So zeigen zum Beispiel mehr als 10 Prozent der Jugendlichen an Realschulen höhere Leistungen als der Durchschnitt der Gymnasien. Mehr als 25 Prozent der Fünfzehnjährigen an Hauptschulen übertreffen den Mittelwert, der für die Integrierten Gesamtschulen gemessen wurde. Die leistungsstärksten 5 Prozent der Hauptschülerinnen und -schüler liegen mit ihrer mathematischen Kompetenz im mittleren Kompetenzbereich der Gymnasiastinnen und Gymnasiasten.

Betrachtet man die Verteilungen auf die Stufen mathematischer Kompetenz differenziert nach Schulart, dann interessieren insbesondere die Anteile auf und unter der ersten Kompetenzstufe (Tabelle 2.4). Diese Kompetenzstufe weist auf Probleme für die Ausbildungskarriere und für die weitere gesellschaftliche Teilhabe hin. Die Anteile von Fünfzehnjährigen auf dieser Stufe an den Hauptschulen (49.4 Prozent), Schulen mit mehreren Bildungsgängen (28.0 Prozent), Integrierten Gesamtschulen (23.3 Prozent) und selbst Realschulen (10.7 Prozent) sind allesamt deutlich zu hoch. Anlass zu Besorgnis

geben nicht nur die sehr großen Anteile an Hauptschulen (etwa jeder zweite Jugendliche), sondern auch die Anteile an den anderen Schularten und speziell den Realschulen. Hier besteht weiterhin dringender Handlungsbedarf, auch im Sinne eines Erreichens der Bildungsstandards, obschon sich ja insgesamt die Anteile der Schülerinnen und Schüler auf und unter der ersten Kompetenzstufe seit PISA 2003 in Deutschland um nennenswerte vier Prozentpunkte verringert haben.

Die Anteile der Schülerinnen und Schüler auf den Kompetenzstufen V und VI, die eine besonders starke mathematische Kompetenz erkennen lassen, verteilen sich erwartungsgemäß unterschiedlich auf die Schularten. An den Gymnasien können 40.3 Prozent der Fünfzehnjährigen diesem sehr hohen Leistungsniveau zugeordnet werden. An den Realschulen erreichen 8.3 Prozent der Schülerinnen und Schüler dieses sehr hohe Leistungsniveau, die Anteile an den Schulen mit mehreren Bildungsgängen und an den Integrierten Gesamtschulen bewegen sich in einer Größenordnung um 3.5 Prozent.

Tabelle 2.4: Prozentuale Anteile von Schülerinnen und Schülern auf den Stufen mathematischer Kompetenz in Deutschland nach Schulart

Schulart		Kompetenzstufe						
		Unter I	I	II	III	IV	V	VI
Hauptschule	%	17.1	32.3	28.0	16.3	5.7	0.5	0.1
	(SE)	(1.9)	(4.5)	(3.0)	(4.0)	(1.4)	(0.5)	(0.0)
Schule mit mehreren Bildungsgängen	%	8.5	19.5	31.9	25.1	11.7	2.8	0.5
	(SE)	(1.4)	(2.5)	(2.3)	(2.1)	(2.2)	(0.9)	(0.1)
Integrierte Gesamtschule	%	4.5	18.8	32.9	25.1	15.0	3.6	0.1
	(SE)	(1.2)	(1.9)	(2.7)	(2.8)	(2.5)	(0.9)	(0.1)
Realschule	%	1.8	8.9	25.7	33.0	22.3	7.2	1.1
	(SE)	(0.4)	(0.7)	(1.9)	(1.8)	(2.3)	(0.8)	(0.2)
Gymnasium	%	0.0	0.8	5.1	20.1	33.7	28.0	12.3
	(SE)	(0.0)	(0.3)	(0.6)	(1.3)	(1.8)	(1.7)	(0.7)
Gesamtstichprobe	%	5.5	12.2	19.4	23.7	21.7	12.8	4.7
	(SE)	(0.6)	(0.7)	(0.9)	(0.8)	(1.2)	(0.7)	(0.3)

Kompetenzunterschiede zwischen Mädchen und Jungen

Bei der Verteilung auf die Kompetenzstufen sind außerdem Geschlechterdifferenzen zu beobachten. Deutschlandweit befinden sich 18,7 Prozent der Mädchen auf Stufe I und darunter, bei den Jungen sind es 16,8 Prozent. Auf den beiden höchsten Kompetenzstufen (V und VI) wiederum fällt der Anteil der Jungen höher aus: 14,9 Prozent der Mädchen erreichen die Stufen V und VI; für die Jungen beträgt der Anteil 20,0 Prozent (vgl. Abbildung 2.15 und Abschnitt 2.4.3).

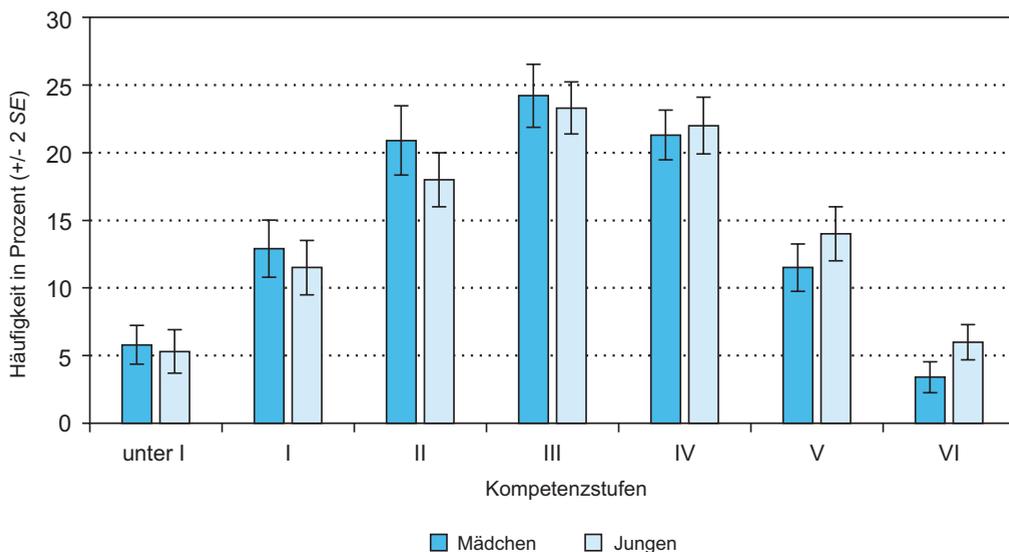


Abbildung 2.15: Prozentuale Anteile von Schülerinnen und Schülern auf den Stufen der mathematischen Kompetenz in Deutschland

2.5.2 Mathematische Kompetenz in Deutschland im Trend betrachtet

Während der OECD-Durchschnitt bei der mathematischen Kompetenz über die Messzeitpunkte hinweg seit PISA 2003 gesunken ist (zum Teil bedingt durch eine veränderte Staatszusammensetzung), zeigt der Verlauf in Deutschland eine kontinuierliche Verbesserung für die Gesamtstichprobe an. Im Vergleich zu PISA 2003 liegt der Mittelwert der mathematischen Kompetenz in PISA 2012 signifikant höher.

Seit PISA 2003 hat sich die Schullandschaft auf der Ebene der Sekundarstufe in Deutschland verändert, wie die Vielzahl von Bezeichnungen für Schulformen in den Ländern andeutet. Einige Länder haben auf ein zweigliedriges Schulsystem umgestellt, das neben dem Gymnasium eine bestimmte Variante einer „Schule mit mehreren Bildungsgängen“ anbietet. In anderen Ländern existieren neben dem Gymnasium zwei oder gar drei weitere Schularten (vgl. Pant et al., 2013). Ergänzt werden diese allgemei-

nen Angebote (für die PISA-Zielgruppe der Fünfzehnjährigen) durch Förderschulen und berufliche Schulen, die unter bestimmten Voraussetzungen besucht werden (vgl. Kapitel 1). Diese sich ständig wandelnden Konstellationen im Sekundarschulbereich erschweren eine Betrachtung der Veränderungen in der mathematischen Kompetenz für die einzelnen Schularten seit PISA 2003. Nach wie vor gibt es jedoch in allen Ländern in Deutschland eine Schulart, die sich einer gemeinsamen Bildungstradition verpflichtet sieht und nach unterschiedlich geregelten Übertrittsverfahren die Teilgruppe der leistungsstärksten Schülerinnen und Schüler versammelt: das Gymnasium. Damit besteht die Möglichkeit, neben der Entwicklung für Deutschland insgesamt, auch Veränderungen in der mathematischen Kompetenz in Bezug auf die Gymnasien nachzuzeichnen. Dabei muss berücksichtigt werden, dass der Anteil an Schülerinnen und Schülern, die nach der Grundschule ein Gymnasium besuchen, seit Jahrzehnten stetig ansteigt. Im Zeitraum zwischen PISA 2003 und PISA 2012 hat dieser Anteil um fast 5 Prozentpunkte zugenommen. Eine steigende Bildungsbeteiligung sollte sich dabei in den Testresultaten niederschlagen. Wenn anstelle von 31.3 Prozent eher leistungsstarken Fünfzehnjährigen (PISA 2003) neun Jahre später 36.0 Prozent des gleichen Altersjahrgangs verglichen werden, so ist damit zu rechnen, dass der Mittelwert in einem Leistungstest aufgrund der erweiterten Zusammensetzung sinkt – es sei denn, das Leistungsniveau hätte aus anderen Gründen in diesem Zeitraum zugenommen. Diesen Fragen wird im Folgenden nachgegangen.

In einem ersten Schritt stellt Abbildung 2.16 (unbereinigt um Effekte einer veränderten Zusammensetzung) dar, wie sich von PISA 2003 bis PISA 2012 der Gesamtmittelwert der mathematischen Kompetenz einerseits für Deutschland insgesamt und andererseits für die Gymnasien in Deutschland verändert hat. Die Ausgangswerte in PISA 2000 (Deutschland 490 Punkte, Gymnasien 579 Punkte) lassen sich nicht direkt in die statistische Trendbetrachtung einbeziehen, da diese beim Zeitpunkt der ersten Erhebung als Hauptdomäne startete (PISA 2003) und den Mittelwert der OECD-Staaten nach PISA 2000 erneut auf 500 Punkte normierte. Während der Verlauf der Kurve für Deutschland seither insgesamt eine (statistisch abgesicherte) Verbesserung der mathematischen Kompetenz erkennen lässt (um 11 Punkte), nimmt die Leistung für die Gymnasien tendenziell ab (seit PISA 2003 um 5 Punkte, nämlich von 594 auf 589 Punkte). Somit scheint es, als könne die Leistungssteigerung der letzten Jahre in Deutschland nicht den Gymnasien zugerechnet werden. Betrachtet man die Entwicklungen im Detail, dann verdienen in diesem Zusammenhang zwei Ergebnisse besondere Aufmerksamkeit. Zum einen hat sich der Anteil leistungsschwacher Schülerinnen und Schüler (auf der Kompetenzstufe I und darunter) seit PISA 2003 von 21.6 Prozent auf 17.7 Prozent in PISA 2012 verringert. Es ist unwahrscheinlich, dass dieser Kompetenzzuwachs im unteren Bereich am Gymnasium stattgefunden hat. Plausibler ist, dass hier gerade die Schularten mit einem hohen Anteil leistungsschwacher Schülerinnen und Schüler gute Arbeit geleistet haben. Zum anderen hat sich der Anteil von Schülerinnen und Schülern auf den Kompetenzstufen V und VI für Deutschland seit 2003 leicht erhöht (von 16.3 Prozent auf 17.5 Prozent). Hier wird man zunächst an die Gymnasien denken, doch ist der

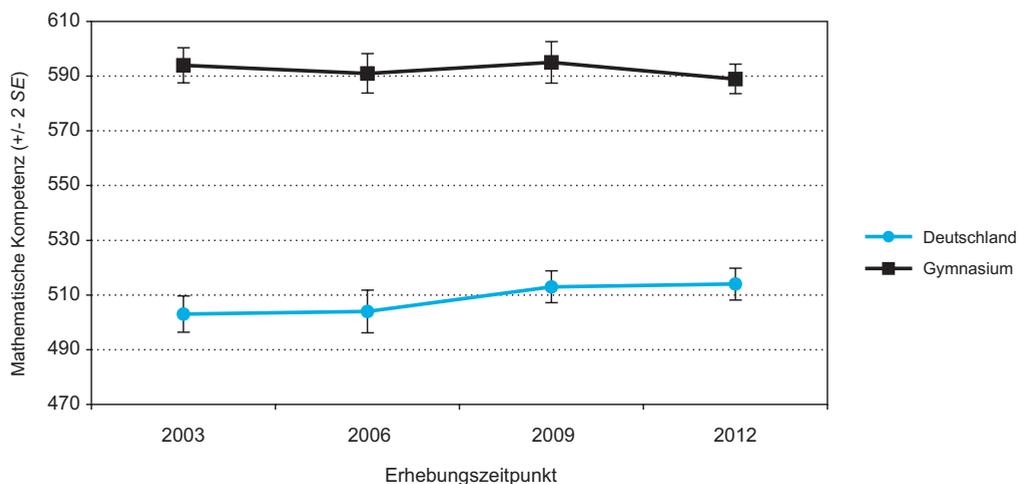


Abbildung 2.16: Mathematische Kompetenz in Deutschland und an Gymnasien von PISA 2003 bis PISA 2012

relative Anteil dieser Spitzengruppe in dieser Schulart von 42.6 Prozent (PISA 2003) auf 40.3 Prozent (PISA 2012) gesunken. Dieses uneinheitliche Bild könnte schlicht eine Folge der höheren Übergangsquoten an das Gymnasium sein.

Um also der Frage nachzugehen, ob die Gymnasien mehr oder weniger zur Verbesserung der mathematischen Kompetenz seit PISA 2003 beigetragen haben, kann man versuchen, bei PISA 2012 die 5 Prozent Schülerinnen und Schüler „herauszurechnen“, die durch die steigende Bildungsbeteiligung seit PISA 2003 zusätzlich das Gymnasium besuchen. Für diese Abschätzung wird (konservativ und zugunsten der Schulform) unterstellt, dass die hinzugekommenen Schülerinnen und Schüler im Vergleich zur sonstigen Gymnasialpopulation leistungsschwach sind. Dieser Annahme folgend kann die Gymnasialstichprobe um die Schülerinnen und Schüler der unteren 5 Perzentile der Leistungsverteilung reduziert werden, die maximal 478 Punkte im Mathematiktest erreicht haben. Diese künstliche Reduktion der Gymnasialbeteiligung durch die leistungsschwächsten Schülerinnen und Schüler von 36 auf 31 Prozent erhöht den Mittelwert für das Gymnasium auf 596 Punkte. Dieser um die Bildungsbeteiligung korrigierte Gymnasialmittelwert für PISA 2012 unterscheidet sich nicht signifikant von dem in PISA 2003 gemessenen Mittelwert für die Gymnasien von 594 Punkten. Sicherlich stellt die steigende Bildungsbeteiligung genauso wie etwa die Umstellung von neun auf acht Schuljahre (das sogenannte G8) die Gymnasien vor neue Anforderungen. Vor diesem Hintergrund gilt es, weitere Entwicklungen in diesem Bereich sorgfältig zu beobachten.

Diese Modellrechnung verstärkt den Eindruck, dass Zuwächse der mathematischen Kompetenz in Deutschland seit PISA 2003 weitgehend in den Schulen außerhalb des Gymnasiums entwickelt wurden. Der Leistungsstand der Gymnasien in PISA 2012 unterscheidet sich offensichtlich nicht vom Leistungsstand in PISA 2003. Die Fortschritte, die in Deutschland – ablesbar an den signifikant höheren Mittelwerten – erzielt

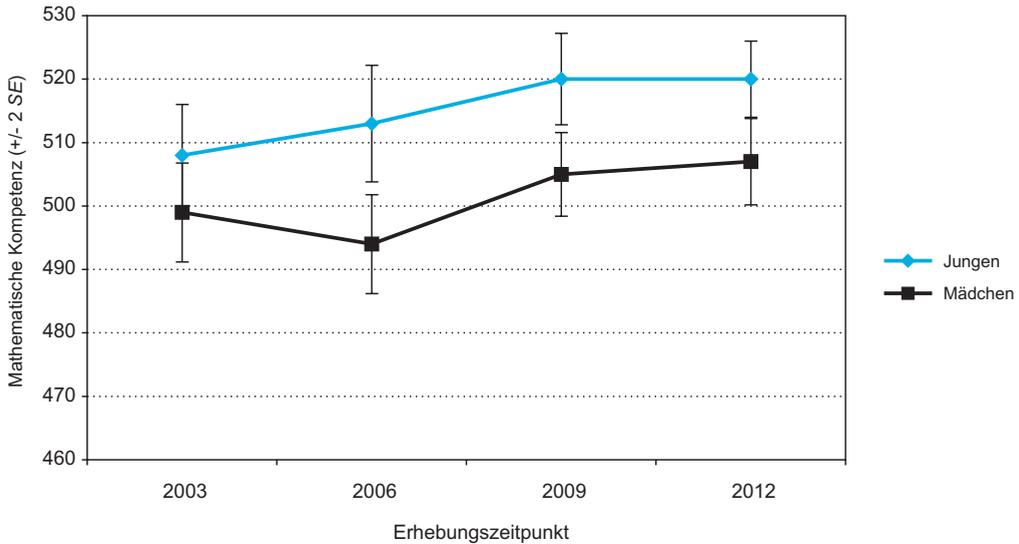


Abbildung 2.17: Mathematische Kompetenz von Jungen und Mädchen in Deutschland von PISA 2003 bis PISA 2012

wurden, beruhen demnach insbesondere auf einer besseren Förderung schwächerer Schülerinnen und Schüler. Es könnte also durchaus ein Ziel der kommenden Jahre sein, auch leistungsstarke Schülerinnen und Schüler noch besser zu fördern.

In Abbildung 2.17 wird der Blick nochmals auf den Aspekt der Geschlechterdifferenzen in Deutschland gerichtet. In PISA 2003 und 2006 befand sich die durchschnittliche mathematische Kompetenz von Schülerinnen und Schülern in Deutschland im Bereich des OECD-Durchschnitts. In PISA 2009 lag dieser Wert für Deutschland bereits signifikant über dem OECD-Mittelwert, was in PISA 2012 konsolidiert werden konnte. Bemerkenswert hierbei ist, dass sich in Deutschland die Jungen im Vergleich zu PISA 2003 signifikant verbessert haben (um 12 Punkte von 508 auf 520 Punkte), während die Mädchen in PISA 2012 zwar auch 8 Punkte mehr erreichen als in PISA 2003 (von 499 auf 507 Punkte), diese Differenz jedoch nicht statistisch signifikant ist (Abbildung 2.17). Im OECD-Durchschnitt haben sich im Vergleich zu PISA 2003 sowohl Jungen als auch Mädchen signifikant verschlechtert.

2.6 Zusammenfassung und Diskussion

Wie bereits bei PISA 2009, so lässt sich auch in PISA 2012 feststellen, dass die Schülerinnen und Schüler in Deutschland in Bezug auf ihre mathematische Kompetenz signifikant über dem OECD-Durchschnitt liegen. Sie knüpfen damit an die stetige Verbesserung an, die sich seit PISA 2003 abzeichnet. Insbesondere im Vergleich zu dieser Studie, bei der Mathematik schon einmal die Hauptdomäne war, zeigen die Fünfzehnjährigen

in Deutschland im Durchschnitt eine deutlich höhere mathematische Kompetenz. Die Anstrengungen um eine verbesserte Bildung insbesondere im letzten Jahrzehnt haben sich folglich gelohnt. Es wäre allerdings nicht richtig anzunehmen, dass die bisherigen Maßnahmen ausreichen. Der detaillierte Blick auf die Ergebnisse von PISA 2012 offenbart genau wie in den Studien zuvor Schwächen im deutschen Bildungssystem, die es in den kommenden Jahren weiter anzugehen gilt.

Zunächst ist es eine gute Nachricht, dass zwischen den vier Inhaltsbereichen *Veränderung und Beziehungen*, *Quantität*, *Raum und Form* sowie *Unsicherheit und Daten* nur geringe Unterschiede in der durchschnittlich erreichten Kompetenz der Jugendlichen bestehen. Während bei PISA 2003 der Unterschied zwischen dem am besten bearbeiteten Inhaltsbereich *Quantität* und dem wesentlich schwächeren Inhaltsbereich *Unsicherheit und Daten* bei 21 Punkten lag, trennen 2012 nur noch 10 Punkte die beiden Extreme *Quantität* und *Raum und Form*. Mittlerweile beherrschen die Schülerinnen und Schüler also Teilgebiete der Mathematik konsistenter als noch neun Jahre zuvor. Darüber hinaus gibt es in allen inhaltlichen Bereichen einen Zuwachs an Punkten gegenüber PISA 2003, der allerdings nur bei *Veränderung und Beziehungen*, *Raum und Form* sowie *Unsicherheit und Daten* deutlich ist. Er fällt in Bezug auf *Unsicherheit und Daten* mit 16 Punkten am größten aus. Es ist plausibel, diese Entwicklungen mit der seit etlichen Jahren – unter anderem durch die Bildungsstandards – verstärkten Rolle von Daten, Wahrscheinlichkeitsrechnung und Statistik im Mathematikunterricht zu verbinden. Durch ihre Einführung wurde insbesondere der Inhaltsbereich *Unsicherheit und Daten* gestärkt, dem vorher in der Schule nur wenig Aufmerksamkeit geschenkt wurde. Noch immer werden allerdings Aufgaben aus den Inhaltsbereichen *Raum und Form* sowie *Unsicherheit und Daten* nicht ganz so gut gelöst wie Aufgaben, die den Inhaltsbereichen *Veränderung und Beziehungen* beziehungsweise *Quantität* zuzuordnen sind. Bedauerlich ist insbesondere, dass die Leistungen der Schülerinnen und Schüler in Bezug auf *Raum und Form* weniger angehoben werden konnten. Gerade in diesem Inhaltsbereich lässt sich besonders gut zeigen, dass die Mathematik ein Kulturgut ist und Werkzeuge für die Lösung aktueller Probleme auf einer Tausende Jahre alten Grundlage bereitstellen kann.

Die Abnahme des Anteils von sehr leistungsschwachen Schülerinnen und Schülern, also solchen, deren Kompetenz die Stufe I nicht überschreitet, ist ebenfalls ein erfreulicher Befund für Deutschland. Dabei liegt die Streuung der Leistungen nur unwesentlich über dem Durchschnitt der OECD-Länder. Der Anteil der sogenannten Risikogruppe (Kompetenzstufe I und darunter) ist in Deutschland mit 17,7 Prozent allerdings in etwa gleich groß wie der Anteil sehr guter bis exzellenter Schülerinnen und Schüler (Kompetenzstufen V und VI) mit 17,5 Prozent. Trotzdem muss man feststellen, dass gerade in den leistungsstarken europäischen Ländern der Anteil der Fünfzehnjährigen auf den beiden obersten Kompetenzstufen höher als derjenige in der Risikogruppe ist, so etwa in den Niederlanden, Polen, der Schweiz oder Estland. Es bleibt das Ziel, den Anteil schwächerer Schülerinnen und Schüler in Deutschland weiter zu verringern und

gleichzeitig den Anteil der Schülerinnen und Schüler auf den beiden höchsten Kompetenzstufen zu vergrößern.

Die Dringlichkeit dieser Forderung zeigt sich besonders, wenn man die mathematische Kompetenz von Jungen und Mädchen in Deutschland vergleicht und sich die verschiedenen Schulformen dabei mit ansieht. Die Differenz zwischen Mädchen und Jungen auf der Gesamtskala liegt mit 14 Punkten signifikant über der in den OECD-Staaten durchschnittlichen Geschlechterdifferenz und entspricht in etwa einem Drittel eines Schuljahres. Insbesondere an Gymnasien haben die Jungen gegenüber den Mädchen in Mathematik einen markanten Vorsprung. Der Anteil der Schülerinnen an Gymnasien auf den Kompetenzstufen V und VI (32.6 Prozent) liegt klar unter dem Anteil der Schülerinnen an Gymnasien (53.9 Prozent).

Dieser Aspekt ist auch deshalb besonders brisant, weil sich anhand der Kompetenzstufenverteilung ablesen lässt, dass die Verbesserung der durchschnittlichen mathematischen Kompetenz weniger auf die hohen Werte der Schülerinnen und Schüler an Gymnasien zurückzuführen ist als auf die gestiegenen Punktzahlen der schwächeren Schülerinnen und Schüler. Insofern muss der Blick nun auch in die Gymnasien und die dortige Förderung der potentiell leistungsstärksten Fünfzehnjährigen gehen, damit diese Gruppe sich in Zukunft gut entwickeln kann und entsprechende Unterstützung ihrer Talente erhält. Mit 36 Prozent ist die Schülerschaft der Gymnasiasten die größte Gruppe; das Gymnasium wird als weiterführende Schulform von immer mehr Jugendlichen und deren Eltern angestrebt. Es wäre daher zu kurz gedacht, sich auf den konstant hohen Mittelwert mathematischer Kompetenz in den PISA-Studien zurückzuziehen und sich damit zufrieden zu geben. Des Weiteren könnte dabei insbesondere die Förderung von Mädchen und Jungen gleichermaßen als ein wichtiger Auftrag an diese Schulart gesehen werden.

Die Daten aus PISA 2012 zeigen deutliche Fortschritte in Richtung der Ziele, die 2003 mit der Einführung der Bildungsstandards angestrebt wurden. Möglicherweise hat die Einführung von Bildungsstandards in den Jahren 2003 und 2004 dazu beigetragen, dass sich die Kompetenz in der Mathematik bei Schülerinnen und Schülern in Deutschland seither positiv entwickelt hat. Das zeigt sich insbesondere bei Jugendlichen aus nicht gymnasialen Bildungsgängen, die deutlich aufgeholt haben. Auch wenn PISA eher das Konzept der mathematischen Grundbildung betont und die Bildungsstandards den breiten Blick auf anschlussfähige Lernergebnisse haben, gibt es viele Gemeinsamkeiten in der grundsätzlichen Ausrichtung. In beiden Fällen geht es darum, Mathematik als eine Wissenschaft mit zentraler Bedeutung für technische, natürliche, soziale und kulturelle Prozesse zu sehen. Sowohl PISA als auch die Bildungsstandards stellen kreatives Problemlösen in ganz unterschiedlichen Kontexten in den Vordergrund und betonen so einen wesentlichen Aspekt mathematischen Arbeitens. Insofern sind die Ergebnisse des Vergleichs der Bundesländer, die im Oktober 2013 veröffentlicht wurden (Pant et al., 2013), eine wichtige Ergänzung von PISA. Die vorliegende PISA-Studie zeigt, dass die Verteilung der Leistungen in Deutschland im internationalen Vergleich eher unauffällig ist, der sogenannte Ländervergleich macht allerdings deutlich, dass es dabei erhebliche

regionale Unterschiede gibt (Pant et al., 2013). Teilweise liegen zwischen Bundesländern mit den höchsten und den niedrigsten Kompetenzwerten Differenzen von zwei Schuljahren (Pant et al., 2013).

Es ist keine Frage, dass alle Jugendlichen in Deutschland einen Mathematikunterricht angeboten bekommen sollten, der die Entwicklung ihrer mathematischen Kompetenz optimal unterstützt. Diese Aussage gilt für Jugendliche in allen Schulformen und für Jungen und Mädchen gleichermaßen. Damit muss weiterhin die Verringerung der Disparitäten ein wesentliches Ziel von Bildungsanstrengungen in Deutschland sein. Mit den bundeseinheitlichen Bildungsstandards für Mathematik ist ein guter Anfang gemacht, der von den verschiedenen mit Bildung befassten Institutionen noch konsequenter aufgegriffen und weitergeführt werden sollte. Eine solide mathematische Bildung, die kulturelle und historische Aspekte genauso wie Methoden des Problemlösens und Anwendungen umfasst, ist eine unverzichtbare Grundlage für eine wirksame gesellschaftliche Teilhabe.

Literatur

- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W. et al. (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Lehmann, R., Lehrke, M. S. B., Clausen, M., Hosenfeld, I., Köller, O. et al. (Hrsg.). (1997). *TIMSS – Mathematisch-Naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- BLK Projektgruppe Innovationen im Bildungswesen. (1997). *Gutachten zur Vorbereitung des Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“* (Materialien zur Bildungsplanung und zur Forschungsförderung, Bd. 60). Bonn: Bund-Länder-Kommission, Geschäftsstelle.
- Blum, W. (2002). Applications and modelling in mathematics education. Discussion document of ICME-Study No. 14. *Educational Studies in Mathematics*, 51, 149–171.
- Blum, W., Drüke-Noe, K., Hartung, R. & Köller, O. (Hrsg.). (2010). *Bildungsstandards Mathematik konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsideen und Fortbildungsmöglichkeiten* (4., vollständig überarbeitete Auflage). Berlin: Cornelsen/Scriptor.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfing, F. et al. (2004). Mathematische Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs* (S. 47–92). Münster: Waxmann.
- Brehl, T., Wendt, H. & Bos, W. (2012). Geschlechtsspezifische Unterschiede in mathematischen und naturwissenschaftlichen Kompetenzen. In W. Bos, H. Wendt, O. Köller & C. Selzer (Hrsg.), *Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 203–230). Münster: Waxmann.
- de Lange, J. (1987). *Mathematics, insight and meaning*. Utrecht: OW and OC, Utrecht University.

- Ehmke, T., Blum, W., Neubrand, M., Jordan, A. & Ulfig, F. (2006). Wie verändert sich die mathematische Kompetenz von der neunten zur zehnten Klassenstufe? In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 63–86). 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 et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 153–176). Münster: Waxmann.
- Glei, R. (2006). *Die sieben freien Künste in Antike und Gegenwart*. Trier: Wissenschaftlicher Verlag Trier.
- Hartig, J. & Frey, A. (2012). Validität des Tests zur Überprüfung des Erreichens der Bildungsstandards in Mathematik: Zusammenhänge mit den bei PISA gemessenen Kompetenzen und Varianz zwischen Schulen und Schulformen. *Diagnostica*, 58, 3–14.
- Kahl, R. (2003, 10. Juli). Worauf warten wir noch? *Die ZEIT*, 29.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards. Expertise*. Bonn: BMBF.
- Klieme, E. & Baumert, J. (Hrsg.). (2001). *TIMSS – Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente*. Bonn: BMBF.
- Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 139–190). Opladen: Leske + Budrich.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik. (2003). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss (10. Jahrgangsstufe). Beschluss vom 4.12.2003* (Beschlüsse der Kultusministerkonferenz), Bonn.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik. (2004). *Bildungsstandards im Fach Mathematik für den Hauptschulabschluss (Jahrgangsstufe 9). Beschluss vom 15.10.2004* (Beschlüsse der Kultusministerkonferenz), Bonn.
- Kobarg, M. & Prenzel, M. (2009). Stichwort: Der Mythos der nordischen Bildungssysteme. *Zeitschrift für Erziehungswissenschaft*, 4, 597–615.
- Neubrand, M. & Klieme, E. (2002). Mathematische Grundbildung. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele et al. (Hrsg.), *PISA 2000 – die Länder der Bundesrepublik Deutschland im Vergleich* (S. 95–128). Opladen: Leske + Budrich.
- Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM project. In A. Gagatsis & S. Papstavridis (Hrsg.), *3rd Mediterranean Conference on Mathematical Education* (S. 115–124). Athens, Greece: Hellenic Mathematical Society and Cyprus Mathematical Society.
- Niss, M. & Højgaard, T. (Hrsg.). (2011). *Competencies and mathematical learning: Ideas and inspiration for the development of mathematics teaching and learning in Denmark (English edition)*. Roskilde, Denmark: IMFUFA.
- OECD. (1999). *Measuring student knowledge and skills. A new framework for assessment*. Paris: OECD.

- OECD. (2004). *The PISA 2003 assessment framework. Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publishing. Zugriff am 18.09.2013. Verfügbar unter <http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33694881.pdf>
- OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M. et al. (Hrsg.). (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Prenzel, M., Friedrich, A. & Stadler, M. (Hrsg.). (2009). *Von SINUS lernen – Wie Unterrichtsentwicklung gelingt*. Seelze-Velber: Kallmeyer.
- Turner, R. & Adams, R. (2012). Some drivers of test item difficulty in mathematics: an analysis of the competency rubric. *OECD Programme for International Student Assessment (PISA)*. Zugriff am 01.10.2013. Verfügbar unter <http://research.acer.edu.au/pisa/7>
- Winter, H. (1995). Mathematikunterricht und Allgemeinbildung. *Mitteilungen der Gesellschaft für Didaktik der Mathematik*, 61, 37–46.
- Wößmann, L. & Piopiunik, M. (2009). *Was unzureichende Bildung kostet – Eine Berechnung der Folgekosten durch entgangenes Wirtschaftswachstum*. Studie im Auftrag der Bertelsmann Stiftung. Gütersloh: Bertelsmann Stiftung.

3 **Mathematikbezogene emotionale und motivationale Orientierungen, Einstellungen und Verhaltensweisen von Jugendlichen in PISA 2012**

Anja Schiepe-Tiska und Stefanie Schmidtner

Der Bildungsauftrag der Schule beschränkt sich nicht nur auf die Vermittlung von Wissen und kognitiven Fähigkeiten. Sie soll darüber hinaus junge Menschen stärken und sie darauf vorbereiten, ihr zukünftiges Leben selbstverantwortlich zu gestalten (Fend, 2008; vgl. auch Tenorth, 1994). Das Leben innerhalb und in zunehmendem Maße auch das außerhalb der Schule verlangt unter anderem, sich Herausforderungen zu stellen, Vertrauen in die eigenen Fähigkeiten zu erlangen und Arbeitshaltungen zu erlernen, die helfen, mit den Anforderungen der Umwelt umzugehen. Solche Orientierungen, Einstellungen und Verhaltensweisen basieren auf kulturellen und sozialen Normen einer Gesellschaft, die vor allem in der Schule vermittelt werden. Damit legt die Schule die Basis für ein lebenslanges Lernen, das nicht nur fundiertes Wissen, sondern auch positive Einstellungen und Verhaltensweisen gegenüber Lernen verlangt. Durch das Herausbilden solcher Einstellungen und Verhaltensweisen sind Jugendliche gerüstet, sich auf die Herausforderungen und den zukünftigen Wandel einzulassen und damit selbstbewusst und verantwortlich umzugehen.

Um zu unterstreichen, dass Schule mehr als kognitive Ziele verfolgt, wird von mehrdimensionalen Bildungszielen gesprochen, die auch in Vergleichsstudien Berücksichtigung finden (vgl. Prenzel, 2011). Solche Ziele sind in Lehrplänen wie in Schulprogrammen verankert; sie können fächerübergreifend angelegt oder auf ein spezifisches Fach wie Mathematik bezogen sein (vgl. Kunter, 2005). Mit Mathematik als Hauptdomäne der aktuellen PISA-Erhebung interessieren im Folgenden besonders solche Einstellungen und Verhaltensweisen, die einen Bezug zur Mathematik haben.

Eine ausschlaggebende Funktion für die Entwicklung positiver mathematikbezogener Orientierungen, Einstellungen und Verhaltensweisen hat der Mathematikunterricht, da dieser entscheidende Lerngelegenheiten für die Auseinandersetzung mit Mathematik bereitstellt (vgl. Kapitel 4). Hier können – und sollten – Jugendliche Freude an Mathematik erleben und deren Relevanz für ihr eigenes Leben und ihre Zukunft erfahren. Sie finden heraus, was sie sich selbst in Bezug auf ihre mathematischen Kompetenzen zutrauen können, und haben Gelegenheit, eine zielstrebige und gewissenhafte Arbeitshaltung zu entwickeln. Diese Bedingungen beeinflussen ihre Bereitschaft, sich

auch außerhalb der Schule und über die Schulzeit hinaus mit Mathematik auseinandersetzen zu wollen. Die verschiedenen Ziele – Wissen und Kompetenzen aufzubauen und gleichzeitig Freude an Mathematik und eine positive Selbstwahrnehmung zu entwickeln – beeinflussen sich wechselseitig, denn Lernen ist ein aktiver Prozess (KMK, 2010; Pekrun & Zirngibl, 2004; Reiss & Hammer, 2013). Es gilt, das Lernangebot zu nutzen, um Wissen und Kompetenzen weiterzuentwickeln. Ob und wie Schülerinnen und Schüler jedoch dieses Angebot wahrnehmen, hängt von ihren motivationalen Orientierungen und Einstellungen ab. Jugendliche, die Freude an Mathematik haben und ihre Bedeutung für das eigene Leben schätzen, nutzen die angebotenen Lerngelegenheiten häufiger und intensiver als Jugendliche, die sich wenig kompetent oder gar hilflos fühlen, wenn sie sich mit Mathematik auseinandersetzen (Helmke, 2007). Emotionale wie motivationale Orientierungen und Einstellungen können deshalb nicht nur als mehrdimensionale Bildungsziele, sondern auch als Bedingungs- oder Prozessvariablen für die Leistungsentwicklung betrachtet werden (Ma & Kishor, 1997).

Mitunter können die verschiedenen Ziele in Konkurrenz zueinander stehen. Das ist zum Beispiel dann der Fall, wenn die Entwicklung mathematischer Kompetenz nur unter äußerem Druck erfolgt und mit einem Verlust an Freude und Interesse einhergeht. Doch was nützen letztlich hoch ausgeprägte mathematische Fähigkeiten, wenn sich junge Menschen nicht mehr mit Mathematik befassen wollen? Ein Auseinanderklaffen von Wissen und Interesse ist besonders für die Nachwuchsförderung problematisch. Vor allem Mädchen entscheiden sich seltener für Ausbildungsgänge und Berufskarrieren, die ein mathematisches Verständnis voraussetzen (Wang, Eccles & Kenny, 2013). Um diese Bereitschaft zu fördern, ist es hilfreich, wenn die Schule und der Mathematikunterricht dazu anregen, sich aus eigenen Stücken mit Mathematik zu beschäftigen.

Die Bedeutung mehrdimensionaler Bildungsziele wird auch bei PISA 2012 berücksichtigt. Dabei werden Merkmale, die sich auf diese Ziele beziehen, primär mit Hilfe von Fragebögen erfasst. Als Ausgangspunkt für die Entwicklung der Fragen diente eine theoretische Rahmenkonzeption (OECD, 2013a). Diese wird im Folgenden kurz vorgestellt und es wird die Frage beantwortet, wie die mathematikbezogenen emotionalen und motivationalen Orientierungen sowie Einstellungen und Verhaltensweisen bei fünfzehnjährigen Jugendlichen in Deutschland ausgeprägt sind. Dabei werden die Ergebnisse in den internationalen Vergleich eingeordnet und es wird betrachtet, wie sich einige Merkmale seit 2003 entwickelt haben. Eine Untersuchung der Zusammenhänge der Schülermerkmale untereinander rundet die Darstellung der Ergebnisse ab.

3.1 Die Rahmenkonzeption der PISA-Fragebögen 2012

Bezug nehmend auf Theorien und Befunde der empirischen Bildungsforschung differenziert die Rahmenkonzeption für die PISA-Fragebögen zwischen Ausgangsbedingungen, Prozessmerkmalen sowie Bildungszielen (vgl. Klieme & Vieluf, 2013) und orientiert

sich an sogenannten Kontext-Input-Prozess-Outcome-Modellen (z. B. Purves, 1987). Die Konzeption unterscheidet drei Aggregationsebenen, nämlich die Ebenen *Schulen*, *Klassen* sowie *Schülerinnen und Schüler*. Die im Fragebogen erfassten Merkmale werden, je nach Fragestellung, einer *Input*-, *Prozess*- oder *Ergebnisebene* zugeordnet (OECD, 2013a). Auf der individuellen Ebene (der Ebene der Schülerinnen und Schüler) können zum Beispiel Hintergrundvariablen und strukturelle Bedingungen wie der soziale und familiäre Hintergrund (vgl. Kapitel 8 und 9) als *Input* eingeordnet werden. *Prozesse* stellen Aspekte der Qualität und Quantität des Lehrens und Lernens dar, wie zum Beispiel Schulversäumnisse der Jugendlichen (vgl. Kapitel 5). Als *Ergebnisse* werden neben den Kompetenzen in Mathematik, Lesen und Naturwissenschaften allgemeine sowie fachspezifische emotionale und motivationale Orientierungen, Einstellungen und Verhaltensweisen unterschieden. Mit dieser Modellierung können Forschungsfragen strukturiert werden, die Zusammenhänge zwischen entsprechenden Merkmalen untersuchen. Bei der Interpretation muss jedoch beachtet werden, dass mit dem in PISA verwendeten Querschnittsdesign zwar Zusammenhänge beschrieben, jedoch in der Regel keine Aussagen über Ursache-Wirkungs-Beziehungen getroffen werden können.

Die individuellen, schülerbezogenen Merkmale werden bei PISA mit Hilfe des Schülerfragebogens erfasst, der 2012 zum ersten Mal – ähnlich wie die Testhefte – in einem Rotationsdesign mit drei verschiedenen Fragebogenvarianten eingesetzt wurde. Alle Varianten enthielten zu Beginn einen gemeinsamen Teil mit Fragen zum sozialen und familiären Hintergrund. Anschließend beantworteten die Schülerinnen und Schüler unterschiedlich zusammengestellte Fragebogenteile, die auch emotionale wie motivationale Orientierungen, Einstellungen und Verhaltensweisen betrafen. Zu beachten ist daher, dass bei der Auswertung der Daten immer nur die Antworten von zwei Drittel der Schülerinnen und Schüler berücksichtigt werden konnten. Mit den verwendeten statistischen Verfahren kann aber dennoch eine robuste Schätzung der Schülermerkmale auf Populationsebene vorgenommen werden (Adams, Lietz & Berezner, 2013).

3.2 Beschreibung der erhobenen Schülermerkmale

Die mathematikbezogenen Schülermerkmale lassen sich bei PISA 2012 in drei Gruppen unterteilen: (a) emotionale und motivationale Orientierungen, (b) Selbstbilder sowie (c) mathematikbezogene Einstellungen und Verhaltensweisen. Dabei ist zu berücksichtigen, dass in PISA relativ stabile, über die Zeit entstandene und bereits in der Person verankerte Orientierungen und Erlebens- sowie Verhaltenstendenzen erhoben werden. Die Modellierung der Schülermerkmale orientiert sich an theoretischen Vorstellungen, die in der einschlägigen Forschung als weitgehend akzeptiert gelten können.

3.2.1 Emotionale und motivationale Orientierungen

Emotionen beeinflussen das Lern- und Leistungsverhalten von Schülerinnen und Schülern und sind ein wichtiger Bestandteil ihres Wohlbefindens in der Schule (Pekrun & Linnenbrink-Garcia, im Druck). Wie Jugendliche emotional mit mathematikspezifischen Anforderungssituationen umgehen, wird durch Schule und Unterricht beeinflusst. Positive Emotionen wie Freude erhöhen die Lernbereitschaft für Mathematik und führen dazu, dass sich Schülerinnen und Schüler für Schulen oder Schulzweige entscheiden, die einen mathematisch-naturwissenschaftlichen Schwerpunkt aufweisen (Pekrun, von Hofe, Blum, Frenzel, Götz & Wartha, 2007). Im Gegensatz dazu können negative Emotionen wie Angst das Lernen behindern und bewirken, dass Ausbildungen und Berufe, die ein gewisses Verständnis an Mathematik erfordern, eher vermieden werden (Ashcraft & Moore, 2009). Bei PISA 2012 wurden wie bereits bei PISA 2003 *Freude und Interesse* an Mathematik mit vier Items und mathematikbezogene *Ängstlichkeit* mit fünf Items erhoben (vgl. Tabelle 3.1). Freude und Interesse wurden dabei in einer Skala zusammengefasst, da sie in enger Beziehung zueinander stehen und einander ergänzen (Izard, 2007; Krapp & Prenzel, 2011). Die Schülerinnen und Schüler konnten zwischen den Antwortkategorien „stimme völlig zu“, „stimme eher zu“, „stimme eher nicht zu“ und „stimme überhaupt nicht zu“ auswählen.

Mit den emotionalen Orientierungen eng verknüpft sind motivationale Schülermerkmale. Bei PISA 2012 steht die *instrumentelle Motivation* im Mittelpunkt der Betrachtung. Sie beschreibt, wie Jugendliche die Bedeutung und Relevanz von Mathematik für ihre Zukunft einschätzen beziehungsweise welchen Wert sie der Mathematik für ihre persönliche Zukunft beimessen. Eine hohe Relevanzeinschätzung muss aber nicht unbedingt mit dem Erleben von Freude und Interesse verbunden sein. Vielmehr geht es hier darum, ob die Jugendlichen glauben, dass Mathematik für ihr zukünftiges Ausbildungs- und Berufsleben wichtig ist. Eine hoch ausgeprägte instrumentelle Motivation kann demnach unter anderem die zukünftige Ausbildungs- und Berufswahl beeinflussen (Wang, 2012). Umgekehrt kann sie aber auch durch Berufsvorstellungen und Berufswünsche befördert werden, wenn man zum Beispiel weiß, dass auch sozialwissenschaftliche Studiengänge mathematische Kompetenz voraussetzen. Die instrumentelle Motivation wurde bei PISA 2012 mit vier Items erfasst, die ebenfalls auf einer vierstufigen Zustimmungsskala einzuschätzen waren (vgl. Tabelle 3.1).

Die Wahrnehmung und Einschätzung der Mathematik hängt nicht nur davon ab, welchen Wert Jugendliche selbst der Mathematik beimessen, sondern sie wird auch dadurch beeinflusst, wie ihrer Ansicht nach das persönliche Umfeld, nämlich ihre Eltern sowie Freundinnen und Freunde, mathematikbezogenes Verhalten bewerten oder selbst zeigen (Ajzen & Madden, 1986; Lipnevich, MacCann, Krumm, Burrus & Roberts, 2011). Diese *sozialen Normen* wurden mit sechs Items erfasst (vgl. Tabelle 3.1).

Tabelle 3.1: Skalen der Schülermerkmale bei PISA 2012 mit Beispielitems

Schülermerkmal	Anzahl der Items	Reliabilität*	Beispielitem
Freude und Interesse	4	.89	Mich interessiert das, was ich in Mathematik lerne.
Ängstlichkeit	5	.87	Beim Lösen von Aufgaben in Mathematik werde ich sehr nervös.
Instrumentelle Motivation	4	.84	Mathematik zu lernen lohnt sich, weil es meine Berufs- und Karriereaussichten verbessert.
Soziale Normen	6	.81	Meine Eltern sind der Meinung, dass es für mich wichtig ist, Mathematik zu lernen.
Selbstkonzept	5	.87	Im Mathematikunterricht verstehe ich sogar die schwierigsten Aufgaben.
Selbstwirksamkeitserwartung	8	.81	Ausrechnen, wie viele Quadratmeter Fliesen du bräuchtest, um einen Fußboden damit auszulegen.
Gewissenhafte Arbeitshaltung	9	.87	Ich passe im Mathematikunterricht gut auf.
Intentionen	5	.77	Ich habe beschlossen, einen Beruf zu wählen, der viel mit Mathematik zu tun hat. Ich habe beschlossen, einen Beruf zu wählen, der viel mit Naturwissenschaften zu tun hat.
Beschäftigung mit Mathematik	8	.70	Ich bin Mitglied in einer Mathematik-AG.

Anmerkung: *Die Reliabilität entspricht Cronbachs Alpha. Dieser Wert wurde auf der Basis der deutschen Stichprobe ermittelt.

3.2.2 Selbstbilder

Ein weiteres wichtiges Ziel des Schulbesuchs ist die Herausbildung eines realistisch-positiven *Selbstkonzepts*. Damit sind Vorstellungen gemeint, die ein Jugendlicher von sich selbst hat (Marsh, Xu & Martin, 2012). Im Kontext von Leistungsvergleichen stehen vor allem subjektive Vorstellungen wie das Zutrauen in die eigenen Fähigkeiten im Blickpunkt. Von der Schule stark geprägt wird etwa das mathematische Fähigkeitsselbstkonzept, das von einem sprachlichen Fähigkeitsselbstkonzept abgegrenzt werden kann. Man kann sich demnach für mathematisch begabt halten, aber eine niedrige Selbsteinschätzung für das Lernen von Sprachen haben oder umgekehrt. Selbstkonzepte entstehen in der Schule vor allem durch soziale Vergleiche mit anderen Schülerinnen und Schülern sowie durch Vergleiche der eigenen Leistung in unterschiedlichen Fächern. Das Fähigkeitsselbstkonzept hängt deshalb mit der Leistung, dem Herausbilden von Interessen, Kurswahlentscheidungen sowie dem Wohlbefinden in der Schule zusammen (Hascher, 2004; Köller, Trautwein, Lüdtke & Baumert, 2006; Marsh & Martin, 2011; Möller, Retelsdorf, Köller & Marsh, 2011). Bei PISA 2012 wurde das mathematische Selbstkonzept (also wie gut jemand glaubt, Mathematik bewältigen zu können) mit fünf Items erhoben, bei denen aus den vier Antwortkategorien „stimme völlig zu“, „stimme eher zu“, „stimme eher nicht zu“ und „stimme überhaupt nicht zu“ jeweils eine ausgewählt werden sollte (vgl. Tabelle 3.1).

In engem Zusammenhang mit dem mathematischen Selbstkonzept stehen mathematikbezogene *Selbstwirksamkeitserwartungen* (Ferla, Valcke & Cai, 2009; Lee, 2009). Im Unterschied zum Selbstkonzept, das sich eher auf verallgemeinerte Überzeugungen der eigenen Leistungsfähigkeit bezieht, beschreiben Selbstwirksamkeitserwartungen Überzeugungen, konkrete Handlungen und Problemstellungen erfolgreich bewältigen zu können (Bandura, 1977). Selbstwirksamkeitserwartungen hängen vor allem mit Motivations- und Selbstregulationsstrategien wie erhöhter Anstrengungsbereitschaft und Ausdauer zusammen (Jerusalem & Hopf, 2002; Klassen & Usher, 2010). Wenn Jugendliche der Auffassung sind, dass sie nicht über die Fähigkeiten und Fertigkeiten verfügen, die für das Lösen bestimmter Aufgaben erforderlich sind, dann strengen sie sich bei der Lösung der Aufgabe weniger an. Dadurch sinkt jedoch im Sinne einer sich selbst erfüllenden Prophezeiung die Wahrscheinlichkeit, die Aufgabe tatsächlich richtig zu lösen. Bei PISA 2012 wurde die mathematische Selbstwirksamkeitserwartung mit acht Items erhoben. Die Items repräsentieren unterschiedliche Arten von Mathematikaufgaben. Die Schülerinnen und Schüler gaben auf einer vierstufigen Antwortskala an, wie sicher sie sind, dass sie die jeweilige Mathematikaufgabe lösen können (vgl. Tabelle 3.1).

3.2.3 Mathematikbezogene Einstellungen und Verhaltensweisen

Der Mathematikunterricht bietet Gelegenheit, *gewissenhafte Arbeitshaltungen* zu entwickeln und von diesen zu profitieren. Solche Arbeitshaltungen bestehen etwa darin, dass sich Schülerinnen und Schüler konsequent auf Mathematikprüfungen vorbereiten, im Unterricht mitarbeiten und ihre Anstrengung steigern, um gute Leistungen zu erzielen. Eine gewissenhafte Arbeitshaltung hängt mit einer besseren Leistung in der Schule zusammen (vgl. de Fruyt, van Leeuwen, de Bolle & de Clercq, 2008). Die mathematikbezogenen Arbeitshaltungen wurden mit neun Items erfasst. Die Schülerinnen und Schüler schätzten auf einer vierstufigen Zustimmungsskala ein, ob sie sich auf den Mathematikunterricht vorbereiten, sich anstrengen und im Unterricht selbst gut aufpassen.

Darüber hinaus wurde bei PISA 2012 gefragt, ob die Jugendlichen bereit sind, überdurchschnittlich viel in den Mathematikunterricht zu investieren, beziehungsweise ob sie vorhaben, sich auch nach der Schule weiterhin mit Mathematik auseinanderzusetzen. Diese Absichten (auch *Intentionen* genannt) eignen sich, um tatsächliches Verhalten vorherzusagen (Ajzen, 2011). Die Items zur Erfassung der Absichten wurden als Forced-Choice-Items formuliert. Dabei wählten die Fünfzehnjährigen aus jeweils zwei Aussagen diejenige aus, die sie am besten beschreibt. Eine Aussage bezog sich dabei immer auf Mathematik und die andere auf Deutsch oder Naturwissenschaften (vgl. Tabelle 3.1). Die Fragen betrafen das Lernverhalten, die Wahl zusätzlicher Kurse sowie die Karriereplanung.

Im Hinblick auf eine Förderung des MINT¹-Nachwuchses ist es ebenfalls wichtig, ob sich Jugendliche verschiedentlich auch außerhalb des Unterrichts mit Mathematik auseinandersetzen. Deshalb wurden die Schülerinnen und Schüler nach ihrer *Beschäftigung mit Mathematik* über den normalen Mathematikunterricht hinaus befragt. Dazu gehört es, Freundinnen und Freunden bei Mathematik zu helfen, an Mathematikwettbewerben oder Mathematik-AGs teilzunehmen, Schach zu spielen oder Computer zu programmieren. Es wurden acht Items vorgegeben, bei denen die Jugendlichen einschätzen sollten, wie häufig sie an den mathematikbezogenen Aktivitäten innerhalb und außerhalb der Schule beteiligt sind (vgl. Tabelle 3.1).

3.3 Ergebnisse

Im Folgenden werden für jedes Schülermerkmal zunächst die Befunde für Deutschland in den internationalen Vergleich eingeordnet sowie Unterschiede zwischen Mädchen und Jungen betrachtet. Anschließend zeigen vertiefende Analysen für Deutschland, wie sich einige Schülermerkmale seit PISA 2003 verändert haben. Zum Abschluss werden Zusammenhangsanalysen zwischen den Schülermerkmalen dargestellt.

Methodische Vorbemerkung

Die hier berichteten Ergebnisse beruhen auf Selbsteinschätzungen der fünfzehnjährigen Jugendlichen, die mit Hilfe des Schülerfragebogens erhoben wurden. Für jedes Merkmal wurden unter Anwendung der Item-Response-Theorie Skalenwerte geschätzt (vgl. Kapitel 10), die anschließend so standardisiert wurden, dass der Durchschnitt der OECD-Staaten null und die Standardabweichung eins beträgt. Positive Skalenwerte geben an, dass das Merkmal im Vergleich zum OECD-Durchschnitt überdurchschnittlich ausgeprägt ist. Negative Skalenwerte zeigen an, dass das entsprechende Merkmal im OECD-Vergleich unterdurchschnittlich ausgeprägt ist. Damit sind Vergleiche zwischen einzelnen Ländern möglich, es können jedoch keine Aussagen über die absolute Höhe der Merkmalsausprägungen getroffen werden.

Die Reliabilitäten der Skalen werden durch den Koeffizienten Cronbachs Alpha ausgedrückt, der auf der Basis der deutschen Stichprobe berechnet wurde. Die Werte liegen für die Schülermerkmale zwischen .70 und .89, was einer guten bis sehr guten Messgenauigkeit entspricht (vgl. Tabelle 3.1).

Zusätzlich zu den Skalenwerten berichten wir für jedes Merkmal charakteristische Beispielitems und die zugehörigen prozentualen Zustimmungen zu den einzelnen Antwortkategorien. Auf diese Weise soll die jeweilige Skala illustriert und das Antwortverhalten genauer betrachtet werden. Darüber hinaus erleichtert die Darstellung der Beispielitems die inhaltliche Deutung der Skalenwerte, da sie mit dem standardisierten

1 Mathematik, Informatik, Naturwissenschaften, Technik.

Mittelwert der Skala in Verbindung gebracht werden können. Die Items wurden so ausgewählt, dass sie die Verteilung der relativen Häufigkeitsangaben für die jeweilige Skala möglichst gut repräsentieren (vgl. Abbildungen 3.1 bis 3.3).

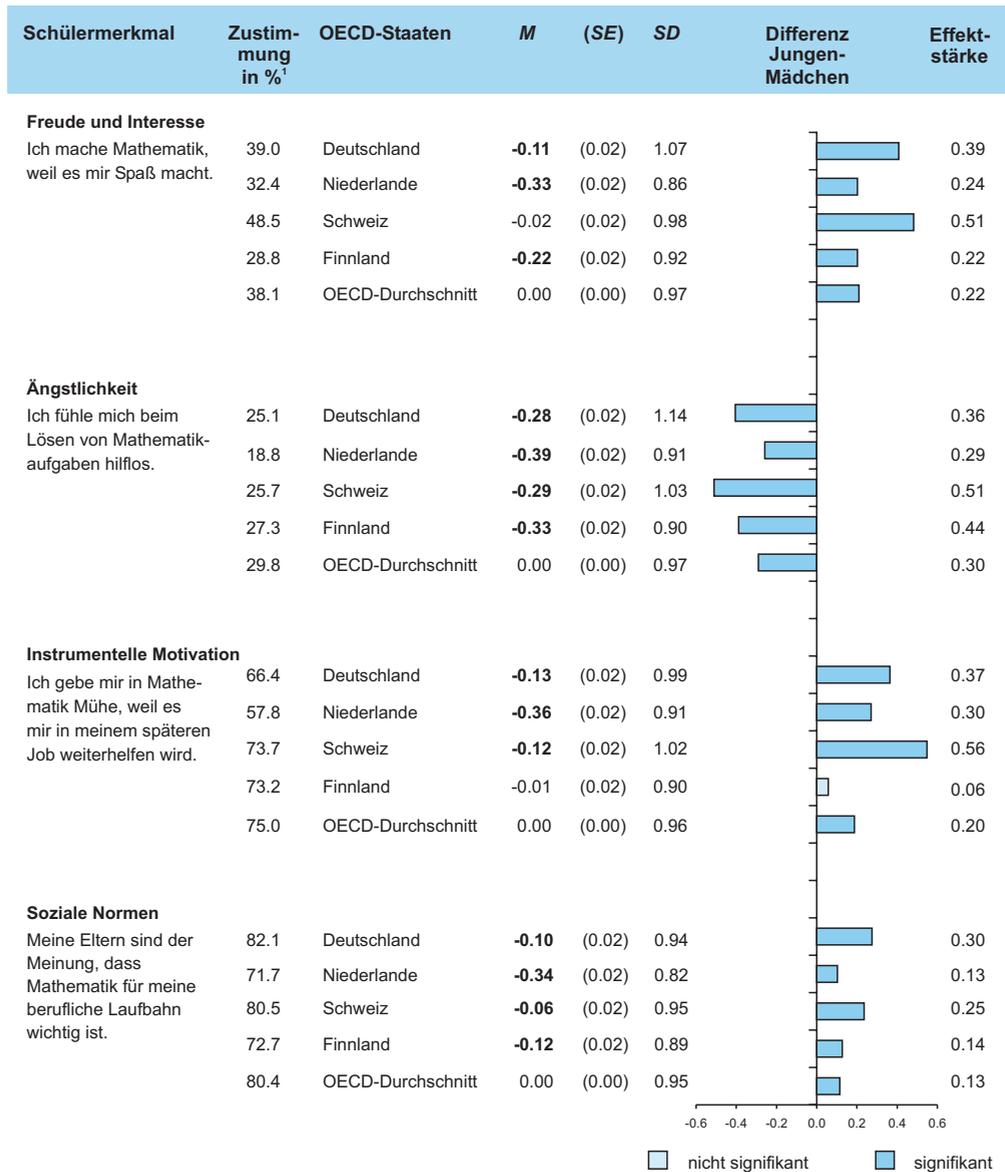
Bei der Einordnung der Ergebnisse für Deutschland in den internationalen Vergleich muss beachtet werden, dass Schülerinnen und Schüler ihre Selbsteinschätzungen nicht losgelöst vom jeweiligen kulturellen Hintergrund treffen. Das kann dazu führen, dass sie Aussagen grundsätzlich eher zustimmen als sie abzulehnen oder extreme Antwortkategorien generell vermeiden (Johnson, Shavitt & Holbrook, 2011). Sie antworten möglicherweise auch so, wie sie glauben, dass es von ihnen erwartet wird (soziale Erwünschtheit). Demnach können kulturspezifische Werte und Normen zu kulturbedingten Unterschieden im Antwortverhalten führen, die eine internationale Vergleichbarkeit beeinträchtigen (van de Gaer, Grisay, Schulz & Gebhardt, 2012). Darüber hinaus ist die Selbsteinschätzung davon abhängig, welche Bezugsgruppe für den Vergleich ausgewählt wird (Heine, Lehman, Peng & Greenholtz, 2002).

Aus diesen Gründen betrachten wir zum einen die Antworten der Schülerinnen und Schüler im Vergleich zum OECD-Durchschnitt und wählen zusätzlich drei Staaten aus, die Kulturräumen mit gewissen Ähnlichkeiten zu Deutschland angehören: die Schweiz, die Niederlande und Finnland. Diese Länder sind auch deshalb von Interesse, weil sie im internationalen Vergleich überdurchschnittliche mathematische Kompetenzwerte erzielten (vgl. Kapitel 2). Zudem gelingt es in diesen Ländern, kompetenzschwache und kompetenzstarke Jugendliche gleichermaßen erfolgreich zu fördern, was sich in einem geringen Anteil von Jugendlichen auf Kompetenzstufe I oder darunter sowie in einem hohen Anteil Jugendlicher auf den Kompetenzstufen V und VI widerspiegelt.

3.3.1 Emotionale und motivationale Orientierungen im internationalen Vergleich

Betrachtet man die emotionalen Orientierungen im OECD-Vergleich (Abbildung 3.1), so fällt auf, dass die Kennwerte für *Freude und Interesse* an Mathematik in Deutschland etwas niedriger ausgeprägt sind als im OECD-Durchschnitt. Dies sagt jedoch noch nichts über die absolute Höhe der Ausprägung der Freude an Mathematik aus. Berücksichtigt man aber in diesem Zusammenhang, dass nur 39,0 Prozent der Schülerinnen und Schüler in Deutschland angeben, dass sie sich mit Mathematik beschäftigen, weil es ihnen Spaß macht, wird deutlich, dass rund 60 Prozent der Fünfzehnjährigen dem „eher nicht“ oder „überhaupt nicht“ zustimmen.

In den Niederlanden und in Finnland berichten die Jugendlichen eine noch geringere Freude an Mathematik als die Fünfzehnjährigen in Deutschland. Nur die Schülerinnen und Schüler in der Schweiz zeigen mehr Freude und Interesse. Dort gibt immerhin knapp die Hälfte der Jugendlichen an, dass sie sich mit Mathematik beschäftigen, weil es ihnen Spaß macht. In der Schweiz sind jedoch, ebenso wie in Deutschland, deutliche Geschlechterunterschiede zu finden: Jungen berichten über mehr Freude und Interesse



Anmerkung: Statistisch signifikante Unterschiede vom OECD-Durchschnitt sind fett hervorgehoben.

¹ Die Antwortkategorien „stimme eher zu“ und „stimme völlig zu“ wurden zusammengefasst.

Abbildung 3.1: Emotionale und motivationale Orientierungen im internationalen Vergleich

an Mathematik als Mädchen. Auch in Finnland und den Niederlanden zeigen sich diese Geschlechterunterschiede, jedoch sind sie die Effektstärken (vgl. Kapitel 1) dort schwächer ausgeprägt ($d_{\text{Finnland}} = 0.22$ und $d_{\text{Niederlande}} = 0.24$).

Die mathematikbezogene *Ängstlichkeit* ist in Deutschland wie in den drei Vergleichsstaaten (im Verhältnis zum OECD-Durchschnitt) geringer ausgeprägt. Am wenigsten Ängstlichkeit berichten die Schülerinnen und Schüler in den Niederlanden. Eine genauere Betrachtung des Beispielitems zum Teilaspekt der Hilflosigkeit zeigt, dass ein Viertel der Jugendlichen in Deutschland der Aussage, sich hilflos beim Lösen von Mathematikaufgaben zu fühlen, „eher“ oder „völlig“ zustimmt. Wenn gefragt wird, ob die Fünfzehnjährigen sich Sorgen machen, dass es für sie im Mathematikunterricht schwierig sein wird oder dass sie in Mathematik eine schlechte Note bekommen, stimmt ein relativ großer Anteil – nämlich knapp die Hälfte – „eher zu“ oder „völlig zu“. In allen Staaten fühlen sich zudem die Mädchen ängstlicher als die Jungen, wobei dieser Unterschied in der Schweiz am stärksten ausgeprägt ist ($d = 0.51$).

Die *instrumentelle Motivation* ist in Deutschland etwas niedriger ausgeprägt als im OECD-Durchschnitt. Der Frage, inwieweit sie sich Mühe in Mathematik geben, weil es ihnen für den späteren Beruf hilft, stimmen in Deutschland zwei Drittel der Jugendlichen „eher zu“ oder „völlig zu“. Demnach ist dem Großteil der Schülerinnen und Schüler sehr wohl bewusst, dass Mathematik eine wichtige Grundlage für ihr weiteres Ausbildungs- und Berufsleben ist.

In der Schweiz und in Finnland stimmen dieser Aussage sogar knapp drei Viertel der Jugendlichen zu. Lediglich in den Niederlanden ist die instrumentelle Motivation signifikant geringer ausgeprägt. Wie auch bei der Freude an Mathematik berichten Jungen in Deutschland, den Niederlanden und der Schweiz über eine höhere instrumentelle Motivation als Mädchen. In der Schweiz ist der Unterschied erneut am größten ($d = 0.56$). In Finnland ist hier keine signifikante Geschlechterdifferenz zu beobachten.

Neben dem Wert, den Jugendliche selbst der Mathematik beimessen, wird mit den Fragen zu *sozialen Normen* erfasst, welche Bedeutung Eltern und Freunde der Mathematik zuschreiben. Deutschland zeigt hier insgesamt einen etwas niedrigeren Wert als der OECD-Durchschnitt. Die Einzelitems weisen dabei auf einen interessanten Unterschied hin: Die überwiegende Mehrheit der Jugendlichen in Deutschland (82.1 Prozent) gibt an, dass ihre Eltern finden, Mathematik sei wichtig für ihre berufliche Laufbahn. Fragt man sie aber danach, welchen Wert ihre Freundinnen und Freunde Mathematik beimessen, indem diese zum Beispiel hart für Mathematik arbeiten, stimmen dem nur 40.4 Prozent „eher zu“ oder „völlig zu“. Eltern scheinen demnach (basierend auf der Wahrnehmung der Jugendlichen) die Bedeutung von Mathematik höher einzuschätzen als Gleichaltrige.

Ein ähnliches Bild findet man in der Schweiz und in Finnland. In den Niederlanden scheinen Eltern und Freunde Mathematik noch einen signifikant geringeren Wert beizumessen als in Deutschland, Finnland und der Schweiz. Die Geschlechterunterschiede weisen in den berichteten Ländern eher schwache Effektstärken auf. In Deutschland unterscheiden sich Mädchen und Jungen am stärksten ($d = 0.30$). Mädchen schätzen

demnach die Bedeutung, die Mathematik für ihre Eltern und Freunde hat, noch geringer ein als Jungen.

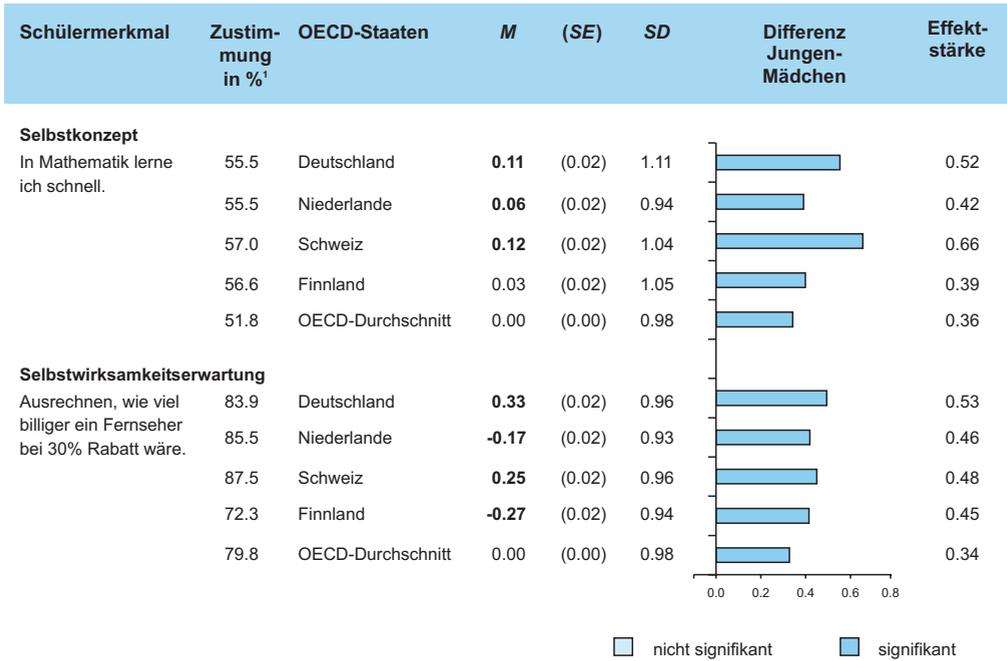
3.3.2 Selbstbilder im internationalen Vergleich

Bei der Betrachtung der Selbstbilder (Abbildung 3.2) fällt auf, dass Schülerinnen und Schüler in Deutschland ein höheres *Selbstkonzept* ihrer mathematischen Fähigkeiten berichten als im OECD-Durchschnitt. 55.5 Prozent der Jugendlichen stimmen dem Item ‚In Mathematik lerne ich schnell‘ „eher zu“ oder „völlig zu“.

Dieser Wert ist in den Vergleichsstaaten ähnlich hoch ausgeprägt. Das bedeutet zugleich, dass knapp die Hälfte der Jugendlichen ein geringes Zutrauen in ihre mathematikbezogenen Fähigkeiten hat. Etwa ein Drittel stimmt darüber hinaus der Aussage ‚Ich bin einfach nicht gut in Mathematik‘ „eher zu“ (20.4 Prozent) oder „völlig zu“ (14.7 Prozent). Vor diesem Hintergrund besteht die Gefahr, dass eine so pessimistische Einschätzung der eigenen mathematischen Fähigkeiten die Schülerinnen und Schüler davon abhält, durch verstärkte Anstrengung ihr mathematisches Potential zu entwickeln. Aus diesem Grund ist es wichtig, durch adäquate Aufgabenschwierigkeiten und informierende Rückmeldungen (unter einer individuellen Bezugsnorm) darauf hinzuwirken, dass Schülerinnen und Schüler Vertrauen in ihre mathematischen Fähigkeiten entwickeln können (vgl. Rheinberg, 2008). Dies gilt wiederum im Besonderen für die Mädchen, denn auch hier zeigen sich für Deutschland (und die Vergleichsstaaten) ausgeprägte Geschlechterdifferenzen zugunsten der Jungen. Sie berichten durchweg ein höheres Selbstkonzept als Mädchen. Für Deutschland kann dies als mittlerer Effekt eingeschätzt werden ($d = 0.52$). Im Vergleich zu den emotionalen und motivationalen Orientierungen ist hier der größte und bedeutsamste Unterschied zwischen Mädchen und Jungen zu finden.

Die *Selbstwirksamkeitserwartungen* sind in Deutschland deutlich höher ausgeprägt als in den Vergleichsstaaten. Die Jugendlichen haben demnach signifikant höhere Überzeugungen bezüglich der Umsetzung konkreter mathematischer Handlungen und der Bewältigung konkreter Aufgabenstellungen. Die Items zur Selbstwirksamkeitserwartung erfragen sowohl die Überzeugung, innermathematische Aufgaben als auch Aufgaben mit einfachem Anwendungsbezug lösen zu können (vgl. Kapitel 4).

In Deutschland geben 89.4 Prozent der Jugendlichen bei einer innermathematischen Beispielaufgabe (‚Eine Gleichung wie $3x + 5 = 17$ lösen‘) an, dass sie „eher sicher“ oder „sehr sicher“ sind, diese lösen zu können. In der Schweiz sind es 87.2 Prozent, in Finnland 83.7 Prozent und in den Niederlanden 77.4 Prozent. Bei einer Beispielaufgabe mit einfachem Anwendungsbezug sollten die Schülerinnen und Schüler einschätzen, wie sicher sie die Preisersparnis beim Kauf eines Fernsehers bei 30 Prozent Rabatt berechnen können. 83.9 Prozent trauen sich in Deutschland zu, diese Aufgabe „eher sicher“ oder „sehr sicher“ zu lösen, ähnlich hohe Werte sind in den Niederlanden und der Schweiz zu finden, nur in Finnland ist die Überzeugung, diese Aufgabe zu lösen, mit



Anmerkung: Statistisch signifikante Unterschiede vom OECD-Durchschnitt sind fett hervorgehoben.
¹Die Antwortkategorien „stimme eher zu“ und „stimme völlig zu“ bzw. „eher sicher“ und „sehr sicher“ wurden zusammengefasst.

Abbildung 3.2: Mathematikbezogene Selbstbilder im internationalen Vergleich

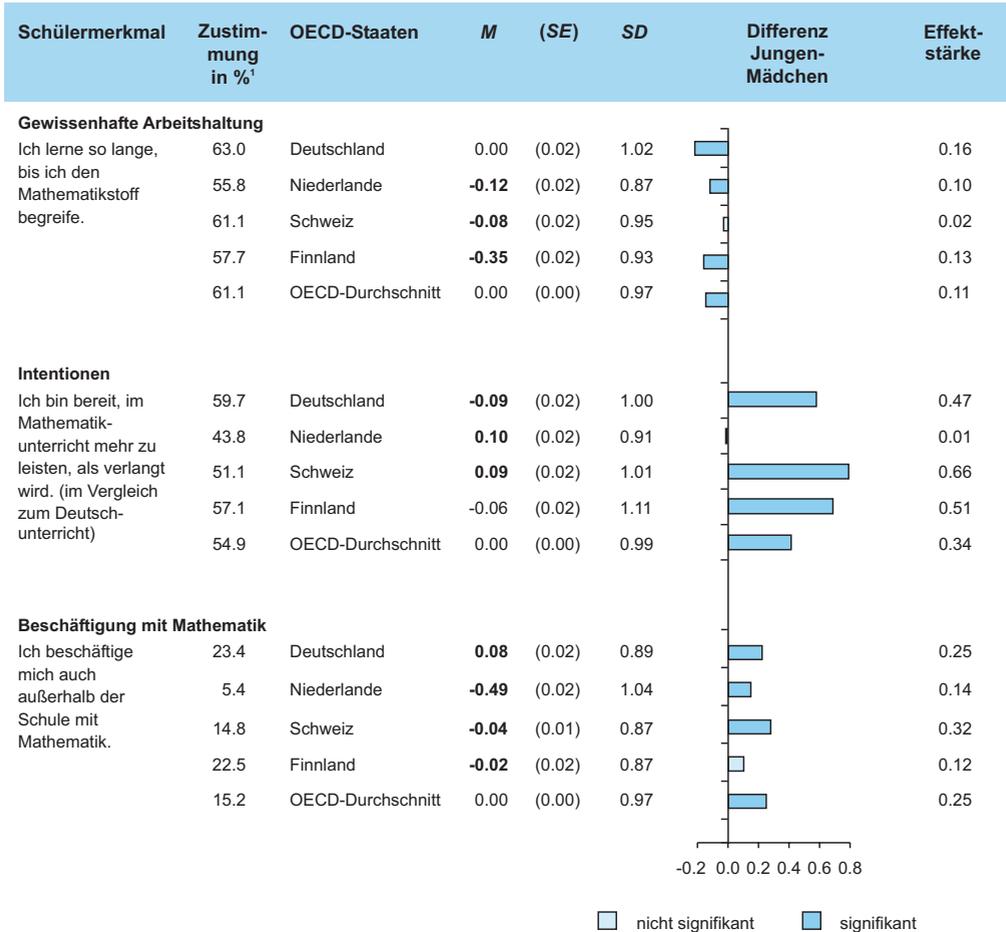
72.3 Prozent etwas geringer ausgeprägt. Da die Selbstwirksamkeitserwartungen darüber erfasst werden, wie sicher sich Schülerinnen und Schüler fühlen, konkrete Mathematikaufgaben lösen zu können, muss beachtet werden, dass die Antworten auch davon abhängen, wie häufig die jeweiligen Aufgabentypen im Mathematikunterricht vorkommen (vgl. Kapitel 4). Für Deutschland zeigt sich, dass Jugendliche, die angeben, dass die jeweilige Aufgabe „häufig“ oder „manchmal“ in ihrem Unterricht vorkam, sich auch sicherer fühlen, diese Aufgabe lösen zu können (OECD, 2013b). Insgesamt zeigt sich, dass Jugendliche in Deutschland sich relativ sicher fühlen, einfache Anwendungsaufgaben lösen zu können, auch wenn im Mathematikunterricht eher innermathematische Aufgaben (z. B. Gleichungen) überwiegen. Jungen berichten darüber hinaus eine höhere Selbstwirksamkeitserwartung als Mädchen ($d = 0.53$). Das gleiche Muster ist auch in den berichteten Vergleichsstaaten mit Effektstärken im mittleren Bereich zwischen $d = 0.45$ und $d = 0.53$ zu beobachten.

3.3.3 Mathematikbezogene Einstellungen und Verhaltensweisen im internationalen Vergleich

Im Rahmen der mathematikbezogenen Einstellungen und Verhaltensweisen zeigt sich bei der Skala *gewissenhafte Arbeitshaltung*, dass Jugendliche in Deutschland im Vergleich zu den hier berücksichtigten Staaten höhere Werte angeben (Abbildung 3.3). Eine gewissenhafte Arbeitshaltung zeichnet sich zum Beispiel dadurch aus, dass man sich so lange mit dem Mathematikstoff auseinandersetzt, bis man ihn verstanden hat. Dem stimmen in Deutschland 63.0 Prozent der Schülerinnen und Schüler „eher zu“ oder „völlig zu“. Der Aussage, sich immer auf Mathematikprüfungen vorzubereiten, stimmen sogar 70.7 Prozent „eher zu“ oder „völlig zu“. Hier zeigen sich in den vier Staaten keine oder nur schwache Geschlechterunterschiede, die diesmal jedoch zugunsten der Mädchen ausfallen. Demnach berichten Mädchen eine etwas gewissenhaftere Arbeitshaltung als Jungen, wenngleich die Effektstärken der Geschlechterdifferenz schwach ausgeprägt sind.

Bei der Frage nach den *Intentionen*, sich in der Schulzeit aber auch darüber hinaus mit Mathematik anstelle von Deutsch oder Naturwissenschaften auseinanderzusetzen, fällt auf, dass sich in Deutschland die Jugendlichen etwas weniger mit Mathematik befassen wollen als die Jugendlichen in der Schweiz und den Niederlanden. Betrachtet man genauer den Vergleich zwischen Mathematik und Deutsch sowie zwischen Mathematik und Naturwissenschaften, stimmen 59.7 Prozent der Jugendlichen in Deutschland „eher zu“ oder „völlig zu“ im Mathematikunterricht bereit zu sein, mehr zu leisten, als verlangt wird (im Vergleich zum Deutschunterricht). In den Niederlanden sind es hingegen nur 43.8 Prozent. Wenn die Wahl zwischen einem Beruf, der viel mit Mathematik zu tun hat, und einem, der viel mit Naturwissenschaften zu tun hat, getroffen werden müsste, entscheidet sich in Deutschland etwas mehr als die Hälfte der Jugendlichen für die Naturwissenschaften. Zu beachten ist hier, dass sich die Fünfzehnjährigen durch die angebotenen Antwortoptionen zwischen Mathematik und Naturwissenschaften beziehungsweise Deutsch entscheiden mussten. Es können daraus keine Aussagen über die absolute Höhe der Absicht getroffen werden, sich weiterhin mit Mathematik zu befassen. Außer in den Niederlanden zeigen sich insgesamt deutliche Geschlechterunterschiede, die darauf hinweisen, dass Jungen eher als Mädchen die Absicht haben, sich weiterhin mit Mathematik auseinanderzusetzen. Die Effektstärke für diesen Unterschied liegt in Deutschland bei $d = 0.47$.

Ob sich Jugendliche auch außerhalb des Unterrichts mit Mathematik beschäftigen, wurde anhand der Skala *Beschäftigung mit Mathematik* ermittelt. Im Vergleich zu den drei anderen Staaten befassen sich Fünfzehnjährige in Deutschland häufiger außerhalb des Unterrichts mit Mathematik. Hier fallen insbesondere die Niederlande auf, wo die Angaben der Schülerinnen und Schüler zur Beschäftigung mit Mathematik deutlich geringer ausfallen. Dort berichten nur 5.4 Prozent, dass sie sich außerhalb der Schule mit Mathematik beschäftigen, in den drei anderen Staaten sind es mindestens 14 Prozent. In Deutschland stimmen 23.4 Prozent der Jugendlichen dieser Aussage „eher zu“



Anmerkung: Statistisch signifikante Unterschiede vom OECD-Durchschnitt sind fett hervorgehoben.
¹ Bei der gewissenhaften Arbeitshaltung und der Beschäftigung mit Mathematik wurden die Antwortkategorien „stimme eher zu“ und „stimme völlig zu“ zusammengefasst.

Abbildung 3.3: Mathematikbezogene Einstellungen und Verhaltensweisen im internationalen Vergleich

oder „völlig zu“. Besonders intensive Beschäftigung mit Mathematik zum Beispiel im Rahmen einer Mathematik-AG oder durch die Teilnahme an Mathematikwettbewerben wird lediglich von 2.1 Prozent beziehungsweise 5.1 Prozent der Jugendlichen in Deutschland berichtet. Erneut beschäftigen sich eher Jungen mit Mathematik als Mädchen. Einzig in Finnland ist dieser Unterschied nicht statistisch bedeutsam. Die Effektstärken liegen zwischen $d = 0.12$ und $d = 0.32$ ($d_{\text{Deutschland}} = 0.25$).

Exkurs: Ausprägungen der Schülermerkmale am Gymnasium

Die relativ große Homogenität der Gymnasien in Deutschland, sowie die ausreichend große Stichprobe dieser Schulart bei PISA 2012, erlaubt eine gesonderte Betrachtung. Dabei zeigt sich, dass im Vergleich zum nationalen Durchschnitt Jugendliche am Gymnasium etwas mehr Freude und Interesse an Mathematik haben ($M = -0.05$), deutlich weniger ängstlich sind ($M = -0.42$) und eine höhere Selbstwirksamkeitserwartung berichten ($M = 0.69$). Das Selbstkonzept ist ebenfalls etwas höher ausgeprägt ($M = 0.16$), ebenso die gewissenhafte Arbeitshaltung in Bezug auf Mathematik ($M = 0.04$). Etwas schlechter ausgeprägt ist die instrumentelle Motivation ($M = -0.19$). Auch Eltern und Freunde messen Mathematik weniger Bedeutung bei ($M = -0.29$). Ebenso ist die Beschäftigung mit Mathematik ($M = 0.04$) etwas geringer ausgeprägt und die Fünfzehnjährigen berichten wesentlich geringere Intentionen, sich weiterhin – auch über die Schulzeit hinaus – mit Mathematik auseinanderzusetzen ($M = -0.25$).

Diese Befunde lassen das Mathematiklernen am Gymnasium nicht als emotional und motivational besonders belastend erscheinen, so wie es in Diskussionen – speziell auch um die verkürzte Gymnasialzeit – verschiedentlich kolportiert wird. Jedoch zeigt sich gerade in Bezug auf eine zukunftsorientierte, instrumentelle Motivation noch Verbesserungspotential. Da die Jugendlichen am Gymnasium die besten mathematischen Leistungsvoraussetzungen besitzen, sollte vor allem hier die Bereitschaft erhöht werden, sich verstärkt in und nach der Schulzeit mit Mathematik auseinanderzusetzen.

3.3.4 Veränderungen der Schülermerkmale seit PISA 2003

Da bei PISA 2012 zum zweiten Mal Mathematik als Hauptdomäne im Mittelpunkt der Betrachtungen steht, kann auch für einige Schülermerkmale, die bereits 2003 erfasst wurden, eine Veränderung berichtet werden. Dies ist für Freude und Interesse, mathematikbezogene Ängstlichkeit, instrumentelle Motivation, Selbstkonzept und Selbstwirksamkeitserwartungen möglich. Abbildung 3.4 zeigt die Entwicklung dieser Merkmale zwischen den Jahren 2003 und 2012.

Besonders hervorzuheben ist, dass sich die mathematikbezogene Selbstwirksamkeitserwartung der Jugendlichen über die Zeit hinweg signifikant verbessert hat. Dies trifft sowohl auf die Jungen (von $M = 0.28$ auf $M = 0.58$) als auch auf die Mädchen (von $M = -0.15$ auf $M = 0.08$) zu. Um zu überprüfen, ob diese positive Entwicklung möglicherweise auf die Überzeugungen, bestimmte Aufgabentypen lösen zu können, zurückzuführen ist, zeigt Tabelle 3.2 die Veränderungen der prozentualen Zustimmungen zu den einzelnen Items der Skala. Die Befunde zeigen, dass sich Jugendliche im Vergleich zu 2003 vor allem sicherer fühlen, Aufgaben zu lösen, die einen einfachen Anwendungsbezug aufweisen (Diagramme in Zeitungen verstehen; anhand des Zugfahrplans ausrechnen, wie lange die Fahrt von einem Ort zu einem anderen dauern würde; ausrechnen, wie viel billiger ein Fernseher bei 30% Rabatt wäre). Eine mögliche Erklärung ist,

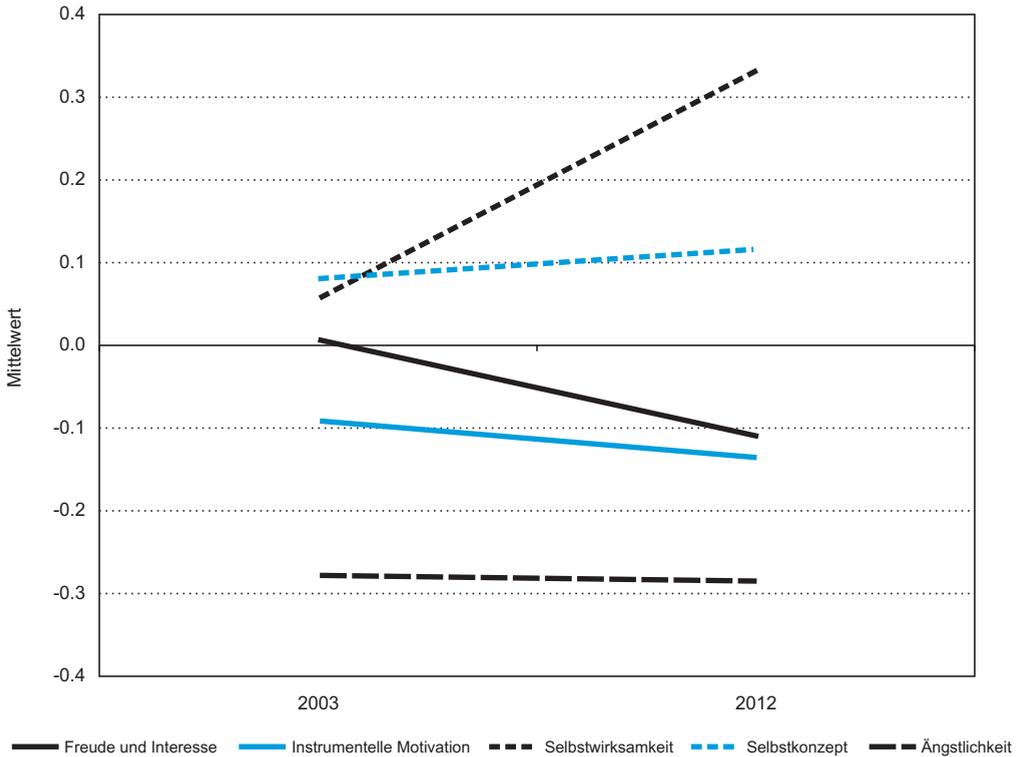


Abbildung 3.4: Veränderung der Schülermerkmale in Deutschland zwischen den Jahren 2003 und 2012

dass durch die Einführung der länderübergreifenden Bildungsstandards für den Mittleren Schulabschluss für das Fach Mathematik (KMK, 2003) anwendungsbezogene Aufgaben häufiger im Unterricht behandelt werden und sich die Schülerinnen und Schüler deshalb sicherer fühlen könnten, auch solche Aufgaben lösen zu können. Diese Hypothese kann mit den vorliegenden Daten jedoch nicht überprüft werden, da bei PISA 2003 die Auftretenshäufigkeit der Aufgabentypen nicht erhoben wurde. Keine signifikanten Veränderungen sind in Deutschland bei der Ängstlichkeit, der instrumentellen Motivation und dem Selbstkonzept zu beobachten.

Betrachtet man die Veränderungen in den ausgewählten Vergleichsstaaten, so sind dort zwar keine bedeutsamen Veränderungen in der Selbstwirksamkeit zu finden, allerdings zeigen sich überall Verbesserungen im Selbstkonzept (Differenz von 0.05 in der Schweiz, 0.08 in Finnland und 0.12 in den Niederlanden). Diese sind in allen drei Ländern auf einer Verbesserung des Selbstkonzepts bei den Mädchen zurückzuführen (in der Schweiz von $M = -0.29$ auf $M = -0.21$, in Finnland von $M = -0.28$ auf $M = -0.17$; in den Niederlanden von $M = -0.33$ auf $M = -0.14$). Auch in Deutschland hat sich das Selbstkonzept der Mädchen etwas verbessert von $M = -0.19$ auf $M = -0.16$, diese Verän-

Tabelle 3.2: Veränderung der mathematikbezogenen Selbstwirksamkeitserwartungen

	PISA 2003		PISA 2012		Veränderung 2003–2012	
	Zustimmung in % ¹	(SE)	Zustimmung in % ¹	(SE)	Zustimmung in % ¹	(SE)
Anhand des Zugfahrplans ausrechnen, wie lange die Fahrt von einem Ort zu einem anderen dauern würde.	83.3	(0.7)	92.1	(0.5)	8.8	(0.8)
Ausrechnen, wie viel billiger ein Fernseher bei 30% Rabatt wäre.	77.4	(0.7)	83.9	(0.7)	6.5	(1.0)
Ausrechnen, wie viele Quadratmeter Fliesen du bräuchtest, um einen Fußboden damit auszulegen.	75.0	(0.8)	79.2	(0.9)	4.3	(1.2)
Diagramme in Zeitungen verstehen.	79.0	(0.9)	89.0	(0.7)	10.0	(1.1)
Eine Gleichung wie $3x + 5 = 17$ lösen.	86.2	(0.7)	89.4	(0.6)	3.2	(1.0)
Auf einer Karte mit einem Maßstab von 1:10.000 die tatsächliche Entfernung zwischen zwei Orten bestimmen.	54.5	(0.8)	59.6	(1.0)	5.1	(1.3)
Eine Gleichung wie $2(x+3)=(x+3)(x-3)$ lösen.	73.0	(0.9)	73.4	(1.0)	0.5	(1.3)
Den Benzinverbrauch eines Autos berechnen.	58.5	(0.8)	64.4	(1.1)	5.9	(1.3)

*Statistisch signifikante Unterschiede zwischen PISA 2003 und PISA 2012 sind fett hervorgehoben.

¹ Die Antwortkategorien „eher sicher“ und „sehr sicher“ wurden zusammengefasst.

derung kann jedoch nicht als statistisch bedeutsam abgesichert werden. Hier zeichnen sich demnach in Deutschland Möglichkeiten einer Verbesserung ab.

Beachtung verlangt auch die signifikante Abnahme der Kennwerte für Freude und Interesse an Mathematik in Deutschland. Diese bedeutsame Abnahme zeigt sich sowohl für die Mädchen (von $M = -0.19$ auf $M = -0.32$) als auch für die Jungen (von $M = 0.21$ auf $M = 0.09$) mit einer Differenz von -0.13 beziehungsweise -0.12 . Eine ähnliche Verschlechterung zeigt sich in den Niederlanden und der Schweiz, wobei dies jedoch auf eine Abnahme von Freude und Interesse bei den Jungen zurückzuführen ist (Differenz von -0.14 in der Schweiz und -0.10 in den Niederlanden). Lediglich in Finnland wird eine leichte Verbesserung berichtet, die durch eine Zunahme der Freude an Mathematik bei den Mädchen begründet werden kann (von $M = -0.42$ auf $M = -0.33$). Hier besteht also in Deutschland ebenfalls Handlungsbedarf, denn Freude und Interesse sind wichtige Indikatoren für eine intrinsische Motivation (Krapp & Prenzel, 2011). Angesetzt werden muss jedoch nicht nur bei den Mädchen, wie es die erreichten Kompetenzwerte vermuten lassen würden (vgl. Kapitel 2), sondern auch bei den Jungen.

3.3.5 Zusammenhänge der Schülermerkmale untereinander

Im Folgenden berichten wir bivariate Korrelationen, um Zusammenhänge zwischen den Schülermerkmalen zu untersuchen. Bei der Betrachtung der Ergebnisse gilt es zu berücksichtigen, dass hier Zusammenhänge angezeigt werden und diese nicht als Ursache-Wirkungs-Beziehung interpretiert werden dürfen. Welches Merkmal zu einer Veränderung in einem anderen Merkmal führt, kann nur mit weiterführenden Studiendesigns wie zum Beispiel durch die Auswertung von Längsschnittdaten oder experimenteller Studien untersucht werden.

Emotionale und motivationale Orientierungen sowie Selbstbilder, Einstellungen und Verhaltensweisen sind nicht unabhängig voneinander, sondern innerhalb einer Person eng miteinander verzahnt und stehen in einem komplexen Wechselspiel zueinander (z. B. Ahmed, Minneart, Kuyper & van der Werf, 2012; Lee, 2009). Die beiden Emotionen Freude und Angst korrelieren in Deutschland mit $r = -.43$ ($p < .001$). Umso stärker demnach die Freude an Mathematik ausgeprägt ist, desto weniger ängstlich sind Jugendliche in Bezug auf Mathematik. Darüber hinaus zeigt sich zwischen Freude an Mathematik und instrumenteller Motivation ein deutlicher, positiver Zusammenhang, $r = .65$ ($p < .001$). Zwischen instrumenteller Motivation und Ängstlichkeit ergibt sich ein negativer Zusammenhang von $r = -.28$ ($p < .001$). Wer also Freude an Mathematik hat, der misst ihr gleichzeitig eine hohe Bedeutung für die eigene Zukunft bei und ist wenig ängstlich. Die eigene instrumentelle Motivation hängt außerdem positiv mit dem Wert zusammen, den Eltern und Freunde Mathematik zuschreiben, $r = .46$ ($p < .001$). Wenn Eltern und Freunde Mathematik wichtig finden, berichten die Jugendlichen zugleich selbst auch mehr Freude an Mathematik, $r = .36$ ($p < .001$).

Das mathematische Selbstkonzept korreliert mit der Selbstwirksamkeitserwartung zu $r = .49$ ($p < .001$). Darüber hinaus hängt das Selbstkonzept stärker als die Selbstwirksamkeitserwartung mit den emotionalen und motivationalen Orientierungen zusammen ($r_{\text{Freude}} = .68$, $p < .001$; $r_{\text{Ängstlichkeit}} = -.75$, $p < .001$; $r_{\text{instrumentelle Motivation}} = .54$, $p < .001$). Fünfzehnjährige, die demnach viel Zutrauen in ihre mathematischen Fähigkeiten haben, erleben gleichzeitig mehr Freude an Mathematik, eine höhere instrumentelle Motivation und weniger Ängstlichkeit.

Die mathematikbezogenen Einstellungen und Verhaltensweisen hängen vor allem mit Freude ($r_{\text{Arbeitshaltung}} = .45$, $p < .001$; $r_{\text{Intentionen}} = .50$, $p < .001$; $r_{\text{Beschäftigung mit Mathematik}} = .43$, $p < .001$) und instrumenteller Motivation ($r_{\text{Arbeitshaltung}} = .41$, $p < .001$; $r_{\text{Intentionen}} = .44$, $p < .001$; $r_{\text{Beschäftigung mit Mathematik}} = .36$, $p < .001$) zusammen. Umso höher die Freude an Mathematik und die instrumentelle Motivation ausgeprägt sind, umso höher sind auch die mathematikbezogenen Einstellungen und Verhaltensweisen. Besonders auffällig ist zusätzlich, dass die Absicht, sich auch über die Schulzeit hinaus mit Mathematik zu beschäftigen, mit dem mathematischen Selbstkonzept zu $r = .50$ ($p < .001$) korreliert.

3.4 Zusammenfassung und Diskussion

Im Rahmen dieses Kapitels wurden mathematikbezogene Schülermerkmale der fünfzehnjährigen Jugendlichen in Deutschland untersucht. Mehrdimensionale Bildungsziele berücksichtigen, dass Schule nicht nur eine fundierte Wissensbasis für das spätere Leben schafft, sondern gleichzeitig auch emotionale wie motivationale Orientierungen, Einstellungen und Verhaltensweisen gegenüber Mathematik prägt, die bedeutsame Grundlagen für ein lebenslanges Lernen sind.

Die Befunde zeichnen für Deutschland ein Bild mit Stärken aber auch einigen Schwächen, bei deren behutsamer Interpretation die diskutierten Einschränkungen des Designs berücksichtigt werden müssen. Erfreulich scheint, dass die Selbstbilder der Jugendlichen im internationalen Vergleich überdurchschnittlich positiv ausgeprägt sind. Demnach sind Fünfzehnjährige in Deutschland der Überzeugung, dass sie mathematische Aufgaben sicher lösen können und sie besitzen Vertrauen in ihre eigenen mathematischen Fähigkeiten. Gerade die Selbstwirksamkeitserwartungen haben sich im Vergleich zu 2003, als mathematikbezogene Schülermerkmale das letzte Mal bei PISA untersucht wurden, sowohl bei den Jungen als auch bei den Mädchen deutlich verbessert. Dies lässt sich darauf zurückführen, dass Jugendliche sich sicherer fühlen, Anwendungsaufgaben lösen zu können. Zwar erlauben die vorliegenden Daten keine kausalen Schlüsse, dennoch scheint es plausibel, dass durch die Einführung der länderübergreifenden Bildungsstandards (KMK, 2003) der Anwendungsbezug im Mathematikunterricht gestärkt wurde. Demnach können sich Jugendliche durch das häufigere Vorkommen dieser Aufgaben im Unterricht auch sicherer im Umgang mit ihnen fühlen.

Gleichzeitig erleben die Schülerinnen und Schüler in Deutschland – wie auch 2003 – im Mittel vergleichsweise wenig Ängstlichkeit in Bezug auf Mathematik. Leicht überdurchschnittlich ausgeprägt ist die verschiedentliche Beschäftigung mit Mathematik außerhalb des Unterrichts. Eine besonders intensive Auseinandersetzung wie zum Beispiel im Rahmen der Teilnahme an Mathematik-AGs oder Mathematikwettbewerben berichtet hingegen nur ein geringer Anteil der Jugendlichen. Hier besteht Verbesserungspotential gerade im Hinblick auf die Förderung von Talenten, die in besonderem Maße in der Lage sein sollten, Studiengänge zu wählen sowie Berufe auszuüben, die ein hohes mathematisches Verständnis voraussetzen. Dazu muss jedoch die Schule ein entsprechendes Angebot bereitstellen.

Die Beschäftigung mit Mathematik hängt vor allem mit einer ausgeprägten instrumentellen Motivation sowie mit Freude und Interesse an Mathematik zusammen. Die Freude an Mathematik ist bei Schülerinnen und Schülern in Deutschland jedoch eher gering ausgeprägt. Mehr als die Hälfte der Jugendlichen berichtet keine Freude und Interesse an Mathematik. Zusätzlich hat sich die Freude an Mathematik im Vergleich zu 2003 bedeutsam verringert. Sie ist sowohl bei den Jungen als auch bei den Mädchen gesunken. Dies ist gerade vor dem Hintergrund, dass Freude nicht nur die Lernbereitschaft, sondern eben auch die Beschäftigung mit Mathematik fördert, bedauerlich. Hier ist vor allem der Mathematikunterricht gefragt. Wie die Erfahrungen aus SINUS zeigen,

kann Mathematikunterricht, der sich durch herausfordernde Aufgaben auszeichnet und dem es gelingt, Schülerinnen und Schüler zu aktivieren, dazu beitragen, dass Jugendliche Freude an Mathematik entwickeln (Prenzel, Carstensen, Senkbeil, Ostermeier & Seidel, 2005).

Gestaltungsräume für Verbesserungen zeichnen sich auch bei der instrumentellen Motivation ab, die sich im Vergleich zu 2003 vor allem bei den Jungen verringert hat, auch wenn sie noch immer deutlich stärker ausgeprägt ist als bei den Mädchen. Die genauere Betrachtung zeigt zwar, dass zwei Drittel der Fünfzehnjährigen die besondere Bedeutung von Mathematik für ihr zukünftiges Ausbildungs- und Berufsleben erkennen. Sie geben ebenfalls an, dass ihre Eltern der Meinung sind, Mathematik sei wichtig für ihre Zukunft. Jedoch messen ihre Freundinnen und Freunde Mathematik weniger Bedeutung bei. Dies ist eher kritisch zu beurteilen, da in diesem Alter der Einfluss der Gleichaltrigen im Vergleich zur dem der Eltern immer größer wird (Brinthaup & Lipka, 2002). Am Gymnasium drückt sich dieses eher ungünstige Profil noch einmal deutlicher aus. Gerade hier sollten Maßnahmen ergriffen werden, die darauf abzielen, die Bedeutung und Relevanz von Mathematik zu vermitteln, um damit auch die Bereitschaft zu erhöhen, sich in der Schulzeit und danach verstärkt mit Mathematik auseinandersetzen zu wollen.

Weiterhin auffällig ist, dass es in den mathematikbezogenen Schülermerkmalen noch immer deutliche Unterschiede zwischen Mädchen und Jungen gibt. Dabei zeigen Jungen außer bei einer gewissenhaften Arbeitshaltung günstigere Merkmalsausprägungen als Mädchen. Sie haben mehr Freude, erkennen eher die Bedeutung von Mathematik, sind weniger ängstlich, haben die positiveren Selbstbilder und können sich eher vorstellen, sich auch außerhalb des Unterrichts und nach der Schulzeit weiterhin mit Mathematik zu beschäftigen. Besonders begünstigt sind Jungen bei den Selbstwirksamkeitserwartungen sowie dem Selbstkonzept und der Ängstlichkeit, auch wenn bei den letzteren beiden keine bedeutsamen Veränderungen im Vergleich zu 2003 zu beobachten sind.

Es gilt somit weiterhin das Ziel, günstige Ausprägungen bei den Schülermerkmalen gerade bei den Mädchen zu erreichen. Wenn es gelingt, die angesprochenen Verbesserungspotentiale umzusetzen, können sowohl Mädchen als auch Jungen über leistungsbezogene Voraussetzungen hinaus auf die zukünftigen Herausforderungen wie lebenslanges Lernen und gesellschaftliche und technische Entwicklungen vorbereitet werden, damit sie diese selbstbewusst und verantwortungsvoll meistern können.

Literatur

- Adams, R. J., Lietz, P. & Berezner, A. (2013). On the use of rotated context questionnaires in conjunction with multilevel item response models. *Large-scale Assessments in Education*, 1 (1). Zugriff am 28.10.2013. Verfügbar unter: <http://link.springer.com/article/10.1186%2F2196-0739-1-5>

- Ahmed, W., Minnaert, A., Kuyper, H. & van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. *Large-scale Cross-cultural Studies of Cognitive and Noncognitive Constructs*, 22, 385–389.
- Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology & Health*, 26, 1113–1127.
- Ajzen, I. & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology*, 22, 453–474.
- Ashcraft, M. H. & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment*, 2, 197–205.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.
- Brinthaupt, T. M., & Lipka, R. P. (2002). *Understanding early adolescent self and identity: Applications and interventions*. Albany, NY: State University of New York Press.
- de Fruyt, F., van Leeuwen, K., de Bolle, M. & de Clercq, B. (2008). Sex differences in school performance as a function of conscientiousness, imagination and the mediating role of problem behaviour. *European Journal of Personality*, 22, 167–184.
- Fend, H. (2008). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen* (2. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ferla, J., Valcke, M. & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, 19, 499–505.
- Hascher, T. (2004). *Wohlbefinden in der Schule* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 40). Münster: Waxmann.
- Heine, S. J., Lehman, D. R., Peng, K. & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, 82, 903–918.
- Helmke, A. (2007). *Unterrichtsqualität erfassen, bewerten, verbessern* (5. Aufl.). Seelze: Klett Kallmeyer.
- Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, 2, 260–280.
- Jerusalem, M. & Hopf, D. (2002). *Selbstwirksamkeit und Motivationsprozesse in Bildungsinstitutionen*. 44. Beiheft der Zeitschrift für Pädagogik. Weinheim: Beltz.
- Johnson, T. P., Shavitt, S. & Holbrook, A. L. (2011). Survey response styles across cultures. In D. R. Matsumoto & F. J. R. d. van Vijver (Hrsg.), *Cross-cultural research methods in psychology* (Culture and psychology, S. 130–178). New York: Cambridge University Press.
- Klassen, R. M. & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging direction. In T. C. Urdan & S. A. Karabenick (Hrsg.), *The decade ahead: Theoretical perspectives on motivation and achievement* (Advances in motivation and achievement, Bd. 16, S. 1–33). Bingley, U.K.: Emerald.
- Klieme, E. & Vieluf, S. (2013). Schulische Bildung im internationalen Vergleich. Ein Rahmenmodell für Kontextanalysen in PISA. In N. Jude & E. Klieme (Hrsg.), *PISA 2009 – Impulse für die Schul- und Unterrichtsforschung*. 59. Beiheft der Zeitschrift für Pädagogik (S. 229–246). Weinheim: Beltz Juventa.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik. (2003). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss (10. Jahrgangsstufe)*. Beschluss vom 4.12.2003 (Beschlüsse der Kultusministerkonferenz), Bonn.

- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik. (2010). *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Köln: Carl Link.
- Köller, O., Trautwein, U., Lüdtke, O. & Baumert, J. (2006). Zum Zusammenspiel von schulischer Leistung, Selbstkonzept und Interesse in der gymnasialen Oberstufe. *Zeitschrift für Pädagogische Psychologie*, 20, 27–39.
- Krapp, A. & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education*, 33, 27–50.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 51). Münster: Waxmann.
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Large-scale Cross-cultural Studies of Cognitive and Noncognitive Constructs*, 19, 355–365.
- Lipnevich, A. A., MacCann, C., Krumm, S., Burrus, J. & Roberts, R. D. (2011). Mathematics attitudes and mathematics outcomes of U.S. and Belarusian middle school students. *Journal of Educational Psychology*, 103, 105–118.
- Ma, X. & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, 28, 26–47.
- Marsh, H. W. & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, 81, 59–77.
- Marsh, H. W., Xu, K. & Martin, A. J. (2012). Self-concept: A synergy of theory, method, and application. In K. R. Harris, S. Graham & T. C. Urdan (Hrsg.), *APA educational psychology handbook* (S. 427–458). Washington, DC: American Psychological Association.
- Möller, J., Retelsdorf, J., Köller, O. & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, 48, 1315–1346.
- OECD. (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
- OECD. (2013b). *Ready to learn: Student engagement, attitudes and motivation*. Paris: OECD.
- Pekrun, R. & Linnenbrink-Garcia, L. (Hrsg.). (2014). *Handbook of emotions and education* (Educational psychology handbook series). New York: Francis & Taylor / Routledge.
- Pekrun, R., Vom Hofe, R., Blum, W., Frenzel, A. C., Götz, T. & Wartha, S. (2007). Development of mathematical competencies in adolescence. The PALMA longitudinal study. In M. Prenzel (Hrsg.), *Studies on the educational quality of schools: The final report on the DFG priority programme* (S. 17–38). Münster: Waxmann.
- Pekrun, R. & Zirngibl, A. (2004). Schülermerkmale im Fach Mathematik. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs* (S. 191–210). Münster: Waxmann.
- Prenzel, M. (2011). Empirische Bildungsforschung morgen: Reichen unsere bisherigen Forschungsansätze aus? In M. Gläser-Zikuda, T. Seidel, C. Rohlf, A. Gröschner & S. Ziegelbauer (Hrsg.), *Mixed Methods in der empirischen Bildungsforschung* (S. 273–286). Münster: Waxmann.
- Prenzel, M., Carstensen, C. H., Senkbeil, M., Ostermeier, C. & Seidel, T. (2005). Wie schneiden SINUS-Schulen bei PISA ab? Ergebnisse der Evaluation eines Modellversuchsprogramms. *Zeitschrift für Erziehungswissenschaft*, 8, 487–501.

- Purves, A. C. (1987). The evolution of the IEA: A memoir. *Comparative Educational Review*, 31, 10–28.
- Rheinberg, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch Pädagogische Psychologie* (S. 178-186). Göttingen: Hogrefe.
- Reiss, K. & Hammer, C. (2013). *Grundlagen der Mathematikdidaktik. Eine Einführung für den Unterricht in der Sekundarstufe*. Basel: Birkhäuser.
- Tenorth, H.-E. (1994). „Alle alles zu lehren“. *Möglichkeiten und Perspektiven allgemeiner Bildung*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- van de Gaer, E., Grisay, A., Schulz, W. & Gebhardt, E. (2012). The reference group effect: An explanation of the paradoxical relationship between academic achievement and self-confidence across countries. *Journal of Cross-Cultural Psychology*, 43, 1205–1228.
- Wang, M.-T. (2012). Educational and career interests in math: A longitudinal examination of the links between classroom environment, motivational beliefs, and interests. *Developmental Psychology*, 48, 1643–1657.
- Wang, M.-T., Eccles, J. S. & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science*, 24, 770–775.

4 Mathematikunterricht in Deutschland: Befunde aus PISA 2012

Anja Schiepe-Tiska, Kristina Reiss, Andreas Obersteiner,
Jörg-Henrik Heine, Tina Seidel und Manfred Prenzel

Mathematik lernt man kaum in informellen Kontexten, sondern hauptsächlich im Unterricht. Während der allgemeinen Schulpflichtzeit von mindestens neun Jahren kommen weit mehr als 1000 Schulstunden zusammen, in denen die Grundlagen mathematischer Bildung aufgebaut werden. Eine zentrale Rolle spielt dabei die Lehrperson: Erfolgreiches Lernen hängt von gutem Unterricht und damit entscheidend vom Verhalten der Lehrerin oder des Lehrers ab. Folgt man den von Hattie (2009) zusammengefassten Metaanalysen und seiner Rangreihe der Variablen, die mit Schulerfolg in Verbindung gebracht werden können, dann lassen sich acht der zwölf einflussreichsten Faktoren der Lehrperson beziehungsweise dem Lehren zuordnen.

Unterricht folgt allerdings nicht dem einfachen Muster, dass Lehren auch Lernen bewirkt. Vielmehr stellt Unterricht ein Angebot bereit, das von den Lernenden aktiv und konstruktiv genutzt werden muss. Diese Auffassung von Unterricht als Lernangebot liegt auch PISA 2012 zugrunde. Merkmale der Unterrichtsqualität werden durch die Befragung der Fünfzehnjährigen über das von ihnen wahrgenommene Angebot erfasst. Wie aber nehmen Schülerinnen und Schüler in Deutschland den von ihren Lehrerinnen und Lehrern angebotenen Mathematikunterricht wahr? Das vorliegende Kapitel stellt hierzu die theoretischen Annahmen vor und informiert über die Ergebnisse der PISA-Studie.

4.1 Merkmale guten Unterrichts

4.1.1 Befunde aus Metaanalysen

Erfolgreiches Lernen kann nicht einfach von außen (z. B. durch eine Lehrperson) „angeschaltet“ werden (Prenzel, 1995), sondern setzt sich aus einem Wechselspiel vielfältiger Faktoren zusammen. So üben individuelle Voraussetzungen wie Vorwissen, Interesse und selbstbezogene Vorstellungen einen starken Einfluss auf das Lernen aus (Helmke & Weinert, 1997). Aber auch viele kontextuelle und systembezogene Faktoren, allen voran die positive Unterstützung in Familie, Schule und Unterricht, spielen eine wichtige Rolle. In den von Hattie (2009) zusammengefassten Metaanalysen wurden entsprechend potentielle Einflussfaktoren hinsichtlich ihrer Wirksamkeit für den Lernerfolg – gemessen

über die Effektstärke – untersucht und angeordnet. Ausgangspunkt war die Zuordnung der Variablen zu einer von sechs thematischen Gruppen, nämlich Lernende, Elternhaus, Schule, Lehrperson, Curricula und Unterrichten. Im Ergebnis erwiesen sich die Variablen, die die Lehrperson betrafen, als stärkste Prädiktoren für den Lernerfolg, gefolgt von denen zu den Curricula und zum Unterrichten.

In Bezug auf die Lehrperson und das Unterrichten fällt auf, dass viele der wirksamsten Variablen mit der Interaktion zwischen Lehrenden und Lernenden zusammenhängen. So nimmt die formative Evaluation des Unterrichts, also die Bewertung von unterrichtlichen Situationen im Hinblick auf notwendige Interventionen oder Korrekturen der eigenen Arbeit, einen vorderen Platz ein. Ebenso wichtig ist es, Schülerinnen und Schülern Rückmeldungen zu geben und eine positive Erwartungshaltung gegenüber ihren möglichen Kompetenzfortschritten zum Ausdruck zu bringen. Darüber hinaus führt ein kognitiv aktivierender Unterricht zu guten Lernerfolgen. Auch wenn es aus unterschiedlichen Richtungen Kritik an dem von Hattie gewählten Vorgehen gibt – beispielsweise hinsichtlich der Methodik oder bezüglich der Auswahl der einbezogenen Studien – stimmen die Erkenntnisse im Wesentlichen mit dem Forschungsstand zu Unterrichtseffektivität überein und betonen die bedeutende Rolle der Lehrenden und des Unterrichts. Terhart (2011) stellt in diesem Zusammenhang fest: „Hattie rehabilitiert den dominanten, redenden Lehrer – der aber ebenso auch genau weiß, wann er zurücktreten und schweigen muss“ (S. 289). Damit nimmt Hattie eine lehrerzentrierte Perspektive auf den Unterricht ein, in dem die Schülerinnen und Schüler im Mittelpunkt stehen. Diese Sichtweise bestätigen auch neuere Studien, in denen die Bedeutung der aktiven Rolle von Lehrenden in der Begleitung des Lernens ihrer Schülerinnen und Schüler thematisiert wird (vgl. Furtak, Seidel, Iverson & Briggs, 2012). Auf die Lehrerinnen und Lehrer und ihr Verhalten kommt es demnach in besonderer Weise an.

Um Unterrichtsvariablen zu gruppieren, stützt sich Hattie (2009) auf die Schlüsselwörter, die von den Autoren der einbezogenen Studien genannt wurden. Leider ist es nicht so, dass diese Schlüsselwörter von allen Autoren gleich verstanden und verwendet werden. Deshalb besteht die Gefahr, dass in einer Unterrichtsrubrik sehr unterschiedliche Umsetzungen von Unterrichtspraktiken zusammengefasst werden. Diese generelle Problematik von Metaanalysen kann zu kaum interpretierbaren Ergebnissen führen (vgl. Cooper & Hedges, 1994). Eine Metaanalyse von Seidel und Shavelson (2007) kategorisiert aus diesem Grund die auf den Unterricht bezogenen Variablen auf der Basis eines theoretischen Lehr-Lern-Modells. Diese Studie bezieht außerdem nicht nur kognitive, sondern auch motivational-affektive Ergebnisse des Lernens ein. Die Befunde verdeutlichen, dass unterrichtsbezogene Aktivitäten, die vor allem auf die Vermittlung fachbezogenen Wissens (etwa mathematisches Argumentieren und Modellieren, naturwissenschaftliches Denken und Arbeiten) abzielen und bei denen die Schülerinnen und Schüler in ihren Lernprozessen durch die Lehrenden positiv unterstützt und begleitet werden, deutliche Effekte auf das Lernen der Schülerinnen und Schüler haben. Diese Effekte zeigen sich in Bezug auf Lernprozesse wie zum Beispiel das Erleben intrinsischer Motivation, auf das Ausführen von kognitiven Lernaktivitäten wie Elaborieren und Organi-

sieren sowie auf längerfristige Lernergebnisse. Deutlich wird dabei, dass diese Faktoren kognitive *und* motivational-affektive Aspekte des Lernens beeinflussen.

4.1.2 Unterrichtsmuster

Es ist umstritten, ob Rangreihen effektiver Unterrichts- und Lehrervariablen nicht allzu isoliert einzelne Faktoren herausstellen, die eigentlich besser im Zusammenspiel und in einem größeren Zusammenhang betrachtet werden müssten (Cohen, Raudenbush & Loewenberg Ball, 2003; Floden, 2001). So zeigen manche Variablen wie Klassenmanagement und Strukturiertheit deutliche Effekte auf kognitive Lernergebnisse und gleichzeitig keine beziehungsweise schwache Effekte auf motivational-affektive Aspekte des Lernens (Kunter, Baumert & Köller, 2007; Lipowsky, Rakoczy, Pauli, Drollinger-Vetter, Klieme & Reusser, 2009; Rakoczy, Klieme, Drollinger-Vetter, Lipowsky, Pauli & Reusser, 2007). Für andere Variablen ergibt sich ein spiegelbildliches Muster. So zeigen zum Beispiel schüleraktivierende Unterrichtsformen positive Effekte auf die Motivation und das Interesse der Lernenden, aber schwächere bis keine Effekte auf kognitive Lernergebnisse (Seidel, Prenzel & Rimmel, 2003; Seidel, Prenzel, Wittwer & Schwindt, 2007).

Besonders relevant für die Betrachtung solcher Unterrichtsvariablen sind die in den vergangenen Jahren durchgeführten Videosurveys, die den Einsatz von Fragebögen mit Videoaufnahmen des Unterrichts kombinieren (Pauli & Reusser, 2006; Seidel, 2011). Diese Studien konnten nicht nur spezifische Qualitätsmerkmale bezüglich Lehrer-Schüler-Interaktionen oder effizienter Zeitnutzung ausmachen, sondern auch komplexere Muster von Unterricht identifizieren. Hier zeigte sich, dass guter Unterricht nur unzureichend durch das einfache Nebeneinanderstellen als wirksam geltender Maßnahmen zu charakterisieren ist. Entscheidend ist vielmehr, wie im Unterricht unterschiedliche Faktoren arrangiert werden und zusammenwirken (Oser & Baeriswyl, 2001; Seidel et al., 2007).

Der TIMSS¹-Videostudie (Stigler, Gonzales, Kawanaka, Knoll & Serrano, 1996) kommt dabei eine besondere Bedeutung zu, da in dieser Studie erstmals Unterschiede im Mathematikunterricht zwischen Deutschland, Japan und USA über Videoaufzeichnungen erfasst und beschrieben wurden. Der prototypische Mathematikunterricht in Deutschland erwies sich dabei als fragend-entwickelnd mit hohen Redeanteilen der Lehrpersonen und eher geringen Beiträgen der Schülerinnen und Schüler (Blum, 2001; Klieme, Schümer & Knoll, 2001; vgl. Baumert, Kunter, Brunner, Krauss, Blum & Neubrand, 2004). Die Betonung lag auf Algorithmen und Routineaufgaben, weniger auf Problemlösen, welches etwa im japanischen Unterricht eine wichtige Rolle spielte. Obwohl diese Aspekte der Unterrichtsgestaltung jeweils für sich genommen nicht unbedingt negativ zu bewerten sind, erweist sich ein solches Profil des Mathematikunterrichts in der Summe als problematisch, vor allem dann, wenn verständnisvolles Lernen

1 *Third International Mathematics and Science Study.*

und der Aufbau mathematischer Kompetenzen (bzw. mathematischer Grundbildung, vgl. Kapitel 2) Ziele des Unterrichts sein sollen. Tatsächlich waren die Mathematikleistungen der Schülerinnen und Schüler in Deutschland in TIMSS genauso wie in der ersten PISA-Erhebung enttäuschend, weil die Jugendlichen vor allem Schwierigkeiten hatten, ihr Wissen anzuwenden und anspruchsvollere mathematische Aufgaben zu lösen.

4.1.3 Unterricht als Lernangebot

Aus den bisherigen Betrachtungen könnte man den Eindruck gewinnen, dass es allein die Lehrperson ist, die durch ihre Gestaltung des Unterrichts den Lernfortschritt der Schülerinnen und Schüler beeinflusst. Der lehrerbezogene Blick auf den Unterricht blendet aber aus, dass Lehren und Lernen ein Wechselspiel zwischen unterschiedlichen Akteuren ist. Deutlich wird dieser Aspekt in den so genannten Angebots-Nutzungs-Modellen (z.B. Helmke & Weinert, 1997). Sie gehen von Angebotsstrukturen im Sinne von Lerngelegenheiten aus und verbinden sie mit Formen der Nutzung durch die Schülerinnen und Schüler. Auf der Seite der Lehrenden stehen somit beispielsweise Ziele, Lehrinhalte, Darbietungsformen, Aufgaben oder die Erfolgskontrolle; auf der Seite der Lernenden findet man etwa deren Lernaktivitäten, Interesse und Motivation. Beide Aspekte bedingen eine weitere Komponente des Modells, nämlich die Ergebnisse von Unterricht. Da die Lehrperson letztlich nur ein *Lernangebot* bereitstellen kann, kommt in diesem Modell den Lernenden die aktive Rolle zu, das Lernangebot auch zu nutzen. Von dieser aktiven Nutzung der Lernangebote hängt es ab, wie sehr die Schülerinnen und Schüler vom Unterricht profitieren.

Um welche Lernangebote geht es nun speziell im Mathematikunterricht? Zwei Aspekte stehen im Vordergrund, einerseits, welche mathematischen Themen im Unterricht behandelt werden, und andererseits, auf welche Art und Weise mit diesen Themen umgegangen wird, um die Ziele des Lehrplans zu erreichen und die gewünschten Kompetenzen aufzubauen. Während die zu behandelnden Themen weitgehend durch das Curriculum vorgegeben sind, hat die Lehrperson Entscheidungsspielräume, wenn es darum geht, sie den Schülerinnen und Schülern zum Beispiel in Form konkreter Aufgabenstellungen zugänglich zu machen. Hier gilt es, die Lernvoraussetzungen der Schülerinnen und Schüler zu berücksichtigen und zielbezogene Lernprozesse anzuregen. So kann es einen großen Unterschied machen, ob ein mathematischer Inhalt mit oder ohne Anwendungsbezug präsentiert wird. Bei einer linearen Gleichung (z.B. $0,05x + 5 = 0,08x$) könnte es beispielsweise vorwiegend um das Durchführen algorithmischer Schritte bis zur Lösung gehen („Variablen auf die eine Seite, Zahlen auf die andere Seite“) oder aber um die Mathematisierung eines realen Kontextes („Wie lange muss man telefonieren, damit es sich lohnt, von einem Tarif ohne Grundgebühr und einem Minutenpreis von 8 Cent in einen Tarif mit 5 Euro monatlicher Grundgebühr und einem Minutenpreis von 5 Cent zu wechseln?“).

Aufgaben können auch in unterschiedlichem Maße geeignet sein, mathematische Inhalte zu vernetzen oder Problemlöseprozesse zu initiieren und damit auf sehr unterschiedliche Facetten mathematischer Kompetenz (vgl. Kapitel 2) abzielen. Die Art der Aufgabenstellung spielt im Mathematikunterricht eine zentrale Rolle, weil Mathematik oft über Aufgaben kommuniziert wird und diese üblicherweise in allen Phasen des Unterrichts eingesetzt werden. Aus fachdidaktischer Sicht scheint insbesondere die Verwendung solcher Aufgaben sinnvoll, bei denen Schülerinnen und Schüler zum Nachdenken angeregt werden, weil es beispielsweise mehrere Lösungswege gibt oder die richtige Lösung nicht nur gefunden, sondern auch begründet werden muss (vgl. Reiss & Hammer, 2013).

4.2 Fragen und Erwartungen

Unterrichtsmuster ändern sich nicht in kurzen Zeiträumen, weil sie auf eingespielten Routinen beruhen. Seit TIMSS und PISA 2003, als der Mathematikunterricht das letzte Mal schwerpunktmäßig untersucht wurde (Baumert et al., 2004; Prenzel et al., 2006), sind aber nicht nur einige Jahre vergangen, sondern es wurden auch verschiedene Maßnahmen zur Sicherung und Weiterentwicklung der Unterrichtsqualität ergriffen. Sowohl die länderübergreifenden Bildungsstandards (KMK, 2003) als auch Schulentwicklungs- und Fortbildungsprogramme wie SINUS („Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“; vgl. Prenzel, Friedrich & Stadler, 2009) waren insbesondere darauf ausgerichtet, einen kognitiv aktivierenden, abwechslungsreichen und anspruchsvollen Mathematikunterricht weiterzuentwickeln. Im Zuge dieser Entwicklung gab es eine breite Diskussion über eine neue Aufgabenkultur im Mathematikunterricht. Auswirkungen dieser Maßnahmen könnten sich deshalb bei PISA 2012 in einer verstärkten Verwendung anwendungsbezogener und kognitiv aktivierender Aufgaben zeigen.

Aus einer theoretischen Perspektive sollte qualitativ hochwertiger Mathematikunterricht aber nicht nur kognitiv anspruchsvoll, sondern gleichzeitig störungsarm, zielorientiert und auf die individuellen Bedürfnisse der Schülerinnen und Schüler abgestimmt sein. Diese Merkmale werden deshalb anhand der bei PISA 2012 erhobenen Daten ebenfalls berücksichtigt und zunächst einzeln betrachtet, um ein differenziertes Bild zu erhalten. Außerdem wird durch die zusammenhängende Betrachtung einzelner Items untersucht, wie diese Merkmale im Sinne von Unterrichtsmustern zusammenwirken (Dalehefte, 2007; Hugener, Pauli, Reusser, Lipowsky, Rakoczy & Klieme, 2009; Kobarg et al., 2011). Qualitativ hochwertiger Mathematikunterricht sollte sich in einem Muster zeigen, bei dem die eben genannten Kriterien gleichzeitig in hohem Maße erfüllt sind. Nun ist nicht zu erwarten, dass fünfzehnjährige Schülerinnen und Schüler allesamt ihren Mathematikunterricht genau so erleben und beschreiben können. Es wäre aber wünschenswert, wenn die Analysen zeigten, dass es einen solchen Mathematikun-

terricht in Deutschland gibt beziehungsweise dass er von einem erheblichen Anteil der bei PISA 2012 befragten Jugendlichen so erlebt wird.

4.3 Erfassung des Unterrichts bei PISA

PISA erhebt und vergleicht in erster Linie Kompetenzen im Sinne von Bildungsergebnissen in großen, repräsentativen Stichproben und kann aufgrund des gewählten Forschungszugangs nur einen begrenzten Blick auf den Unterricht eröffnen (Kobarg et al., 2011; Kunter, Dubberke, Baumert, Blum & Jordan, 2006). Insbesondere ist die vertiefende Betrachtung von Unterricht, so wie sie beispielsweise über Videostudien und über wiederholte Befragungen und Tests möglich ist (Pauli & Reusser, 2006), nicht das primäre Anliegen von PISA. Vielmehr fokussiert das internationale Erhebungskonzept von PISA darauf, wie sich das Angebot von Unterricht in fachlicher, didaktischer und pädagogischer Hinsicht *in der individuellen Wahrnehmung der Jugendlichen* abbildet (OECD, 2013). Alle Unterrichtsmerkmale werden daher mit Hilfe des Schülerfragebogens erfasst, der bei PISA 2012 erstmals in einem Rotationsdesign eingesetzt wurde (vgl. Kapitel 1 und 3). Dementsprechend wurde jede Frage nur von zwei Drittel der Schülerinnen und Schüler beantwortet. Die eingesetzten statistischen Verfahren ermöglichen aber dennoch eine robuste Beschreibung des erlebten Mathematikunterrichts (Adams, Lietz & Berezner, 2013).

Methodisch ist die Erfassung von Unterrichtsmerkmalen durch Schülereinschätzungen, wie es bei PISA geschieht, nicht unproblematisch, da sicherzustellen ist, dass die Schülerinnen und Schüler die Qualität des Lernangebots auch zuverlässig beurteilen können (vgl. Prenzel & Lankes, 2013). Bei PISA 2012 wurden zum einen bestimmte Merkmale und Praktiken des Mathematikunterrichts erfragt, beispielsweise wie oft die Lehrperson Rückmeldungen gibt oder wie häufig Störungen auftreten. Zum anderen wurde die Vertrautheit mit fachspezifischen Merkmalen wie mathematischen Begriffen oder unterschiedlichen Aufgabentypen im Unterricht erhoben. Diese Verfahren ermöglichen eine einigermaßen zuverlässige Beschreibung des Unterrichts durch Schülerinnen und Schüler (Ditton, 2000). Dies trifft vor allem dann zu, wenn die Jugendlichen, die die Fragen beantworten, einen gemeinsamen kulturellen Hintergrund beziehungsweise vergleichbare schulische Erfahrungen haben. Das Bezugssystem, anhand dessen zum Beispiel Schülerinnen und Schüler in der Türkei ihren Unterricht beurteilen, unterscheidet sich vermutlich von dem, das Schülerinnen und Schüler in Deutschland verinnerlicht haben, und noch mehr von dem Jugendlicher in Mexiko oder Korea. Wegen des variierenden kulturellen Hintergrunds der Schülerinnen und Schüler bei PISA unterliegen Schülerbefragungen zum Unterricht im internationalen Vergleich besonderen Herausforderungen. Auch wenn man innerhalb von Deutschland verschiedene Schularten miteinander vergleicht, zeigen sich Unterschiede in den Vergleichsmaßstäben, die auf unterschiedlichen subjektiven Erfahrungswelten beruhen (z. B. Hauptschülerinnen und

Hauptschüler versus Gymnasiastinnen und Gymnasiasten; Baumert et al., 2004). Deshalb sind entsprechende Vergleiche behutsam zu interpretieren.

Eine weitere Einschränkung im Hinblick auf die Interpretation der Ergebnisse hängt mit der Stichprobe von PISA zusammen: International werden Zufallsstichproben von fünfzehnjährigen Schülerinnen und Schülern betrachtet, die sich in unterschiedlichen Klassen und zum Teil in unterschiedlichen Jahrgangsstufen befinden. In der aktuellen Erhebung 2012 besuchten die Fünfzehnjährigen in Deutschland die Jahrgangsstufen 7 bis 11, etwas mehr als die Hälfte besuchte die Jahrgangsstufe 9 (vgl. Kapitel 1). Es werden somit individuelle – und keine auf der Ebene von Klassen aggregierten – Wahrnehmungen des Unterrichts untersucht. Daten, die auf vollständigen Klassensätzen beruhen, wären jedoch für die Beschreibung des von Schülerinnen und Schülern wahrgenommenen Unterrichts aussagekräftiger (vgl. Prenzel & Lankes, 2013). Insofern werden die in Deutschland zusätzlich erhobenen Daten in vollständigen 9. Klassen (vgl. Kapitel 1) sicherlich das nationale Bild ergänzen, wenn sie auch für den internationalen Vergleich nicht herangezogen werden können. Da in diesem Beitrag der Mathematikunterricht in Deutschland international eingeordnet werden soll, konzentrieren sich die Analysen hier auf den internationalen Datensatz, der auf der Stichprobe der Fünfzehnjährigen basiert. Die nationalen Daten der zusätzlichen Stichprobe von zwei 9. Klassen werden im Zusammenhang mit einer Follow-up-Erhebung (auf der Klassenstufe 10) systematisch ausgewertet.

Darüber hinaus sind die Angaben der Schülerinnen und Schüler zur Wahrnehmung der Unterrichtsmerkmale eine Momentaufnahme des im laufenden Schuljahr erlebten Unterrichts. Das Abschneiden beim Leistungstest ist hingegen das Ergebnis eines mehrjährigen Lernprozesses. Die kumulierten Unterrichts- und Lernerfahrungen aus den vergangenen Schuljahren können in Bezug auf die Einschätzungen zum Unterricht, wie sie hier vorliegen, nicht berücksichtigt werden. Demnach sollten Zusammenhänge zwischen Unterrichtsmerkmalen und mathematischer Kompetenz nur begrenzt analysiert und mit Vorsicht interpretiert werden. Es wird deshalb auf die Darstellung einfacher Korrelationen verzichtet, die Fehlinterpretationen dieser Zusammenhänge nahelegen könnten. Was hingegen berichtet wird, sind einige anspruchsvollere Analysen, die Beziehungen zwischen Unterrichtsmustern und mathematischer Kompetenz sowie mathematikbezogener Freude und Selbstwirksamkeitserwartungen darstellen.

Eingesetzte Skalen bei PISA 2012

Die Erfassung von Unterrichtsmerkmalen konzentriert sich bei PISA 2012 auf solche Variablen, die sich in früheren Studien als zuverlässig und aussagekräftig für die Lernentwicklung der Schülerinnen und Schüler erwiesen haben (Klieme, Lipowsky, Rakoczy & Ratzka, 2006; Klieme, Pauli & Reusser, 2009; Klieme & Rakoczy, 2008; vgl. Waldis, Grob, Pauli & Reusser, 2010). In der Unterrichtsforschung konnten im Zuge der ersten PISA-Erhebungsrunden in den Jahren 2000 und 2003 sowie im Rahmen der daran angebandenen COACTIV-Studie (vgl. Kunter et al., 2011) Fortschritte in der Entwicklung

empirisch gestützter Modelle der Unterrichtsqualität und entsprechender Messinstrumente gemacht werden. Einige der dort entwickelten Skalen wurden auch bei PISA 2012 verwendet, andere wurden neu entwickelt.

Die Darstellungen in diesem Kapitel beschränken sich auf diejenigen erhobenen Unterrichtsmerkmale, welche speziell den Mathematikunterricht betreffen. Eine Übersicht über diese Skalen bietet Tabelle 4.1, die zur Illustration ein Beispielitem für jede Skala enthält. Ferner enthält die Tabelle Angaben zur Anzahl der Items pro Skala und zur Reliabilität. Mit Werten zwischen .67 und .90 für Cronbachs Alpha liegen die Reliabilitäten für alle Skalen im akzeptablen Bereich, sodass davon ausgegangen werden kann, dass die Items einer jeweiligen Skala tatsächlich das gleiche Konstrukt erfassen. Die Skalenwerte, die Gegenstand der folgenden Analysen sein werden, sind sogenannte WL-Schätzer („Weighted Likelihood Estimates“), also Werte, die auf der Basis von psychometrischen Modellen der Item-Response-Theorie geschätzt wurden (vgl. Kapitel 10).

Drei der Skalen, nämlich die von Schülerinnen und Schülern wahrgenommene *Disziplin* im Klassenzimmer, die *kognitive Aktivierung* und die *Unterstützung durch die Lehrperson*, können als Voraussetzungen für wirksamen Unterricht angesehen werden; sie werden deshalb mitunter auch als Basisdimensionen der Unterrichtsqualität bezeichnet (Klieme et al., 2009). Zur Erfassung der *Disziplin*, die einen Aspekt der Klassenführung anzeigt (Lipowsky, 2009; Seidel, 2009), wurden die Schülerinnen und Schüler beispielsweise danach gefragt, wie häufig es vorkommt, dass es im Klassenzimmer laut ist und drunter und drüber geht. Bei der Skala *kognitive Aktivierung* geht es im Wesentlichen darum, ob die von der Lehrperson im Mathematikunterricht gestellten Aufgaben in den Augen der Schülerinnen und Schüler geeignet sind, sie zum Nachdenken anzuregen, also kognitive Prozesse zu aktivieren. Die von den Schülerinnen und Schülern wahrgenommene konstruktive *Unterstützung durch die Lehrperson* wurde mit Items erfasst, die beispielsweise danach fragen, ob die Lehrperson die Schülerinnen und Schüler beim Lernen unterstützt.

Die Skalen *Lehrersteuerung*, *Schülerorientierung* und *Rückmeldung* fokussieren noch stärker auf das Verhalten der Lehrperson und spiegeln eingesetzte „Unterrichtspraktiken“ wider (vgl. Klieme & Vieluf, 2013). *Lehrersteuerung* erfasst, inwieweit die Lehrperson klare Ziele setzt, den Stoff vergangener Unterrichtsstunden zusammenfasst und das Unterrichtsgespräch aktiv lenkt, indem sie die Schülerinnen und Schüler auffordert, ihre Gedanken einzubringen, und das Verständnis des Stoffes mit gezielten Fragen prüft. Die Skala *Schülerorientierung* bezieht sich darauf, wie häufig die Klasse in Kleingruppen arbeitet oder längere Projekte durchführt, inwieweit die Schülerinnen und Schüler in die Planung von Lernaktivitäten eingebunden sind und wie häufig Maßnahmen der Binnendifferenzierung stattfinden, bei denen unterschiedlich schwere Aufgaben an leistungsstärkere und -schwächere Jugendliche vergeben werden. Die Skala *Rückmeldung* erhebt, wie häufig die Lehrperson den Schülerinnen und Schülern Rückmeldungen über deren individuelle Stärken und Schwächen sowie Tipps zum Weiterlernen gibt.

Tabelle 4.1: Skalen der Unterrichtsmerkmale bei PISA 2012 mit Beispielitems

Unterrichtsmerkmal	Anzahl der Items	Reliabilität*	Beispielitem
Disziplin	5	.89	Im Klassenzimmer ist es oft laut, und es geht drunter und drüber.
Kognitive Aktivierung	9	.79	Die Lehrerin/der Lehrer gibt uns Aufgaben, bei denen wir einige Zeit darüber nachdenken müssen.
Unterstützung durch die Lehrperson	5	.84	Unsere Lehrerin/ unser Lehrer unterstützt uns beim Lernen.
Lehrersteuerung	5	.68	Die Lehrerin/der Lehrer setzt uns klare Lernziele.
Schülerorientierung	4	.67	Die Lehrerin/der Lehrer lässt uns in Kleingruppen arbeiten, damit wir gemeinsam Lösungswege erarbeiten.
Rückmeldung	4	.72	Die Lehrerin/der Lehrer sagt mir, was ich tun kann, um mich in Mathematik zu verbessern.
Vertrautheit mit mathematischen Begriffen	13	.84	Wurzeln
Häufigkeit innermathematischer Aufgaben	3	.90	Eine Gleichung wie $2(x+3)=(x+3)(x-3)$ lösen.
Häufigkeit einfacher Anwendungsaufgaben	6	.72	Den wöchentlichen Verbrauch eines Elektrogerätes berechnen.
Häufigkeit außermathematischer Anwendungsaufgaben	1	–	siehe Abbildung 4.1

Anmerkung: *Die Reliabilität entspricht Cronbachs Alpha. Dieser Wert wurde auf der Basis der deutschen Stichprobe ermittelt.

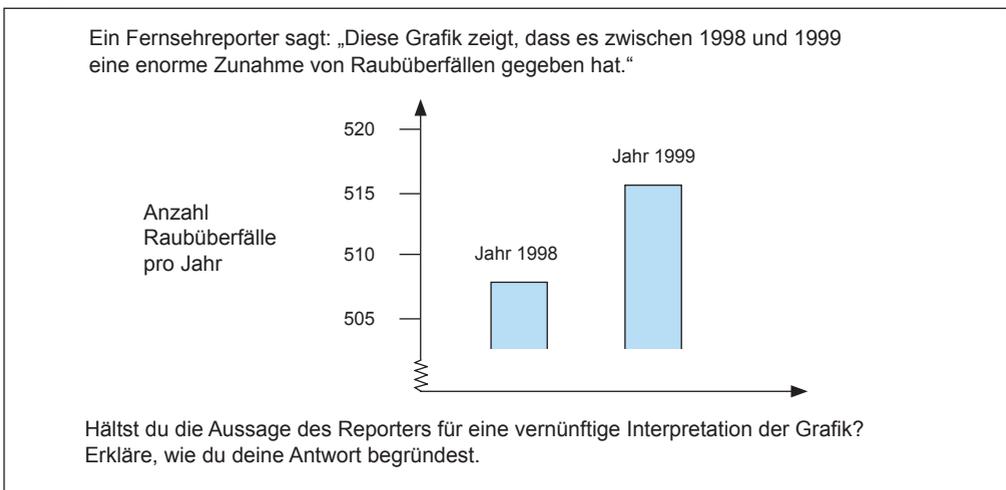
Mit Hilfe der Skala *Vertrautheit mit mathematischen Begriffen* wird erfasst, welchen mathematischen Inhalten die Schülerinnen und Schüler in ihrem Unterricht begegnet sind. Dazu wurde ihnen eine Liste mit 13 mathematischen Begriffen (z. B. „Quadratfunktion“, „Wurzeln“, oder „komplexe Zahl“) vorgelegt. Die Schülerinnen und Schüler sollten für jeden Begriff den Grad der Vertrautheit angeben. Die abgestuften Antwortmöglichkeiten reichten von der Aussage, dass man von einem Begriff noch nie gehört hat, bis zu der Feststellung, dass man ihn kennt und versteht.

Um zu erfassen, wie häufig bestimmte Typen von Aufgaben im Mathematikunterricht vorkamen, wurde den Befragten außerdem eine Liste mit kurzen Beispielaufgaben vorgelegt. Die Schülerinnen und Schüler sollten angeben, ob Aufgaben vom jeweiligen Typ „häufig“, „manchmal“, „selten“ oder „nie“ in ihrem Unterricht vorkamen. Bei drei dieser Aufgaben handelte es sich um innermathematische Aufgaben ohne Anwendungsbezug, etwa zum rein technischen Lösen einer Gleichung. Sechs weitere Aufgaben enthielten jeweils einen einfachen Anwendungskontext (einfache Anwendungsaufgaben) wie etwa die Berechnung des Kaufpreises eines Computers, wenn die Mehrwertsteuer hinzugerechnet werden muss.

Darüber hinaus wurde bei PISA 2012 die Häufigkeit bestimmter Typen von Mathematikaufgaben erstmals mit zusätzlichen Einzelitems untersucht, bei denen jeweils zwei Aufgabenbeispiele vorgegeben waren und die Schülerinnen und Schüler wiederum ange-

ben sollten, wie häufig solche Aufgabentypen in ihrem Mathematikunterricht vorkommen. Für die folgenden Analysen wird auf eines dieser Einzelitems zurückgegriffen, da es einen authentischen außermathematischen Anwendungsbezug aufweist und sich auf das aus mathematikdidaktischer Sicht bedeutende mathematische Argumentieren fokussiert. Abbildung 4.1 zeigt die beiden in diesem Item vorgegebenen Aufgabenbeispiele. Diese Beispiele sollten von den Schülerinnen und Schülern nicht gelöst werden, sondern dienen lediglich als Vorlage für die Häufigkeitseinschätzung des Aufgabentyps. Im ersten Beispiel geht es um die Interpretation einer grafischen Darstellung. Es soll die Aussage eines Reporters beurteilt werden und es ist dafür eine Begründung verlangt. Im zweiten Beispiel müssen die angegebenen Formeln verwendet werden, um sie auf den Kontext der maximalen empfohlenen Herzfrequenz zu beziehen. In beiden Fällen geht es also darum, Mathematik als Werkzeug zu verwenden, um eine reale Situation zu beurteilen. Die Aufgaben können damit insbesondere nicht durch ein fest einstudiertes Verfahren gelöst werden.

Beispiel 1:



Beispiel 2:

Das Verhältnis zwischen empfohlener maximaler Herzfrequenz und dem Alter einer Person wurde lange mit der folgenden Formel beschrieben:

$$\text{Maximale empfohlene Herzfrequenz} = 220 - \text{Alter}$$

Neuere Forschungsergebnisse haben nun gezeigt, dass diese Formel leicht angepasst werden sollte. Die neue Formel lautet:

$$\text{Maximale empfohlene Herzfrequenz} = 208 - (0,7 \cdot \text{Alter})$$

Ab welchem Alter beginnt die maximale empfohlene Herzfrequenz gemäß der neuen Formel zu steigen? Gib deinen Lösungsweg an.

Abbildung 4.1 Aufgabenbeispiele für ein Item zur Erfassung des Vorkommens außermathematischer Anwendungsaufgaben im Unterricht

4.4 Ergebnisse

Wie wird der Mathematikunterricht in Deutschland aus der Sicht der Schülerinnen und Schüler wahrgenommen? Um diese Frage zu beantworten, werden zunächst die Einschätzungen der Unterrichtsmerkmale anhand ausgewählter Items beschrieben. Im Anschluss daran werden die Ergebnisse in den internationalen Vergleich eingeordnet. Weiterführende Analysen untersuchen das Zusammenspiel einzelner Unterrichtsmerkmale, indem Muster des Mathematikunterrichts identifiziert werden. Ferner wird betrachtet, wie häufig diese Muster in Deutschland und in ausgewählten Vergleichsstaaten auftreten. Abschließend soll der Zusammenhang dieser Unterrichtsmuster mit mathematischer Kompetenz, Freude und Interesse an Mathematik sowie mathematikbezogenen Selbstwirksamkeitserwartungen in Deutschland dargestellt werden.

4.4.1 Charakterisierung des Mathematikunterrichts in Deutschland anhand ausgewählter Items

In den Tabellen 4.2 bis 4.4 ist dargestellt, wie häufig Schülerinnen und Schüler einzelne Merkmale und Vorkommnisse in ihrem Mathematikunterricht erleben. Diese Darstellung verhaltens- und ereignisnaher Häufigkeitseinschätzungen auf der Ebene einzelner Items ist geeignet, gewisse Stärken und Schwächen des Mathematikunterrichts in Deutschland zu erkunden. Die Tabelle zeigt für jedes Unterrichtsmerkmal ein Beispielitem sowie die prozentualen Zustimmungen der Fünfzehnjährigen in Deutschland zu den möglichen Antwortkategorien. Die Beispielitems sind so ausgewählt, dass sie die Häufigkeitsangaben für die jeweilige Skala möglichst gut repräsentieren. Zunächst werden die Basisdimensionen der Unterrichtsqualität und die Unterrichtspraktiken betrachtet, anschließend die Vertrautheit mit mathematischen Begriffen und Aufgabentypen.

Disziplin

Die Unterrichtsdisziplin wird von den Schülerinnen und Schülern in Deutschland recht positiv eingeschätzt (vgl. Tabelle 4.2). Mehr als zwei Drittel der Jugendlichen geben an, dass höchstens „in einigen Stunden“ gewartet werden muss, bis alle Schülerinnen und Schüler in der Klasse ruhig werden. Allerdings muss auch gesehen werden, dass anhaltende Unruhe für immerhin 10,9 Prozent offenbar die Regel ist und für über ein Fünftel der Schülerinnen und Schüler zumindest „in den meisten Stunden“ vorkommt. Insgesamt kann man aber den Eindruck gewinnen, dass Mathematikunterricht in Deutschland eher diszipliniert stattfindet.

Tabelle 4.2: Prozentuale Antworthäufigkeiten für Beispielitems der Skalen zu den Basisdimensionen der Unterrichtsqualität

Unterrichtsmerkmal mit Beispielitem	Prozentuale Zustimmung			
	In jeder Stunde % (SE)	In den meisten Stunden % (SE)	In einigen Stunden % (SE)	Nie oder fast nie % (SE)
Disziplin				
Die Lehrerin/der Lehrer muss lange warten, bis die Schülerinnen und Schüler ruhig werden.	10.9 (0.7)	21.2 (0.8)	38.0 (1.2)	30.0 (1.2)
Kognitive Aktivierung	Immer oder fast immer % (SE)	Häufig % (SE)	Manchmal % (SE)	Selten oder nie % (SE)
Die Lehrerin/der Lehrer gibt uns Aufgaben, für die es keinen sofort ersichtlichen Lösungsweg gibt.	12.4 (0.5)	32.3 (1.1)	41.2 (1.0)	14.0 (0.7)
Unterstützung durch die Lehrperson	In jeder Stunde % (SE)	In den meisten Stunden % (SE)	In einigen Stunden % (SE)	Nie oder fast nie % (SE)
Unsere Lehrerin/unsere Lehrperson erklärt etwas so lange, bis wir es verstehen.	34.4 (1.2)	28.6 (1.0)	22.8 (0.9)	14.2 (0.8)

Kognitive Aktivierung

Aufgrund der anhaltenden Bemühungen um eine Verbesserung des Mathematikunterrichts seit TIMSS und PISA 2000 würde man erwarten, dass im Mathematikunterricht in Deutschland mittlerweile verstärkt kognitiv aktivierende Aufgaben eingesetzt werden. Die Angaben der Schülerinnen und Schüler sind in dieser Hinsicht nicht einheitlich. Mehr als die Hälfte (55.2 Prozent) gibt an, dass Aufgaben, bei denen es keinen sofort ersichtlichen Lösungsweg gibt, „manchmal“ beziehungsweise „selten oder nie“ im Unterricht vorkommen. Dagegen wählen nur 12.4 Prozent der Befragten die Kategorie „immer oder fast immer“ und knapp ein Drittel der Schülerinnen und Schüler kreuzt die Kategorie „häufig“ an. Es zeichnet sich hier also eine gemischte Situation ab. Die Bearbeitung von Aufgaben, deren Lösungsweg unmittelbar ersichtlich ist, hat durchaus ihre Berechtigung, wenn Prozeduren geübt werden sollen. Sie sind allerdings weniger geeignet, um Schülerinnen und Schüler zum Nachdenken über mathematische Inhalte anzuregen. Kognitiv aktivierende Aufgaben wären sicherlich eine bessere Wahl, scheinen aber aus Schülersicht eher seltener im Mathematikunterricht in Deutschland vorzukommen.

Unterstützung durch die Lehrperson

Die individuell wahrgenommene Unterstützung durch die Lehrperson stellt eine zentrale Variable für effektiven Unterricht dar. Betrachtet man vor diesem Hintergrund die Einschätzungen der befragten Schülerinnen und Schüler, zeichnet sich auch hier ein durchwachsendes Bild ab. Zwar berichtet ungefähr ein Drittel der Jugendlichen, dass ihr Leh-

rer oder ihre Lehrerin in jeder Stunde den Unterrichtsstoff so lange erklärt, bis ihn die Klasse verstanden hat, und für weitere 28.6 Prozent ist dies „in den meisten Stunden“ der Fall. Hingegen sagen 22.8 Prozent, dass dies nur „in einigen Stunden“ vorkommt, und 14.2 Prozent erleben dies „nie oder fast nie“. Nimmt man die letzten beiden Antwortkategorien zusammen, so fühlt sich offenbar mehr als ein Drittel der Schülerinnen und Schüler nur in geringem Maße in ihrem Lernprozess unterstützt. Nach den Angaben der Schülerinnen und Schüler erfährt damit ein nennenswerter Anteil keine ausreichende Unterstützung durch die Lehrpersonen.

Lehrersteuerung

Unter dem Aspekt der Lehrersteuerung sind solche Aktivitäten und Handlungsweisen zusammengefasst, bei denen die aktive Rolle der Lehrperson in der Begleitung des Lernens hervorgehoben wird. Dazu zählt unter anderem, ob Lehrkräfte das Verständnis ihrer Schülerinnen und Schüler für die mathematischen Inhalte durch Fragen prüfen. Betrachtet man die in Tabelle 4.3 dargestellten Ergebnisse, scheint eine in diesem Sinne lehrergesteuerte Ausrichtung des Mathematikunterrichts ein wesentliches Merkmal in Deutschland zu sein. Fast 70 Prozent der Jugendlichen geben an, dass ihre Lehrerin oder ihr Lehrer „in jeder Stunde“ oder „in den meisten Stunden“ Fragen zum Verständnis des Unterrichtsstoffs stellt. Für 23.0 Prozent kommt dies „in einigen Stunden“ und für nur 7.4 Prozent „nie oder fast nie“ vor. Interessant wäre hier, ob Verständnisprüfung und Verständnissicherung in gleicher Weise zum Unterrichtsalltag gehören. Zweifel daran legt die Skala Rückmeldung nahe.

Rückmeldung

Auf die Frage, wie häufig Schülerinnen und Schüler Rückmeldungen über ihre Stärken und Schwächen erhalten, gibt etwas mehr als ein Drittel an, dass sie „nie oder fast nie“ solche Rückmeldungen erhalten, und 41.0 Prozent bekommen sie nur „in einigen Stunden“. Für 18.7 Prozent kommt dies „in den meisten Stunden“, für nur 5.9 Prozent tatsächlich „in jeder Stunde“ vor. Demnach vergewissert sich die Lehrperson aus Sicht der Schülerinnen und Schüler zwar relativ häufig durch Fragen, dass der Unterrichtsstoff verstanden wurde, dies erfolgt aber eher selten in Form individueller Rückmeldungen.

Schülerorientierung

In Bezug auf die Schülerorientierung ergibt sich ein nicht unproblematisches Bild: Ein schülerorientierter Mathematikunterricht, der hier beispielhaft über seine Ausrichtung auf die Bedürfnisse einzelner Schülerinnen und Schüler im Sinne einer inneren Differenzierung gesehen wird, scheint an Schulen in Deutschland wenig verbreitet zu sein. Mehr als die Hälfte der Jugendlichen gibt an, dass Mitschülerinnen und Mitschüler mit unterschiedlichem Lerntempo „nie oder fast nie“ unterschiedliche Aufgabenstellungen

Tabelle 4.3: Prozentuale Antworthäufigkeiten für Beispielitems der Skalen zu den Unterrichtspraktiken

Unterrichtsmerkmal mit Beispielitem	Prozentuale Zustimmung			
	In jeder Stunde	In den meisten Stunden	In einigen Stunden	Nie oder fast nie
Lehrersteuerung	% (SE)	% (SE)	% (SE)	% (SE)
Die Lehrerin/der Lehrer stellt uns Fragen, um zu überprüfen, ob wir den Unterrichtsstoff verstanden haben.	30.3 (1.1)	39.4 (0.9)	23.0 (0.9)	7.4 (0.5)
Rückmeldung	% (SE)	% (SE)	% (SE)	% (SE)
Die Lehrerin/der Lehrer gibt mir eine Rückmeldung über meine Stärken und Schwächen in Mathematik.	5.9 (0.4)	18.7 (0.9)	41.0 (0.9)	34.5 (0.9)
Schülerorientierung	% (SE)	% (SE)	% (SE)	% (SE)
Die Lehrerin/der Lehrer gibt Mitschülerinnen und Mitschülern, die Lernschwierigkeiten haben, und/oder Mitschülerinnen und Mitschülern, die schneller vorankommen, unterschiedliche Aufgaben.	9.0 (0.5)	15.6 (0.7)	21.2 (0.9)	54.2 (1.1)

bekommen. Ein weiteres Fünftel berichtet, dass dies nur „in einigen Stunden“ der Fall ist. Lediglich 15.6 Prozent der Jugendlichen geben an, dass dies „in den meisten Stunden“ vorkommt, knapp 10 Prozent erleben es „in jeder Stunde“. Vor dem Hintergrund der häufig geforderten individuellen Förderung und auch angesichts der Bemühungen um Inklusion von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf (vgl. Pant et al., 2013) können diese Angaben noch nicht zufriedenstellen. Hier gibt es offensichtlich erhebliches Potential für eine Weiterentwicklung des Mathematikunterrichts.

Mathematische Begriffe und Aufgabenstellungen

Bei der Frage nach der Vertrautheit mit mathematischen Begriffen hängen die Antworthäufigkeiten davon ab, ob die jeweiligen Inhalte bereits systematisch im Unterricht behandelt worden sind. Beispielsweise werden Wurzeln üblicherweise zu Beginn von Klasse 9 behandelt, sodass sie einem großen Teil der Fünfzehnjährigen zum Testzeitpunkt vertraut sein sollten. Tatsächlich geben auch 79.9 Prozent an, diesen Begriff zu kennen. Hingegen ist der Umgang mit komplexen Zahlen kaum mehr ein Inhalt schulischer Curricula, was sich auch im Antwortverhalten der Befragten widerspiegelt. 33.4 Prozent der Fünfzehnjährigen kreuzen beim Begriff „komplexe Zahl“ an, noch nie davon gehört zu haben, und nur 12.5 Prozent geben „Kenne es und verstehe den Begriff“ an. Das in Tabelle 4.4 dargestellte Beispiel „Quadratfunktion“ bewegt sich

dazwischen: 44.4 Prozent der Jugendlichen geben an, dass sie den Begriff kennen und verstehen, 10.3 Prozent haben „noch nie davon gehört“.

Aufgaben ohne Anwendungsbezug spielen im Mathematikunterricht in Deutschland weiterhin eine deutlich größere Rolle als Anwendungsaufgaben. 67.7 Prozent der Jugendlichen geben an, dass Aufgaben, bei denen eher das technische Anwenden von Algorithmen wie zum Beispiel das Lösen einer Gleichung im Vordergrund steht, „häufig“ im Unterricht vorkommen. Dagegen wählen nur 22.6 Prozent diese Einordnung für die Aufgabe zum Berechnen eines Kaufpreises mit Mehrwertsteuer. Für weitere 20.9 Prozent kommt die Aufgabe, „eine Gleichung wie $3x + 5 = 17$ lösen“ „manchmal“ und für nur 11.4 Prozent „selten“ bis „nie“ vor. Einfache Anwendungsaufgaben wie die Berechnung des Kaufpreises mit Mehrwertsteuer begegnen immerhin 39.2 Prozent der Jugendlichen „manchmal“, einem Viertel aber nur „selten“, und mehr als jedem Zehnten „nie“. Dieses Bild wird auch durch die Angaben zu dem in Abbildung 4.1 gezeigten Einzelitem bestätigt: Lediglich 18.1 Prozent der Schülerinnen und Schüler geben an, dass solche außermathematischen Anwendungsaufgaben „häufig“ vorkommen, immerhin 47.2 Prozent kreuzen „manchmal“ an. Mehr als ein Drittel gibt aber an, solchen Aufgaben „selten“ oder „nie“ begegnet zu sein. Letzteres ist bedauerlich, weil es sich gerade bei den beiden in diesem Einzelitem dargestellten Beispielen um aus mathematikdidaktischer Sicht gute Aufgaben handelt, bei denen die Anwendung von Mathematik und mathematisches Argumentieren im Vordergrund stehen.

Tabelle 4.4: Prozentuale Antworthäufigkeiten für Beispielitems der Skalen zu den mathematischen Begriffen und Aufgabenstellungen

Unterrichtsmerkmal mit Beispielitems	Prozentuale Zustimmung				
	Noch nie davon gehört	Ein- oder zweimal davon gehört	Einige Male davon gehört	Häufig davon gehört	Kenne es und verstehe den Begriff
Vertrautheit mit mathematischen Begriffen	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)
Quadratfunktion	10.3 (0.6)	11.9 (0.7)	14.5 (0.7)	18.9 (0.8)	44.4 (1.2)
Häufigkeit innermathematischer Aufgaben	Häufig % (SE)	Manchmal % (SE)	Selten % (SE)	Nie % (SE)	
Eine Gleichung wie $3x + 5 = 17$ lösen.	67.7 (1.1)	20.9 (0.9)	7.0 (0.5)	4.4 (0.5)	
Häufigkeit einfacher Anwendungsaufgaben	Häufig % (SE)	Manchmal % (SE)	Selten % (SE)	Nie % (SE)	
Ausrechnen, um wie viel teurer ein Computer wird, wenn man die Mehrwertsteuer darauf schlägt.	22.6 (0.9)	39.2 (1.0)	25.0 (0.9)	13.2 (0.7)	
Häufigkeit außermathematischer Anwendungsaufgaben	Häufig % (SE)	Manchmal % (SE)	Selten % (SE)	Nie % (SE)	
siehe Abbildung 4.1	18.1 (0.8)	47.2 (1.0)	28.2 (0.8)	6.4 (0.5)	

Zusammenfassend wird deutlich, dass der Mathematikunterricht in Deutschland von den Schülerinnen und Schülern als weitgehend störungsfrei und gut organisiert wahrgenommen wird. Die Lehrpersonen kontrollieren dabei den Lernfortschritt der Schülerinnen und Schüler häufig. Maßnahmen zur individuellen Unterstützung und Lernbegleitung werden allerdings seltener angeboten, der Unterricht ist wenig schülerorientiert. Noch immer scheinen Anwendungsaufgaben nicht so häufig im Unterricht bearbeitet zu werden wie technische Aufgaben oder Routineprozeduren. Folgt man den Einschätzungen der Schülerinnen und Schüler, dann werden grundlegende Begriffe in einem erwartungsgemäßen Umfang vermittelt.

4.4.2 Mathematikunterricht in Deutschland im internationalen Vergleich

Nachdem nun auf der Basis einzelner, für die Skalen typischer Items ein Blick auf den Mathematikunterricht in Deutschland geworfen wurde, soll in diesem Abschnitt eine Einordnung in den internationalen Zusammenhang vorgenommen werden. Gibt es Besonderheiten der Unterrichtswahrnehmung von Schülerinnen und Schülern in Deutschland im Vergleich zu anderen OECD-Staaten? Um einen solchen Vergleich vorzunehmen, werden die Angaben aus den Befragungen der Jugendlichen ins Verhältnis zum OECD-Durchschnitt gesetzt. Dieser wurde auf null standardisiert, mit einer Standardabweichung von eins (z-Standardisierung). Dabei ist zu beachten, dass der OECD-Durchschnitt vorsichtig zu interpretieren ist, da hier Kennwerte von Staaten aus unterschiedlichen Kulturräumen zusammengefasst sind. Aus diesem Grund erfolgt eine genauere Betrachtung nur anhand ausgewählter europäischer Staaten, die einem ähnlichen Kulturraum wie Deutschland angehören. Von besonderem Interesse sind dabei Staaten, in denen Schülerinnen und Schüler im internationalen Vergleich überdurchschnittliche mathematische Kompetenzwerte erzielen und die sich zusätzlich durch geringe Anteile Jugendlicher auf oder unter der Kompetenzstufe I sowie durch gleichzeitig hohe Anteile Jugendlicher auf den Kompetenzstufen V und VI auszeichnen. Zu diesen Staaten gehören die Niederlande, die Schweiz und Finnland (vgl. Kapitel 2). Anders als etwa bei ostasiatischen Staaten kann davon ausgegangen werden, dass der Vergleich in diesen Fällen kaum durch kulturspezifische Antworttendenzen verzerrt wird (Johnson, Shavitt & Holbrook, 2011). Die Abbildungen 4.2 bis 4.4 zeigen die Ausprägungen der Unterrichtsmerkmale für Deutschland und die ausgewählten Vergleichsstaaten. Positive Werte bedeuten eine im OECD-Vergleich überdurchschnittliche Ausprägung eines Merkmals, negative Werte eine unterdurchschnittliche Ausprägung. Zu beachten ist, dass keine Aussagen über die absolute Höhe der Merkmalsausprägungen getroffen werden können, sondern nur Vergleiche der Länder untereinander möglich sind.

Vergleicht man zunächst die *Basisdimensionen der Unterrichtsqualität* (Disziplin, kognitive Aktivierung und Unterstützung durch die Lehrperson) mit dem OECD-Durchschnitt, so liegt Deutschland bei der Skala Unterstützung durch die Lehrperson

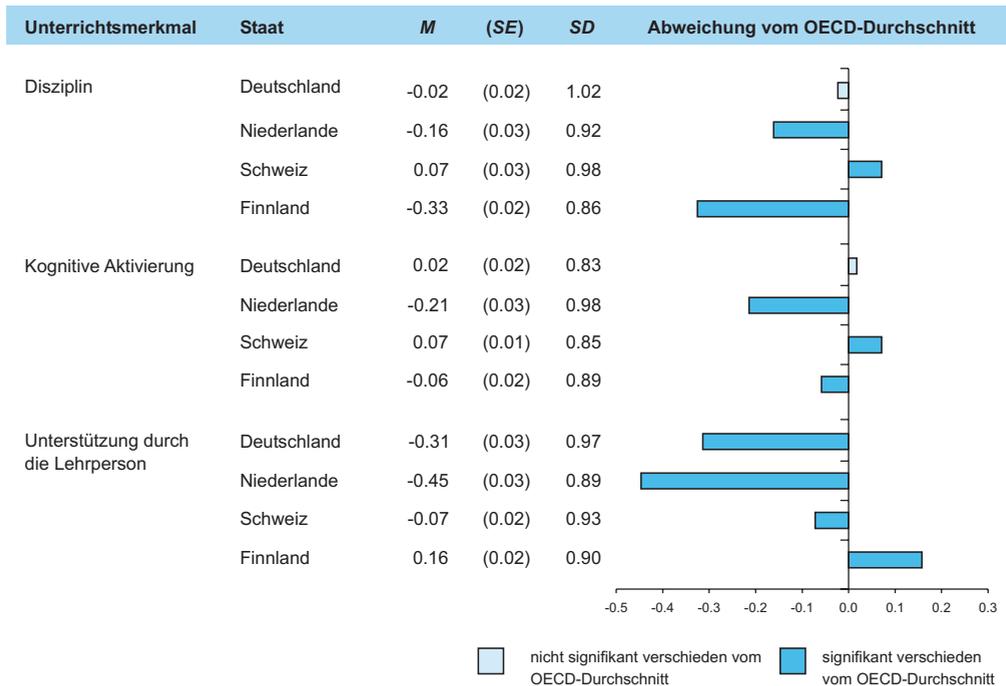


Abbildung 4.2: Basisdimensionen der Unterrichtsqualität im internationalen Vergleich

signifikant unterhalb des OECD-Durchschnitts (vgl. Abbildung 4.2). Die fünfzehnjährigen Jugendlichen fühlen sich in deutlich geringerem Maße von ihren Lehrerinnen und Lehrern unterstützt als Gleichaltrige in anderen OECD-Ländern. Noch niedriger fällt diese Wahrnehmung in den Niederlanden aus. Die Schülerinnen und Schüler in der Schweiz berichten ebenfalls eine signifikant geringere wahrgenommene Unterstützung durch ihre Lehrkräfte als im Mittel der OECD-Staaten, jedoch ist die Abweichung deutlich kleiner als in Deutschland und den Niederlanden. Nur die Schülerinnen und Schüler in Finnland nehmen in ihrem Mathematikunterricht eine überdurchschnittliche Unterstützung durch ihre Lehrerinnen und Lehrer wahr. Dafür berichten sie, ebenso wie die Jugendlichen in den Niederlanden, eine signifikant unterdurchschnittliche Disziplin und geringe kognitive Aktivierung in ihrem Mathematikunterricht. Hier zeichnet sich die Schweiz durch überdurchschnittliche Disziplin und häufige kognitive Aktivierung aus. Für Deutschland liegen die wahrgenommene Disziplin und kognitive Aktivierung im Mittel der OECD-Staaten. Da die Disziplin im Unterricht bereits bei PISA 2003 mit den gleichen Items erfasst wurde, kann hier die Veränderung von 2003 zu 2012 betrachtet werden. Es zeigt sich für Deutschland eine statistisch bedeutsame Abnahme in der wahrgenommenen Disziplin (vgl. OECD, 2013). Die Schülerinnen und Schüler scheinen sich im Mathematikunterricht etwas stärker gestört zu fühlen, als es noch 2003 der Fall war. In der Schweiz und in den Niederlanden hat sich die wahrgenommene Disziplin

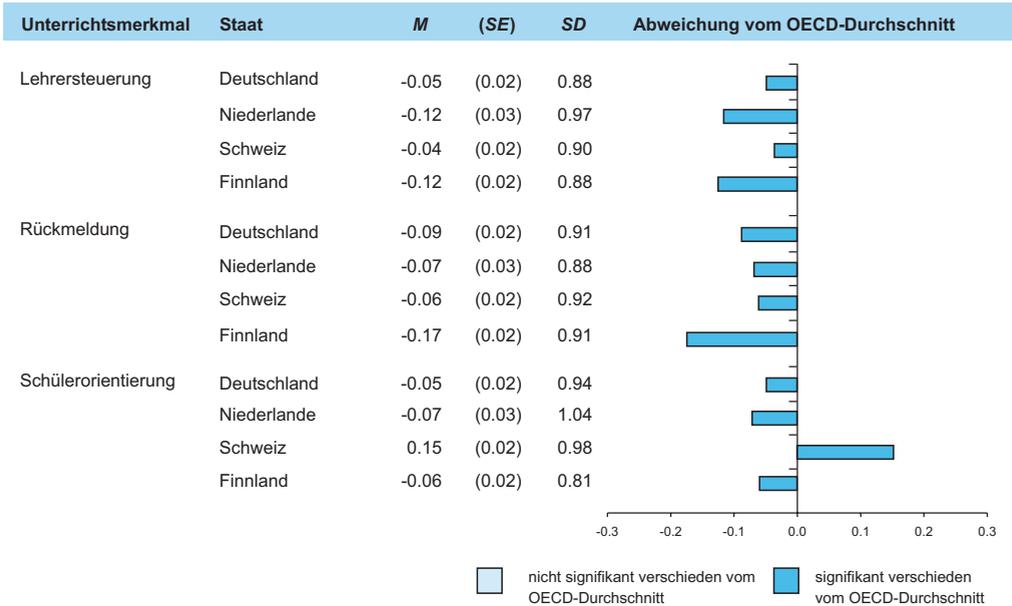


Abbildung 4.3: Wahrgenommene Unterrichtspraktiken im internationalen Vergleich

dagegen verbessert. Auch in Finnland zeigte sich eine leichte Verbesserung, die jedoch nicht statistisch bedeutsam ist.

Bei den wahrgenommenen *Unterrichtspraktiken* zeigt sich für Deutschland, dass alle drei Skalenwerte statistisch bedeutsam unterhalb des OECD-Durchschnitts liegen (vgl. Abbildung 4.3). Der Mathematikunterricht wird in Deutschland weniger lehrergesteuert und weniger schülerorientiert wahrgenommen als im Durchschnitt der OECD-Staaten. Auch individuelle Rückmeldungen kommen im Vergleich seltener vor. Ein ähnliches Bild zeigt sich für die europäischen Vergleichsstaaten. Auch hier weichen alle drei Unterrichtspraktiken signifikant vom OECD-Durchschnitt ab. So liegen in den Niederlanden und Finnland die Werte für alle drei Praktiken ebenfalls unterhalb des OECD-Durchschnitts. Nur die Schweiz stellt bei der Schülerorientierung eine Ausnahme dar: Dort wird der Unterricht statistisch bedeutsam häufiger als schülerorientiert wahrgenommen und scheint vergleichsweise größere Freiheitsgrade für Selbststeuerung und Partizipation zu gewährleisten. Hierin wird möglicherweise die in der Schweiz stärker verbreitete Praxis sogenannter erweiterter Lernformen wie zum Beispiel Wochenplan- und Freiarbeit sichtbar (Reusser, Pauli & Waldis, 2010).

Im Vergleich zu Deutschland berichten Schülerinnen und Schüler aus Finnland, dass ihr Mathematikunterricht seltener lehrergesteuert ist, aber auch, dass sie weniger individuelle Rückmeldungen erhalten. Auch diese Unterschiede sind statistisch bedeutsam. Der im OECD-Vergleich niedrige Wert für Rückmeldungen und Schülerorientierung in Finnland scheint bisherigen Erkenntnissen zum finnischen Unterricht zu widerspre-

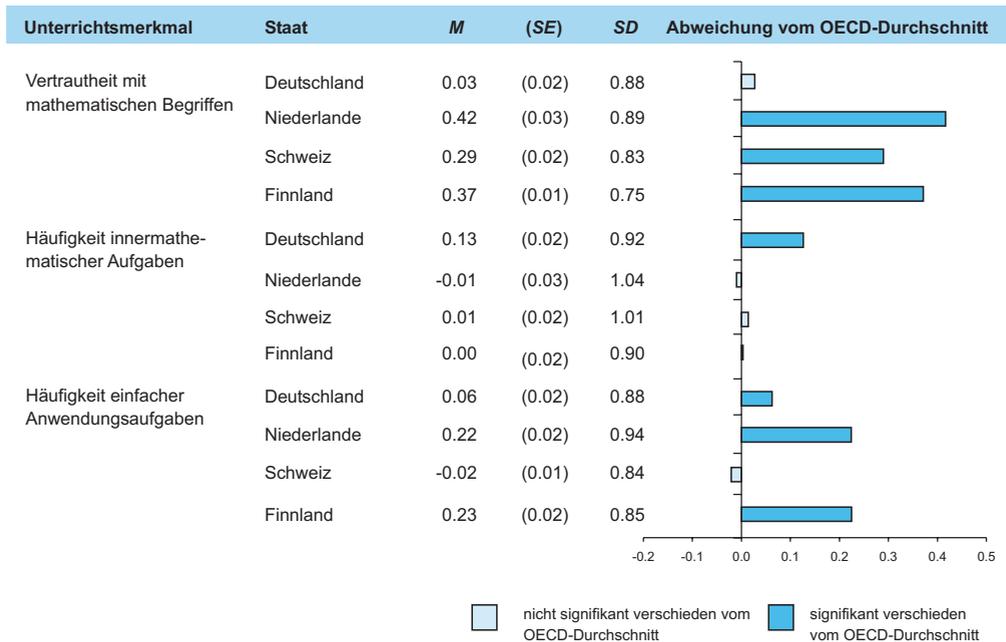


Abbildung 4.4: Mathematische Begriffe und Aufgabenstellungen im internationalen Vergleich

chen. Es könnte sein, dass individuelle Förderung dort mit anderen Mitteln als den hier beschriebenen Praktiken realisiert wird. Allerdings berichteten Beobachter finnischer Schulen oftmals, dass zwar ein sehr anerkennendes Schulklima herrscht, die Unterrichtspraktiken aber eher traditionell wirken (vgl. etwa von Freymann, 2003).

Betrachtet man die Antworten zur *Vertrautheit mit mathematischen Begriffen* in Abbildung 4.4, so sind die Werte für Deutschland im OECD-Vergleich eher unauffällig. Hinsichtlich der Häufigkeit bestimmter Aufgabentypen kommen nach den Angaben der Jugendlichen sowohl eher technische, innermathematische Aufgaben als auch einfache Anwendungsaufgaben signifikant häufiger vor als im Mittel der OECD-Staaten. Allerdings ist nach Angaben der Jugendlichen in den Niederlanden, der Schweiz und Finnland deren Vertrautheit mit mathematischen Begriffen im OECD-Vergleich und im Vergleich zu Deutschland signifikant stärker ausgeprägt.

Vergleicht man die wahrgenommene Häufigkeit von außermathematischen Anwendungsaufgaben im Unterricht (Tabelle 4.5), so zeigt sich, dass diese in Deutschland (18.1 Prozent geben die Antwortkategorie „häufig“ an) im Vergleich zum OECD-Durchschnitt (21.2 Prozent) etwas geringer ist. Dies trifft auch für Finnland (12.4 Prozent) und die Schweiz (14.8 Prozent) zu, nur in den Niederlanden scheint diese Art von Aufgaben häufiger vorzukommen (36.4 Prozent). Letzteres könnte in der starken Tradition von „Realistic Mathematics“ begründet sein – einem in den Niederlanden weit verbreiteten Reformansatz mit Betonung eines anwendungsbezogenen Mathematikunterrichts

Tabelle 4.5: Prozentuale Antworthäufigkeiten für die außermathematische Anwendungsaufgabe

Staat	Antwortkategorie							
	Häufig		Manchmal		Selten		Nie	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)
Deutschland	18.1	(0.8)	47.2	(1.0)	28.2	(0.8)	6.4	(0.5)
Niederlande	36.4	(1.3)	48.6	(1.2)	12.1	(0.6)	2.9	(0.3)
Schweiz	14.8	(0.7)	45.7	(0.8)	31.4	(0.8)	8.1	(0.4)
Finnland	12.4	(0.6)	47.7	(0.8)	31.3	(0.7)	8.5	(0.4)
OECD-Durchschnitt	21.2	(0.1)	45.2	(0.2)	25.8	(0.1)	7.8	(0.1)

Anmerkung: Statistisch signifikante Unterschiede vom OECD-Durchschnitt sind fett hervorgehoben.

(z. B. Gravemeijer, 1994). In allen Ländern und auch im OECD-Durchschnitt geben die meisten Fünfzehnjährigen an, dass diese Aufgaben „manchmal“ im Mathematikunterricht vorkommen. In den Niederlanden berichten nur 2.9 Prozent, dass diese Art von Aufgaben „nie“ im Unterricht vorkommt. Deutschland liegt mit 6.4 Prozent geringfügig unterhalb des OECD-Durchschnitts von 7.8 Prozent.

Exkurs: Wahrnehmung des Unterrichts am Gymnasium

Die relativ große Homogenität des Gymnasiums in Deutschland gestattet eine genauere Betrachtung des Unterrichts in dieser Schulart. Im Vergleich zum nationalen Durchschnitt wird die wahrgenommene Unterstützung durch die Mathematiklehrkraft als noch geringer angesehen ($M = -0.46$), der Unterricht wird seltener als lehrergesteuert ($M = -0.13$) und schülerorientiert ($M = -0.28$) eingeschätzt, individuelle Rückmeldungen kommen weniger häufig vor ($M = -0.30$). Der Unterricht wird jedoch als störungsärmer empfunden ($M = .08$). Obwohl man gerade am Gymnasium ein relativ hohes Niveau an kognitiver Aktivierung erwarten würde, wird dies von den Schülerinnen und Schülern nicht so wahrgenommen ($M = -.01$). Der inhaltliche Schwerpunkt liegt auf innermathematischen ($M = .44$) und nicht auf Anwendungsaufgaben ($M = .05$). Dabei ist die Vertrautheit mit mathematischen Begriffen im Vergleich zum gesamtdeutschen Profil deutlich höher ausgeprägt ($M = .40$). Außermathematische Anwendungsaufgaben, die besonders wünschenswert für den Mathematikunterricht wären, kommen für 14.6 Prozent der Gymnasiasten „häufig“ und für 46.7 Prozent „manchmal“ vor. 33.5 Prozent geben an, dass diese Art von Aufgaben „selten“ im Unterricht vorkommt und 5.2 Prozent sagen, dass sie solchen Aufgaben „nie“ begegnen.

4.4.3 Zwischenfazit

Setzt man die Ergebnisse des internationalen Vergleichs in den Zusammenhang mit der Beschreibung der Einzelitems, die charakteristisch für die jeweilige Skala sind (vgl. 4.4.1), so zeichnet sich für Deutschland das folgende Bild ab: Die wahrgenommene Disziplin hat sich in Deutschland zwar seit PISA 2003 etwas verschlechtert, liegt aber nach wie vor im OECD-Durchschnitt. Die Betrachtung des charakteristischen Items zeigte, dass etwa zwei Drittel der Jugendlichen ihren Unterricht als eher störungsarm bezeichnen würden, ein Drittel berichtet über Verbesserungspotential. Auch die kognitive Aktivierung ist im internationalen Vergleich durchschnittlich ausgeprägt, wohingegen mehr als die Hälfte der Schülerinnen und Schüler angibt, dass sie nur „manchmal“ beziehungsweise „nie oder fast nie“ Aufgaben bearbeiten, für die es nicht sofort einen ersichtlichen Lösungsweg gibt. Bei der Unterstützung durch die Lehrperson zeichnet sich im internationalen Vergleich ein durchaus problematisches Bild ab: Sie ist in Deutschland deutlich unterdurchschnittlich ausgeprägt. Die Betrachtung des charakteristischen Items zeigt allerdings, dass sich zwei Drittel der Jugendlichen im Großteil der Mathematikstunden unterstützt fühlen. Dennoch bleibt ein Drittel übrig, das sich selten oder nie unterstützt fühlt. Bei den Unterrichtspraktiken, die im OECD-Vergleich alle drei unterdurchschnittlich ausgeprägt sind, zeichnen die charakteristischen Items ein noch deutlicheres Bild. Geben noch knapp zwei Drittel der Jugendlichen an, dass die Lehrperson Fragen stellt, um zu überprüfen, ob sie den Stoff des Unterrichts verstanden haben, berichten drei Viertel der Schülerinnen und Schüler, dass sie nur „in einigen Stunden“ oder „nie oder fast nie“ im Unterricht individuell gefördert werden oder sich der Unterricht an ihren individuellen Bedürfnissen ausrichtet. In Bezug auf die Vertrautheit mit mathematischen Begriffen und Aufgabentypen stimmen die Ergebnisse des internationalen Vergleichs mit denen der Analyse der Einzelitems überein und weisen darauf hin, dass eher innermathematische als Anwendungsaufgaben im Mathematikunterricht überwiegen.

4.4.4 Muster des Mathematikunterrichts

Bisher wurden Merkmale des Mathematikunterrichts auf der Ebene einzelner Items und Skalen betrachtet, um Besonderheiten in Deutschland ausmachen zu können. Wenn aber der Mathematikunterricht als Ganzes charakterisiert werden soll, stößt dieser Ansatz an Grenzen. „Guter“ Unterricht kann nicht durch ein einzelnes Unterrichtsmerkmal beschrieben werden, sondern zeichnet sich durch das Zusammenspiel verschiedener Merkmale aus. Dieses Arrangement von Merkmalen ist besonders interessant, wenn Zusammenhänge mit Lernergebnissen untersucht werden sollen.

Im Folgenden werden Analysen berichtet, die es erlauben, komplexere Unterrichtsmuster zu betrachten und damit typische Lernumwelten von Schülerinnen und Schülern zu beschreiben. Dazu wurden fünf aussagekräftige Items aus den bisher berichte-

ten Unterrichtsmerkmalen ausgewählt, die in der Zusammenschau Indikatoren eines aus theoretischer Sicht „guten“ Mathematikunterrichts sein sollten. Die ausgewählten Items können Abbildung 4.5 entnommen werden. Ein wesentliches Kriterium bei der Auswahl der Items war, dass sie spezifische und voneinander unterscheidbare Muster von Mathematikunterricht abbilden können.

Zunächst wurde für den Mathematikunterricht in Deutschland analysiert, wie viele unterscheidbare Unterrichtsmuster von den Jugendlichen berichtet werden und durch welche Variablen sie sich differenzieren lassen. Im Anschluss wurden diese Analysen ebenfalls für die bereits genannten Staaten Niederlande, Schweiz und Finnland durchgeführt, um die Häufigkeit der dort auftretenden Unterrichtsmuster mit denen in Deutschland zu vergleichen. Aufgrund der Homogenität der gymnasialen Stichprobe wurden nachfolgend Zusammenhänge mit mathematischer Kompetenz, Freude und Interesse an Mathematik sowie mit mathematikbezogenen Selbstwirksamkeitserwartungen für Schülerinnen und Schüler am Gymnasium analysiert.

Methodische Vorbemerkung – die Latent-Class-Analyse

Die Analysen zu Unterrichtsmustern wurden mit dem R-Paket „poLCA“ (Linzer & Lewis, 2013) durchgeführt, um Latent-Class-Analysen (LCA – „latente Klassenanalysen“; Lazarsfeld, 1950; Formann, 1984) zu berechnen. Beim Testmodell der Latent-Class-Analyse werden die Antworten (Antwortvektoren) der einzelnen Schülerinnen und Schüler anhand der Ähnlichkeit des Antwortmusters in eine a priori festzulegende Anzahl von latenten Klassen (also in der Stichprobe identifizierbaren Gruppierungen) aufgeteilt. Ein einzelner Antwortvektor einer Schülerin oder eines Schülers wird dann mit einer bestimmten Wahrscheinlichkeit einer der latenten Klassen zugeordnet, und zwar so, dass sich die Antwortvektoren innerhalb einer latenten Klasse möglichst ähnlich sind und gleichzeitig zwischen den latenten Klassen möglichst unterschiedlich ausfallen (Lazarsfeld, 1950).

Aufgrund des Multi-Matrix-Designs mit drei unterschiedlichen Testheften für den internationalen Schülerfragebogen konnten bei den Berechnungen der LCA jeweils nur zwei Drittel der Teilstichprobe eines jeden Staates berücksichtigt werden. Zunächst wurde bestimmt, wie viele latente Klassen die empirische Datenmatrix am besten beschreiben. Dazu wurden jeweils Lösungen für die Annahme einer bis sechs Klassen berechnet, welche anschließend hinsichtlich ihrer relativen Modellpassung verglichen wurden. Als Kriterium für die Bestimmung der Anzahl der latenten Klassen (Auswahl des am besten passenden Modells) wurde das Bayessche Informationskriterium herangezogen (BIC; Schwarz, 1978). Um die latenten Klassen inhaltlich zu beschreiben, wurden jeweils die Mittelwerte der Antworthäufigkeiten zu den ausgewählten Fragen berechnet. Als Ergebnis resultierten für jede latente Klasse typische Profile der Schülerantworten (vgl. Abbildung 4.5). Zur Untersuchung der Frage, inwieweit die unterschiedlichen gefundenen typischen Antwortmuster (latente Klassen) zum Beispiel mit einer unterschiedlich ausgeprägten Kompetenz in Mathematik assoziiert sind, wurden

die Mittelwerte der Mathematikkompetenz für jede der latenten Klassen berechnet (vgl. Tabelle 4.6).

Es ergeben sich für Deutschland nach dem BIC drei und für die Vergleichsstaaten jeweils vier latente Klassen. Ein systematischer Vergleich der Vierklassenlösungen in den Vergleichsstaaten mit der Dreiklassenlösung in Deutschland ergibt, dass sich die zusätzliche vierte latente Klasse inhaltlich jeweils als Ausdifferenzierung einer der drei latenten Klassen für die deutsche Teilstichprobe interpretieren lässt. Weil für die Länder Finnland, Schweiz, und Niederlande die Vierklassenlösung jeweils eine (geringfügig) bessere Modellanpassung aufweist, diese zusätzliche latente Klasse aber keinen substantiellen Mehrwert in Bezug auf die inhaltliche Interpretation der Antwortmuster der Schülerinnen und Schüler liefert, wird aus Gründen der besseren Vergleichbarkeit hier jeweils die Dreiklassenlösung berichtet. Als Ergebnis einer Latent-Class-Analyse (bei vorgegebener Anzahl latenter Klassen) lässt sich für jede Schülerin und jeden Schüler einer Stichprobe die Wahrscheinlichkeit bestimmen, mit der sie oder er den jeweiligen latenten Klassen zugeordnet werden kann. Diese mittlere Klassenzuordnungswahrscheinlichkeit für die Dreiklassenlösung liegt für Deutschland bei $p_{\text{mean}} = .79$, für die Schweiz bei $p_{\text{mean}} = .78$, für Finnland bei $p_{\text{mean}} = .80$ und für die Niederlande bei $p_{\text{mean}} = .82$. Die mittlere Klassenzuordnungswahrscheinlichkeit kann als Maß für die Reliabilität des jeweiligen Latent-Class-Modells (Rost, 2004, S. 161; vgl. Kapitel 10) interpretiert werden. Insofern weist das Klassifikationsergebnis für die Dreiklassenlösung für alle untersuchten Teilstichproben beziehungsweise Staaten eine zufriedenstellende Reliabilität auf.

Als weiteres Ergebnis einer Latent-Class-Analyse lässt sich die Wahrscheinlichkeit bestimmen, mit der eine beliebige Schülerin oder ein beliebiger Schüler mit einem bestimmten Antwortmuster einer bestimmten latenten Klasse zugeordnet wird. Diese „Klassenwahrscheinlichkeiten“ können auch als relative Klassengrößen interpretiert werden, die sich insgesamt zu 1 addieren und sich daher in Prozentwerten ausdrücken lassen (vgl. Tabelle 4.6).

Unterrichtsmuster in Deutschland

Mit Hilfe der Latent-Class-Analyse konnten drei Unterrichtsmuster identifiziert werden, die in Abbildung 4.5 dargestellt sind. Das *erste Unterrichtsmuster* (Typ 1) ist dadurch gekennzeichnet, dass sich die Lehrperson in jeder oder fast jeder Stunde für den Lernfortschritt der Schülerinnen und Schüler interessiert und ihnen mitteilt, was von ihnen erwartet wird. Der Unterricht ist in allen oder den meisten Stunden störungsarm. Sehr häufig bearbeiten die Jugendlichen Aufgaben, für die es mehrere Lösungswege gibt und die sie zum Nachdenken anregen. Dieses positive Muster eines Mathematikunterrichts berichten 21.0 Prozent der Fünfzehnjährigen in Deutschland (Tabelle 4.6). Es verbindet individuelle Unterstützung und Zielorientierung mit häufiger kognitiver Aktivierung und wenig Störungen. Der Unterrichtstyp lässt sich abgekürzt als „unterstützend und aktivierend“ charakterisieren.

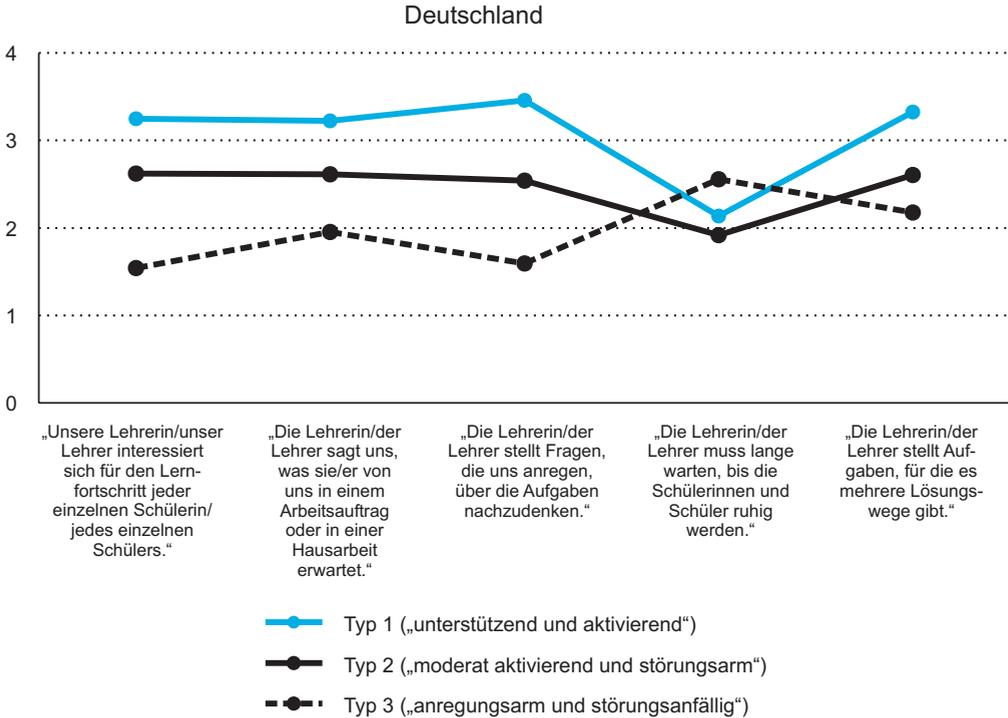


Abbildung 4.5: Profillinien der latenten Klassen für die Unterrichtswahrnehmung der Fünfzehnjährigen in Deutschland

Das *zweite Unterrichtsmuster* (Typ 2) zeichnet sich dadurch aus, dass sich die Lehrperson im Vergleich zu Typ 1 seltener (genauer: in den meisten bis einigen der Mathematikstunden) für den Lernfortschritt der Schülerinnen und Schüler interessiert. Ebenso häufig teilt sie nach Auskunft der Schülerinnen und Schüler mit, was sie von den Fünfzehnjährigen erwartet. Im Vergleich zu Typ 1 muss die Lehrperson noch etwas seltener darauf warten, dass alle Schülerinnen und Schüler ruhig sind. Es werden in etwa der Hälfte der Mathematikstunden Aufgaben mit mehreren Lösungsweisen bearbeitet beziehungsweise solche, die zum Nachdenken anregen. Unterstützung und kognitive Aktivierung sind auf einem passablen Niveau gegeben, die Disziplin ist offensichtlich sehr gut. Dieser Unterrichtstyp wird als „moderat aktivierend und störungsarm“ bezeichnet. Der Großteil der Jugendlichen in Deutschland, nämlich 53,1 Prozent, charakterisiert den Mathematikunterricht in entsprechender Weise (Tabelle 4.6).

Beim *dritten Unterrichtsmuster* (Typ 3) interessiert sich die Lehrperson aus Sicht der Schülerinnen und Schüler nie oder nur in einigen Unterrichtsstunden für deren Lernfortschritt. Sie erklärt auch nur in wenigen Stunden, was sie von ihnen erwartet. In etwa der Hälfte der Mathematikstunden muss zudem gewartet werden, bis alle Schülerinnen und Schüler ruhig sind. Gleichzeitig werden im Unterricht nur manchmal oder selten Aufgaben bearbeitet, die zum Nachdenken anregen und auch nur manchmal Aufgaben,

für die es mehrere Lösungswege gibt. Unterstützung und kognitive Aktivierung sind damit eher seltene Ereignisse, die Disziplin ist unterdurchschnittlich. Dieser Unterrichtstyp wird deshalb als „anregungsarm und störungsanfällig“ bezeichnet. Dieses auf den ersten Blick eher problematische Unterrichtsmuster wird von 26.0 Prozent der Jugendlichen in Deutschland berichtet (Tabelle 4.6).

Tabelle 4.6: Prozentuale Verteilung der drei Unterrichtsmuster in Deutschland und den europäischen Vergleichsstaaten

Staat	Typ 1 („unterstützend und aktivierend“)	Typ 2 („moderat aktivierend und störungsarm“)	Typ 3 („anregungsarm und störungsanfällig“)
Deutschland	21.0	53.1	26.0
Niederlande	35.7	47.8	16.5
Schweiz	24.9	54.5	20.5
Finnland	19.6	56.0	24.4

Anmerkung: In der Tabelle werden gerundete Werte angegeben. Es können demnach auf Rundung zurückzuführende vermeintliche Inkonsistenzen auftreten.

Unterrichtsmuster ausgewählter europäischer OECD-Vergleichsstaaten

In den Niederlanden, der Schweiz und Finnland findet man nach Einschätzungen der Schülerinnen und Schüler ähnliche Unterrichtsmuster wie in Deutschland. Tabelle 4.6 zeigt die jeweilige prozentuale Verteilung der Unterrichtsmuster. Es fällt auf, dass vor allem in den Niederlanden eher über einen unterstützenden und aktivierenden Mathematikunterricht berichtet wird und dass dort das problematische Unterrichtsmuster „anregungsarm und störungsanfällig“ vergleichsweise selten anzutreffen ist.

Zusammenhänge zwischen Unterrichtsmustern und mathematischer Kompetenz, Freude und Interesse an Mathematik sowie mathematikbezogenen Selbstwirksamkeitserwartungen am Gymnasium in Deutschland

Es ist anzunehmen, dass die beschriebenen Unterrichtsmuster in unterschiedlichem Maße geeignet sind, Schülerinnen und Schüler in der Ausführung ihrer Lernaktivitäten zu unterstützen. Es soll daher zunächst analysiert werden, wie diese Muster mit der gemessenen mathematischen Kompetenz zusammenhängen. Dabei beschränkt sich die Betrachtung auf das Gymnasium in Deutschland. Für andere Schularten können keine verlässlichen Aussagen getroffen werden, da die jeweilige Anzahl der teilgenommenen Schulen relativ klein ist und die Ergebnisse aufgrund der weniger homogen zusammengesetzten Schülerschaft schwieriger zu interpretieren wären. Weil Unterricht nicht nur kognitive Leistungen fördern, sondern auch Freude und Interesse an Mathematik sowie positive mathematikbezogene Selbstwirksamkeitserwartungen (vgl. Kapitel 3) vermitteln sollte, werden die bei PISA erhobenen motivational-affektiven Komponenten ebenfalls in die Analysen einbezogen. Ein Mathematikunterricht, der nur kognitive Leistungen

fördert, aber die Entwicklung von Interesse und Selbstwirksamkeitserwartungen unberücksichtigt lässt, wird dem Bildungsauftrag nicht gerecht und schadet der Nachwuchsförderung für Ausbildungsgänge und Berufskarrieren, die ein mathematisches Verständnis voraussetzen.

Betrachtet man für die drei Unterrichtsmuster die Mittelwerte der mathematischen Kompetenz in Tabelle 4.7, dann zeigt sich für die Gymnasien in Deutschland, dass diejenigen Schülerinnen und Schüler die höchsten Kompetenzwerte im PISA-Mathematiktest erzielen, die den Unterricht als „unterstützend und aktivierend“ (Typ 1) erleben. Sie erreichen einen mittleren mathematischen Kompetenzwert von 595 Punkten. Einen ähnlich hohen Wert erzielen auch Jugendliche, die den Unterricht als „moderat aktivierend und störungsarm“ (Typ 2) beschreiben (593 Punkte). Den im Vergleich zu Typ 1 und 2 signifikant geringsten Kompetenzwert (579 Punkte) erreichen Jugendliche, die den Unterricht als „anregungsarm und störungsanfällig“ wahrnehmen (Typ 3). Auch wenn die vorgenommenen Analysen keine kausalen Schlüsse zulassen, so sind die Ergebnisse doch mit der theoretischen Annahme vereinbar, dass ein Mathematikunterricht, der kognitiv anspruchsvoll ist und gleichzeitig in einer störungsarmen und unterstützenden Lernatmosphäre stattfindet, zu höheren Leistungen der Lernenden führt.

Für Freude und Interesse an Mathematik zeigt sich ein ähnliches Bild. Jugendliche, die den Mathematikunterricht als „unterstützend und aktivierend“ (Typ 1) wahrnehmen, berichten die höchste Ausprägung von Freude und Interesse an Mathematik. Signifikant geringer ist diese ausgeprägt, wenn die Fünfzehnjährigen den Unterricht als „anregungsarm und störungsanfällig“ (Typ 3) wahrnehmen. Auch die mathematikbezogenen Selbstwirksamkeitserwartungen sind am stärksten bei den Schülerinnen und Schülern ausgeprägt, die den Mathematikunterricht als „unterstützend und aktivierend“ (Typ 1) wahrnehmen und am schwächsten bei denen, die den Unterricht als „anregungsarm und störungsanfällig“ (Typ 3) erleben.

Die Befunde legen nahe, dass ein von den Jugendlichen als „unterstützend und aktivierend“ wahrgenommener Unterricht (Typ 1) nicht nur mit einer hohen mathematischen Kompetenz zusammenhängt, sondern auch mit der im Vergleich am höchsten ausgeprägten Freude an Mathematik und hohen Selbstwirksamkeitserwartungen einhergeht. Bedauerlich ist, dass dieses Unterrichtsmuster am Gymnasium nur von 15.2 Prozent der Jugendlichen berichtet wird. In einem als „moderat aktivierend und störungsarm“ (Typ 2) erlebten Mathematikunterricht ist die mathematische Kompetenz zwar ähnlich stark ausgeprägt, jedoch regt dieser Unterricht deutlich weniger die Freude für Mathematik an und hängt mit geringer ausgeprägten mathematikbezogenen Selbstwirksamkeitserwartungen zusammen. Dieses Unterrichtsmuster berichten etwas mehr als die Hälfte der Gymnasiastinnen und Gymnasiasten (55.3 Prozent). Zu Bedenken gibt der Befund, dass 29.5 Prozent der fünfzehnjährigen Gymnasiastinnen und Gymnasiasten einen Unterricht erleben, der als „anregungsarm und störungsanfällig“ wahrgenommen wird (Typ 3) und nicht nur mit der niedrigsten mathematischen Kompetenz, sondern auch mit der geringsten Freude an Mathematik sowie den niedrigsten Selbstwirksamkeitserwartungen einhergeht.

Tabelle 4.7: Zusammenhang von Unterrichtsmustern mit mathematischer Kompetenz, Freude und Interesse an Mathematik sowie mathematikbezogene Selbstwirksamkeitserwartungen am Gymnasium

	Mathematikkompetenz			Freude und Interesse			Selbstwirksamkeit		
	<i>M</i>	(<i>SE</i>)	<i>SD</i>	<i>M</i>	(<i>SE</i>)	<i>SD</i>	<i>M</i>	(<i>SE</i>)	<i>SD</i>
Typ 1 („unterstützend und aktivierend“)	595	(6.7)	69.7	0.10	(0.1)	1.1	0.85	(0.1)	0.8
Typ 2 („moderat aktivierend und störungsarm“)	593	(4.0)	68.3	-0.21	(0.1)	1.0	0.63	(0.1)	0.8
Typ 3 („anregungsarm und störungsanfällig“)	579	(4.1)	64.6	-0.57	(0.1)	1.0	0.54	(0.1)	0.9

4.5 Zusammenfassung und Diskussion

In diesem Kapitel wurde der Frage nachgegangen, wie sich der Mathematikunterricht in Deutschland aus der Sicht der fünfzehnjährigen Schülerinnen und Schüler charakterisieren lässt. Die aus dem Schülerfragebogen gewonnenen Daten erlauben eine Rekonstruktion des Unterrichts, bei der allerdings zu berücksichtigen ist, dass die Wahrnehmungen der Jugendlichen nicht für alle Merkmale gleichermaßen zuverlässig sind (vgl. Baumert et al., 2004) und dass sich die befragten Fünfzehnjährigen auf unterschiedliche Klassen, Schularten und Schulen verteilen. Trotz dieser Einschränkungen liefern die Analysen interessante Ergebnisse, die es zu diskutieren gilt.

Im internationalen Vergleich wichen die Skalenwerte für die meisten Unterrichtsmerkmale in Deutschland nur geringfügig vom OECD-Durchschnitt ab. Deutlich unter dem Durchschnitt lag allerdings – wie auch in den Niederlanden – das Ausmaß der von Schülerinnen und Schülern wahrgenommenen Unterstützung durch die Lehrperson. Vor allem auch im Hinblick auf motivational-affektive Ziele des Mathematikunterrichts wie etwa die Förderung der Freude und des Interesses der Schülerinnen und Schüler zeichnen sich hier Verbesserungsmöglichkeiten ab. Dass eine stärkere Unterstützung nicht zu Lasten der Lernergebnisse gehen muss, zeigt sich am Beispiel Finnlands, wo die Schülerinnen und Schüler ein hohes Maß an Unterstützung wahrnehmen und gleichzeitig – auch wenn es im Trend Einbußen gab (vgl. Kapitel 2) – ein hohes Kompetenzniveau erreichen. Wenn man die konkreten Antworthäufigkeiten der Fünfzehnjährigen betrachtet, zeichnen sich zusätzliche Verbesserungsmöglichkeiten bei der individuellen Förderung und der Ausrichtung des Unterrichts an den Bedürfnissen der Schülerinnen und Schüler ab.

Angesichts der in den vergangenen Jahren angestoßenen Maßnahmen zur Weiterentwicklung des Mathematikunterrichts (z. B. KMK-Bildungsstandards, SINUS; vgl. 4.2) war eine Erwartung, dass in Deutschland häufiger ein kognitiv aktivierender Mathematikunterricht stattfindet, in dem anspruchsvolle Aufgabenstellungen eingesetzt werden und in dem die Relevanz von Mathematik für Anwendungssituationen deutlich wird.

Aus der Betrachtung der Wahrnehmung der Fünfzehnjährigen wurden diese Erwartungen nicht vollständig erfüllt, wobei hier ganz besonders die bereits genannten Einschränkungen der Datenlage zu berücksichtigen sind. Vor allem die Tatsache, dass offenbar recht häufig Aufgaben bearbeitet werden, die durch eher technische Verfahren zu lösen sind und bei denen der Lösungsweg mehr oder minder offensichtlich ist, lässt Möglichkeiten zur weiteren Verbesserung des Mathematikunterrichts erkennen. Dabei soll hier nicht der Eindruck entstehen, dass mathematische Arbeitstechniken wie etwa das Lösen einfacher Gleichungen vernachlässigt werden sollten. Verständnisvolles Lernen setzt aber grundsätzlich voraus, dass sich Lernende aktiv mit der Thematik beschäftigen. Die vorgestellten Daten deuten darauf hin, dass sich zu viele Schülerinnen und Schüler im Mathematikunterricht verhältnismäßig wenig kognitiv aktiviert fühlen.

Wie einleitend erwähnt, wirken Unterrichtsmerkmale in komplexer Weise zusammen. Es wurde die Erwartung formuliert, dass sich ein guter Mathematikunterricht als kognitiv anregend und gleichzeitig störungsarm und schülerorientiert charakterisieren lässt. Mit Hilfe der Latent-Class-Analyse ist es gelungen zu zeigen, dass sich ein solcher Unterricht in der Wahrnehmung der Fünfzehnjährigen in Deutschland und auch in ausgewählten Vergleichsstaaten finden lässt. Bedauerlich ist, dass nur etwa ein Fünftel der Schülerinnen und Schüler in Deutschland ihren Unterricht so erlebt. Etwas mehr Schülerinnen und Schüler, nämlich etwa ein Viertel der Fünfzehnjährigen, nehmen den Unterricht hingegen als „störungsanfällig und anregungsarm“ wahr.

Darüber hinaus konnte für das Gymnasium erwartungsgemäß gezeigt werden, dass der als wünschenswert beschriebene unterstützende und aktivierende Mathematikunterricht nicht nur mit den vergleichsweise höchsten Kompetenzwerten, sondern gleichzeitig mit einem hohen Maß an Freude und Interesse sowie stark ausgeprägten Selbstwirksamkeitserwartungen einhergeht. Die Mehrheit der Schülerinnen und Schüler erlebt einen Mathematikunterricht, der verkürzt als „moderat aktivierend und störungsarm“ bezeichnet wurde. Diese Beschreibung sollte nicht zu negativ bewertet werden, denn die konkreten Angaben verweisen auf ein mittleres Niveau an kognitiver Aktivierung wie auch an wahrgenommener Unterstützung, und am Gymnasium ging dieses Muster mit verhältnismäßig hohen Kompetenzwerten einher. In weiteren Analysen wäre zu betrachten, ob die Zugehörigkeit einzelner Schülerinnen und Schüler zu den latenten Klassen durch weitere, hier nicht berücksichtigte Faktoren erklärt werden kann. Erst mit der Follow-up-Studie, die in Deutschland an PISA 2012 angebunden wurde (vgl. Kapitel 1), wird sich herausfinden lassen, welche Bedeutung verschiedene Merkmale des Unterrichts für die Leistungs- und Motivationsentwicklung der Jugendlichen haben, wobei auch Unterschiede zwischen Schulformen in den Blick genommen werden können.

Als Konsequenz aus den jetzt vorliegenden Befunden scheint eine intensivere und individuelle Unterstützung der Lernenden ein sinnvolles Ziel, um den Mathematikunterricht konsequent weiterzuentwickeln. Diese Unterstützung sollte einerseits dem Lernprozess gelten in dem Sinne, dass Schülerinnen und Schüler in kognitiver Hinsicht angeregt werden, beispielsweise durch die Verwendung interessanter Aufgabenstellungen. Gerade leistungsfähige Schülerinnen und Schüler können durch besonders anspruchsvolle Auf-

gabenstellungen noch stärker herausgefordert werden. Andererseits sollte Schülerinnen und Schülern aber auch stets das Gefühl vermittelt werden, dass sie sich bei Fragen jederzeit an ihren Lehrer oder ihre Lehrerin wenden können, dass ihre Lehrkraft sich für die individuellen Bedürfnisse jedes einzelnen Schülers und jeder einzelnen Schülerin interessiert, und dass sie ihnen Rückmeldungen zu ihren Stärken und Schwächen gibt. Wenn dies gelingt, können einerseits Jugendliche, die sich auf den unteren Kompetenzstufen befinden, besser gefördert werden, und andererseits kann der Anteil der Schülerinnen und Schüler auf den oberen Kompetenzstufen vergrößert werden. Neben einer Steigerung der Kompetenz ist die Wahrscheinlichkeit hoch, damit zugleich Freude an Mathematik und ein größeres Zutrauen in die eigenen Fähigkeiten zu wecken.

Literatur

- Adams, R. J., Lietz, P. & Berezner, A. (2013). On the use of rotated context questionnaires in conjunction with multilevel item response models. *Large-scale Assessments in Education*, 1 (1). Zugriff am 28.10.2013. Verfügbar unter: <http://link.springer.com/article/10.1186%2F2196-0739-1-5>
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W. & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 314–354). Münster: Waxmann.
- Blum, W. (2001): Was folgt aus TIMSS für Mathematikunterricht und Mathematiklehrer- und -pädagogenausbildung? In E. Klieme & J. Baumert (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 75–83). Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Cohen, D. K., Raudenbush, S. W. & Loewenberg Ball, D. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25, 119–142.
- Cooper, H. & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Dalehefte, I. M. (2007). *Unterrichtsskripts – ein multikriterieller Ansatz*. Kiel: IPN / CAU Kiel.
- Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. *Zeitschrift für Pädagogik*, 41, 73–92.
- Floden, R. E. (2001). Research on effects of teaching: A continuing model for research on teaching. In V. Richardson (Hrsg.), *Handbook of research on teaching* (S. 3–16). Washington, DC: American Educational Research Association.
- Formann, A. K. (1984). *Die Latent-Class-Analyse: Einführung in Theorie und Anwendung*. Weinheim: Beltz.
- Furtak, E. M., Seidel, T., Iverson, H. & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, 82, 300–329.
- Gravemeijer, K. P. E. (1994). *Developing realistic mathematics education*. Utrecht, The Netherlands: Freudenthal Institute.
- Hattie, J. (2009). *Visible Learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

- Helmke, A. & Weinert, F. E. (1997). Unterrichtsqualität und Leistungsentwicklung: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 241–251). Weinheim: Beltz.
- Hugener, I., Pauli, C., Reusser, K., Lipowsky, F., Rakoczy, K. & Klieme, E. (2009). Teaching patterns and learning quality in Swiss and German mathematics lessons. *Learning and Instruction, 19*, 66–78.
- Johnson, T. P., Shavitt, S. & Holbrook, A. L. (2011). Survey response styles across cultures. In D. R. Matsumoto & F. J. R. d. van Vijver (Hrsg.), *Cross-cultural research methods in psychology* (Culture and psychology, S. 130–178). New York: Cambridge University Press.
- Klieme, E. & Rakoczy, K. (2008). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele et al. (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 334–359). Opladen: Leske + Budrich.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & Allolio-Näcke, L. (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–146). Münster: Waxmann.
- Klieme, E., Pauli, C. & Reusser, K. (2009). The Pythagoras Study: Investigating effects of teaching and learning in Swiss and German classrooms. In T. Janik & T. Seidel (Hrsg.), *The power of video studies in investigating teaching and learning in the classroom* (S. 137–160). Münster: Waxmann.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabekultur“ und Unterrichtsgestaltung. In E. Klieme & J. Baumert (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 43–57). Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Klieme, E. & Vieluf, S. (2013). Schulische Bildung im internationalen Vergleich. Ein Rahmenmodell für Kontextanalysen in PISA. *Zeitschrift für Pädagogik, 59*, 229–246.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik. (2003). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss (10. Jahrgangsstufe)*. Beschluss vom 4.12.2003 (Beschlüsse der Kultusministerkonferenz). Bonn.
- Kobarg, M., Prenzel, M., Seidel, T., Walker, M., McCrae, B., Cresswell, J. & Wittwer, J. (2011). *An international comparison of Science teaching and learning – Further results from PISA 2006*. Münster: Waxmann.
- Kunter, M., Baumert, J. & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction, 17*, 494–509.
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S. & Neubrand, M. (Hrsg.). (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kunter, M., Dubberke, T., Baumert, J., Blum, W., Brunner, M. & Jordan, A. et al. (2006). Mathematikunterricht in den PISA-Klassen 2004: Rahmenbedingungen, Formen und Lehr-Lernprozesse. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 277–306). Münster: Waxmann.

- Lazarsfeld, P. F. (1950). The logical and mathematical foundations of latent structure analysis. In S. A. Stouffler, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Hrsg.), *Studies in social psychology in World War II* (Bd. IV Measurement and Prediction, S. 362–412). Princeton.
- Linzer, D. A. & Lewis, J. (2013). *poLCA: Polytomous Variable Latent Class Analysis*. R-Paket Version 1.4. Zugriff am 25.09.2013. Verfügbar unter <http://dlinzer.github.com/poLCA>
- Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 73–102). Heidelberg: Springer.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E. & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, 19, 527–537.
- OECD (2013). *What Makes Schools Successful? Resources, Policies and Practices*. Paris: OECD.
- Oser, F. & Baeriswyl, F. J. (2001). Choreographies of teaching: Bridging instruction to learning. In V. Richardson (Hrsg.), *Handbook of research on teaching* (S. 1031–1065). Washington, D.C.: American Educational Research Association.
- Pant, H. A., Stanat, P., Pöhlmann, C., Hecht, M., Jansen, M., Kampa, N. et al. (2013). Der Blick in die Länder. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 159–248). Münster: Waxmann.
- Pauli, C. & Reusser, K. (2006). Von international vergleichenden Video Surveys zur video-basierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52, 774–798.
- Prenzel, M. (1995). Zum Lernen bewegen. Unterstützung von Lernmotivation durch Lehre. *Blick in die Wissenschaft*, 4, 58–66.
- Prenzel, M. & Lankes, E.-M. (2013). Was können Schülerinnen und Schüler über ihren Unterricht sagen? Ein Blick in die Schülerfragebogen von internationalen Vergleichsstudien. In N. McElvany & H. G. Holtappels (Hrsg.), *Empirische Bildungsforschung. Theorien, Methoden, Befunde und Perspektiven* (S. 93–107). Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M. et al. (Hrsg.). (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.
- Prenzel, M., Friedrich, A. & Stadler, M. (2009). *Von SINUS lernen – Wie Unterrichtsentwicklung gelingt*. Seelze-Velber: Kallmeyer.
- Rakoczy, K., Klieme, E., Drollinger-Vetter, B., Lipowsky, F., Pauli, C. & Reusser, K. (2007). Structure as quality feature in mathematics instruction. In M. Prenzel (Hrsg.), *Studies on the educational quality of schools. The final report on the DFG Priority Programme* (S. 101–120). Münster: Waxmann.
- Reiss, K. & Hammer, C. (2013). *Grundlagen der Mathematikdidaktik. Eine Einführung für den Unterricht in der Sekundarstufe*. Basel: Birkhäuser.
- Reusser, K., Pauli C. & Waldis, M. (2010). *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht*. Münster: Waxmann.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Berlin: Hans Huber.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6, 461–464.
- Seidel, T. (2009). Unterrichtsführung. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 135–148). Heidelberg: Springer.

- Seidel, T. (2011). Lehrerhandeln im Unterricht. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 605–629). Münster: Waxmann.
- Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77, 454–499.
- Seidel, T., Prenzel, M. & Rimmel, R. (2003). Gelegenheitsstrukturen beim Klassengespräch und ihre Bedeutung für die Lernmotivation – Videoanalysen in Kombination mit Schülerelbsteinschätzungen. *Unterrichtswissenschaft*, 31, 142–165.
- Seidel, T., Prenzel, M., Wittwer, J. & Schwindt, K. (2007). Unterricht in 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. 147–179). Münster: Waxmann.
- Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S. & Serrano, A. (1996). *The TIMSS Video-tape Classroom Study: Methods and preliminary findings*. Los Angeles, CA.
- Terhart, E. (2011). Hat John Hattie tatsächlich den Heiligen Gral der Schul- und Unterrichtsforschung gefunden? Eine Auseinandersetzung mit Visible Learning. In E. Keiner (Hrsg.), *Metamorphosen der Bildung. Historie – Empirie – Theorie. Festschrift für Heinz-Elmar Tenorth* (S. 277–292). Bad Heilbrunn: Klinkhardt.
- von Freyermann, T. (2003). Ein anderes Land, eine andere Schule. Zu den finnischen PISA-Ergebnissen. *Neue Sammlung*, 43, 179–201.
- Waldis, M., Grob, U., Pauli, C. & Reusser, K. (2010). Der Einfluss der Unterrichtsgestaltung auf Fachinteresse und Mathematikleistung. In K. Reusser, C. Pauli & M. Waldis (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S. 209–252). Münster: Waxmann.

5 Schulische Rahmenbedingungen der Kompetenzentwicklung

Christine Sälzer, Manfred Prenzel und Eckhard Klieme

Auch wenn Kinder und Jugendliche im Alltag gelegentlich auf mathematische Probleme treffen, entwickeln sie ihre mathematische Kompetenz vor allem in der Schule. In gewisser Weise hat die Schule das Vermittlungsmonopol für große Teile der mathematischen Kompetenz (Baumert & Stanat, 2010). Das gleiche gilt für naturwissenschaftliche Kompetenz. Und selbst die Lesekompetenz, die von häufigem Lesen in der Freizeit profitiert, wird dort, wo es um das systematische Lesen und kundige Interpretieren bestimmter Textsorten geht, im Wesentlichen in der Schule aufgebaut. In der Schule wiederum findet die Entwicklung fachlicher beziehungsweise domänenspezifischer Kompetenzen nach wie vor insbesondere im Unterricht statt, in Lerngruppen also und unterstützt durch Lehrende, die Lernorte gestalten, gezielt Lerngelegenheiten bereitstellen und Lernprozesse begleiten. Allerdings stellen Schulen häufig auch außerunterrichtliche Lernangebote bereit, die der interessierten Vertiefung oder systematischen Nacharbeit dienen. Ganztagsangebote an Schulen umfassen zusätzliche Möglichkeiten der Freizeitgestaltung und gemeinsamer Aktivitäten.

Jede einzelne Schule (hier die *Einzel­schule* genannt) umfasst damit mehrere, unterschiedlich gestaltete Lernorte; sie integriert soziale Verbände – traditionell Klassen – unter einem Dach; sie dient als sozialer wie kultureller Lebensraum, möglicherweise mit einem bestimmten Profil und Anspruch. Die Schule kann und muss zugleich als organisatorische Einheit verstanden werden, die etwa Lehrkräfte zusammenbringt, Ressourcen bereitstellt, Unterricht über Jahrgangsstufen und Klassen abstimmt, sich auf ein eigenes pädagogisches Programm verständigt, Verantwortlichkeiten definiert und Führungsaufgaben übernimmt, Evaluation und Schulentwicklung organisiert. Die Gestaltungsmöglichkeiten und Entscheidungsspielräume der Einzel­schule (und dort der Schulleitung, der Fachkollegien und Lehrkräfte) hängen jedoch von rechtlichen und administrativen Rahmenvorgaben ab. Hier ist dann die Rede von der Schule als *Institution*. Institutionelle Regelungen betreffen etwa Fragen der Schulstruktur wie auch der grundlegenden Schulorganisation, der Zielsetzungen über Lehrpläne oder verbindliche Standards, der Zuständigkeit und Schulaufsicht bis hin zu den Entscheidungsbefugnissen von Schulleitungen oder Mitwirkungsmöglichkeiten von Eltern.

Diese beiden Aspekte von Schule spielen im Untersuchungsansatz von PISA eine wichtige Rolle. Das Programm dient nämlich nicht allein dazu, Leistungsvergleiche

anzustellen, sondern soll auch Auskunft geben über die Kontexte, in denen die getesteten Kompetenzen entwickelt wurden. Die oben skizzierten Beispiele für Merkmale von Einzelschulen oder für institutionelle Regelungen verweisen auf Sachverhalte, die in einem hohen Maße variieren können, zwischen und innerhalb von Staaten oder anderen Gebietskörperschaften.

In Hinblick auf die erforderliche enge Auswahl von Merkmalen oder Indikatoren, die im Rahmen von PISA berücksichtigt werden können, verfolgt die OECD einige Prinzipien, die zuletzt im Framework für PISA 2012 (OECD, 2013) genauer beschrieben wurden (vgl. Klieme & Vieluf, 2013). Als Rahmen dient ein Modell, das zwischen Input-, Prozess- und Ergebnismerkmalen unterscheidet und diese zueinander in Beziehung setzt. Interessant werden etwa Merkmale auf der *Inputebene*, von denen man annehmen kann, dass sie einen Effekt auf die Gestaltung von Lerngelegenheiten und somit die Prozess- oder auch Ergebnisqualität haben können, wie etwa die Ressourcen einer Schule oder die Zusammensetzung der Schülerschaft. Der *Prozessebene* zugeordnet werden können zum Beispiel die Leistungsorientierung einer Schule, das Engagement der Lehrkräfte in Fortbildungen, das Kurs- und Zusatzangebot oder das Schulklima. Auf der *Ergebnisebene* relevant sind dann etwa Indikatoren für Schulerfolg beziehungsweise Schulversagen oder aggregierte Einschätzungen des emotionalen und sozialen Wohlbefindens der Schülerinnen und Schüler in ihrer Schule. Unter institutioneller Perspektive aufschlussreich wären etwa die Struktur des Schulsystems (etwa Gliederung und Zeitpunkte der Differenzierung), Vorgaben für Unterrichtszeit oder Regelungen, die sich auf die Dauer des Schulbesuchs auswirken können (z. B. Klassenwiederholungen).

Theoretisch stützen sich entsprechende Überlegungen zur Auswahl von Konstrukten und Indikatoren auf erziehungswissenschaftliche und soziologische Arbeiten zur Funktion von Schule, zu institutionellen Bedingungen und sozialen Prozessen in Schulen (z. B. Diederich & Tenorth, 1997; Dreeben & Barr, 1988; Fend, 2006; Meyer & Rowan, 2006). Vielseitige empirische Befunde liefert die Forschung zur Schuleffektivität, die Input-, Prozess- und Outputmerkmale unterscheidet (vgl. z. B. Scheerens & Bosker, 1997; Scheerens, 2004; Klieme, 2013). Aus dieser Forschung ist bekannt, dass bestimmte Merkmale von Schulen gewisse Anteile in der Varianz der Schulleistungen erklären können. Die Größenordnungen variieren zwischen Studien, aber die Effektstärken fallen bei Kontrolle anderer Einflussfaktoren deutlich kleiner aus als die Effektstärken für Unterrichtsmerkmale oder individuelle und familiäre Lernvoraussetzungen. Unter anderem liegt dies daran, dass Lernprozesse hauptsächlich vermittelt über den Unterricht beeinflusst werden. Direkter hängt die Schulkultur hingegen mit Werten und Einstellungen von Kindern und Jugendlichen zusammen, zum Beispiel ihrem Gefühl von sozialer Zugehörigkeit. Der Einfluss von Organisationsmerkmalen lässt sich eher ermessen, wenn man Veränderungen über mehrere Jahre beobachtet, etwa in Ergänzungsstudien zu PISA (Klieme & Steinert, 2008; Bischof, Hochweber, Hartig & Klieme, 2013; Prenzel et al., 2006).

Das internationale Erhebungskonzept von PISA ist als Querschnittstudie sehr viel schlanker gehalten. Es erfasst jede einzelne Schule nur zu einem bestimmten Zeitpunkt.

Dabei beobachtete Unterschiede zwischen Schulen auf nationaler wie internationaler Ebene müssen also mit Vorsicht interpretiert werden. Insbesondere können keine Kausalmodelle geprüft werden. Was die Daten allerdings ermöglichen, ist die Plausibilität von Vermutungen oder Hypothesen zu prüfen, die bestimmte Wirkmechanismen postulieren. In Betracht zu ziehen sind dabei auch Vermutungen, die „auf gesundem Menschenverstand“ oder eigener Anschauung beruhen, und durchaus medienrelevant kommuniziert werden, wie zum Beispiel die Annahme, dass in kleinen Klassen mehr und besser gelernt wird als in großen (vgl. Bromme, Prenzel & Jäger, 2011) oder dass gegliederte Schulsysteme stärkere Disparitäten aufweisen als eingliedrige (Wößmann, 2009). Darüber hinausgehend können PISA-Befunde zu Schulmerkmalen sehr viel mehr unter den beiden übergeordneten Funktionen des *Monitorings* und des *Benchmarkings* beitragen (vgl. Seidel & Prenzel, 2008).

Der internationale Vergleich von Bildungsergebnissen wird durch die Angaben über Unterschiede und Gemeinsamkeiten in Merkmalen von Schulen und Schulsystemen hochgradig angereichert. Auf dieser Datenbasis kann explorierend der Frage nachgegangen werden, ob sich besonders leistungsstarke Staaten hinsichtlich bestimmter Schulmerkmale hervorheben lassen, oder ob bei bestimmten Strukturmerkmalen etwa Ansprüche der Bildungsgerechtigkeit besser erfüllt werden als bei anderen. So interessant die Anstrengungen eines Benchmarkings sind, gilt es dabei jedoch immer auch mögliche Fehlerquellen für Interpretationen zu berücksichtigen (z. B. Wechselwirkungen, Effekte nicht kontrollierter Faktoren wie des kulturellen Hintergrundes; vgl. dazu etwa Vieluf, Kunter & van de Vijver, 2013).

Besonders hilfreich kann ein langfristig angelegtes Monitoring von Schulmerkmalen sein, die auf mögliche Problembereiche hinweisen. So liefern die PISA-Erhebungen im dreijährigen Abstand solide Informationen über verzögerte Schullaufbahnen, die Häufigkeiten von Schülerabsenzen oder Veränderungen in Indikatoren für das Schulklima. Entsprechende Daten vermitteln zusammen mit den Ergebnissen der Tests ein abgerundetes Bild von Stärken oder Schwächen eines Schulsystems.

Auch mit Blick auf die Belastbarkeit der Befunde sind die Informationsgrundlagen zu berücksichtigen, die PISA 2012 nutzt, um Schulmerkmale zu erfassen. Neben einigen strukturellen Grunddaten über die Schulsysteme wird der größte Teil der Informationen aus dem sogenannten *Schulfragebogen* entnommen. Dieser ist an die Leitung der Schule adressiert, die gegebenenfalls zur Beantwortung weitere Kolleginnen und Kollegen hinzuziehen kann. Die Zuverlässigkeit der Auskünfte kann für Fragen, die konkrete Ereignisse oder Fakten an den Schulen betreffen, als hoch bewertet werden. Darüber hinaus enthält der Schulfragebogen auch Fragen, die Einschätzungen aus der Sicht der Schulleitung erbitten, und deren persönliche Perspektive abbilden. Aber auch der *Schülerfragebogen* enthält einige Fragen zu Merkmalen und zur Einschätzung der besuchten Schule, die eine weitere Sicht einbeziehen.

Die Befunde, die im Folgenden über schulische Rahmenbedingungen der Kompetenzentwicklung berichtet werden, sollen einen ersten, weitgehend deskriptiven Überblick geben. Seit PISA 2000 sind Merkmale von Schulen und Schulsystemen immer

wieder berichtet worden, weil mit ihnen bestimmte Vorteile oder Schwierigkeiten für das Lernen assoziiert werden. Deshalb lohnt es sich, im Sinne eines kontinuierlichen internationalen Bildungsmonitorings, regelmäßig diese vermutlichen Erfolgsfaktoren oder Hemmnisse zu beobachten. Im Rahmen der *Follow-up*-Studie in Deutschland (eine zweite Testung der Teilnehmer an PISA 2012 nach einem weiteren Jahr Unterricht) sollen Schulmerkmale systematisch in Mehrebenenanalysen der Klassenstichproben berücksichtigt werden. Der jetzt vorgelegte nationale Bericht braucht damit nicht die Auswertungsstrategie der OECD zu verfolgen, auf der Basis eines einzigen Messzeitpunkts Leistungsvarianz durch Schulmerkmale zu erklären und damit Hinweise auf mögliche Einflussgrößen der Kompetenzentwicklung zu erhalten (vgl. OECD, 2010). Bedenken zu dieser Vorgehensweise, die auch mit Besonderheiten des Schulsystems in Deutschland zu tun haben, werden im nächsten Abschnitt ersichtlich. In den anschließenden Abschnitten werden zunächst Befunde zu institutionellen Merkmalen der Schule im internationalen Vergleich berichtet, um dann über Merkmale der Einzelschule zu berichten, die zur Gesamteinschätzung der Qualität der Schulen in Deutschland beitragen.

5.1 Institutionelle und organisatorische Rahmenbedingungen

5.1.1 Leistungsvarianz und Gliederung des Sekundarschulwesens

Die Stichprobenziehung bei PISA erfolgt in zwei grundlegenden Schritten: Zunächst werden in den teilnehmenden Staaten die Schulen über ein Zufallsverfahren ausgewählt, dann die Schülerinnen und Schüler innerhalb der gezogenen Schulen (vgl. Kapitel 10). Damit kann die in den Staaten beobachtete Gesamtvarianz in den Testwerten statistisch in zwei Varianzanteile zerlegt werden, nämlich die Varianz *zwischen* den Schulen und die Varianz *innerhalb* der Schulen. Über diese Varianzaufteilung kann festgestellt werden, ob beispielsweise die mathematische Kompetenz von Fünfzehnjährigen innerhalb der Schulen stark variiert oder ob die durchschnittliche mathematische Kompetenz der Jugendlichen auf verschiedenen Schulen in einem Staat besonders unterschiedlich ist. Eine entsprechende Aufschlüsselung kann etwas über Qualitätsunterschiede zwischen Schulen aussagen oder auch über unterschiedliche Zusammensetzungen und Rekrutierungsstrategien von Schulen.

In allen bisherigen PISA-Erhebungsrunden wurde festgestellt, dass der zwischen den Schulen liegende Anteil an der Leistungsvarianz von Land zu Land stark variiert (OECD, 2009). Im OECD-Mittel liegen etwa 30 Prozent der Varianz bei Schülerkompetenzen zwischen den Schulen. Wiederholt wurde gefunden, dass sich etwa Schulen in skandinavischen Ländern kaum voneinander unterscheiden (Varianzanteil zwischen Schulen unter 20 Prozent), während dieser Anteil in Ländern wie Belgien, Deutschland oder Österreich bei über 50 Prozent liegt (OECD, 2009).

Ein unmittelbarer Grund dafür ist die unterschiedliche Gliederung der Schulsysteme. Früher oder später differenzieren alle Staaten in ihren Schulsystemen nach Leistung. Die Zielgruppe von PISA, die Fünfzehnjährigen, wird in einer Reihe von Staaten noch gemeinsam in einer Schulart unterrichtet, in anderen Staaten verteilt sich die Untersuchungsgruppe auf zwei oder mehr Schularten. In einem Schulsystem, das in dieser Altersgruppe nach Leistung differenziert (hat), ist die Varianz zwischen Schulen zwangsläufig höher als in nicht differenzierenden Schulsystemen. Andererseits ist zu erwarten, dass in einem gegliederten System die Varianz innerhalb von Schulen geringer ist. Das bedeutet zugleich, dass die Relationen der Varianzanteile *zwischen und innerhalb der Schulen* nicht als Indikator für Chancengerechtigkeit verstanden werden dürfen. Chancengleichheit oder -ungleichheit zeigt sich vielmehr daran, wie stark Lernergebnisse – insbesondere die erworbenen Kompetenzen und Abschlüsse – an die Herkunft der Schüler gekoppelt sind (vgl. hierzu Kapitel 8 dieses Bandes). Auch über die Qualität von Schulen sagen die Varianzanteile nichts aus; hierzu müssen pädagogische Prozesse auf der Ebene der Schule (siehe unten) und des Unterrichts (siehe Kapitel 4) betrachtet werden. Inwiefern sich die Gliederung von Schulsystemen auf die Qualität und Chancengerechtigkeit auswirkt, ist demnach eine empirisch zu untersuchende Frage. In den bisherigen nationalen Berichten zu PISA wurde auf diesen Punkt immer wieder hingewiesen, denn Deutschland ist ein Staat mit einem gegliederten Schulsystem, das gewissermaßen Varianz zwischen den Schulen produziert (Baumert et al., 2001; Prenzel et al., 2004; Prenzel et al., 2007). Die internationale Forschung kommt – auch auf der Basis der PISA-Studien – zu dem Ergebnis, dass die Gliederung der Schulsysteme nicht mit dem mittleren Leistungsniveau der Staaten zusammenhängt, wohl aber mit dem Ausmaß der Kopplung zwischen Leistung und sozialer Herkunft (Duru-Bellat & Suchaut, 2005; Gorard & Smith, 2004).

Was sagen nun die Daten aus PISA 2012? Tabelle 5.1 stellt für die einzelnen OECD-Staaten dar, wie viele allgemeinbildende Schularten für Jugendliche im Alter von fünfzehn Jahren angeboten werden und wann die Ausdifferenzierung in mehrere Schularten erfolgt. Hier handelt es sich also um institutionelle Rahmenbedingungen, innerhalb derer die Schulen eines Landes arbeiten müssen. In der Tabelle wird zudem für jeden Staat die Gesamtvarianz angegeben sowie die prozentuale Aufschlüsselung der Varianzanteile *innerhalb der Schulen* und *zwischen den Schulen*.

Wie der Tabelle zu entnehmen ist, findet in dreizehn von 34 OECD-Staaten eine Leistungsdifferenzierung erst im Alter von sechzehn Jahren statt. Unter den 21 gegliederten Schulsystemen differenzieren acht Staaten im Alter von fünfzehn Jahren, drei weitere Staaten im Alter von vierzehn Jahren. Am frühesten, nämlich im Alter von zehn Jahren, findet die Differenzierung in Deutschland und Österreich statt. Ebenfalls relativ früh, im Alter von elf oder zwölf Jahren, differenzieren die Niederlande, die Schweiz, die Slowakische Republik, die Tschechische Republik, die Türkei und Ungarn. Anzumerken bleibt, dass diese Betrachtung auf einer groben strukturellen Ebene bleibt, die interne, kurz- oder längerfristige, fachbezogene oder fächerübergreifende Differenzierungsmaß-

nahmen (flexible Einschulung, Gruppierung nach Fähigkeit, Bildungsgänge/Tracking) ignoriert.

Betrachtet man vor diesem Hintergrund zunächst die Gesamtvarianz mathematischer Kompetenz, so weist eine Reihe von Staaten Gesamtstreuungen deutlich über dem OECD-Mittel auf. Auf der einen Seite finden sich in dieser Gruppe Staaten wie die Schweiz (106), Deutschland (109), die Slowakische Republik (118) und Belgien (123), die relativ früh differenzieren. Aber auch in Staaten, die erst ab dem Alter von fünfzehn oder sechzehn Jahren differenzieren, wie zum Beispiel Israel (129), Australien (110) und Neuseeland (118), ist eine beträchtliche Gesamtvarianz festzustellen. Eine große Varianz ist also nicht zwingend mit einem mehrgliedrigem Sekundarschulsystem verbunden.

Anders sieht das bei der Aufteilung der Varianz aus: Wie zu erwarten ist, findet man in den OECD-Staaten mit mehreren Sekundarschularten (drei und mehr) relativ viel Varianz *zwischen den Schulen*. Im OECD-Durchschnitt macht dieser Varianzanteil 37 Prozent der Gesamtvarianz aus. Deutlich über diesem Wert befinden sich die mehrgliedrigen Systeme von Deutschland (53 Prozent), Belgien (50 Prozent), Italien (51 Prozent), den Niederlanden (66 Prozent), der Slowakischen Republik (50 Prozent), Slowenien (59 Prozent), der Tschechischen Republik (51 Prozent), der Türkei (62 Prozent) sowie Ungarn (62 Prozent). In Staaten mit einer einzigen Sekundarschulform für Fünfzehnjährige hingegen liegt der Varianzanteil zwischen den Schulen mit Ausnahme von Chile bei lediglich 7 bis 33 Prozent und damit stets unter dem OECD-Durchschnitt.

Setzt man die Angaben aus Tabelle 5.1 schließlich in Beziehung zu den jeweiligen Mittelwerten mathematischer Kompetenz (vgl. Kapitel 2), so zeigt sich: Staaten, die ebenso wie Deutschland drei oder mehr verschiedene Schularten der Sekundarstufe haben, erreichen durchaus unterschiedliche Mittelwerte. Wie in Deutschland liegt der Mittelwert in Mathematik bei PISA 2012 in Belgien und den Niederlanden signifikant über dem OECD-Durchschnitt. Belgien hat ebenso wie Deutschland vier allgemeinbildende Sekundarschulprogramme, die unterschieden werden; in den Niederlanden sind es mittlerweile sogar sieben (PISA 2003: vier; vgl. Prenzel et al., 2004). In Deutschland werden dabei die Integrierten Gesamtschulen und die Schulen mit mehreren Bildungsgängen für den internationalen Vergleich zusammengefasst. Die Schülerinnen und Schüler in Italien, der Slowakischen Republik, in Ungarn und der Türkei jedoch erreichen Mittelwerte der mathematischen Kompetenz, die signifikant unter dem OECD-Durchschnitt liegen, obwohl auch in diesen Staaten drei bis fünf Schularten für die Sekundarstufe angeboten werden.

Auch aus dem Alter des Übertritts in die Sekundarstufe lässt sich nicht auf das mittlere Kompetenzniveau von Schülerinnen und Schülern eines Staates schließen. In den beiden kompetenzstarken Staaten Belgien und den Niederlanden erfolgt der Einstieg in die Sekundarstufe mit 12 Jahren vergleichsweise früh und beide pflegen vier beziehungsweise sieben Schularten, die von Fünfzehnjährigen besucht werden können. Österreich, die Schweiz und Luxemburg differenzieren ähnlich früh wie Deutschland, aber dort wird weniger Leistungsvarianz durch Unterschiede zwischen Schulen erzeugt. Die Differenzierung des Sekundarschulwesens scheint also erwartungsgemäß mit einer stär-

Tabelle 5.1: Varianz der mathematischen Kompetenz im internationalen Vergleich

Staat	Zahl der Schularten/ Programme für 15-Jährige	Alter der Differenzierung	Varianz der Schülerkompetenz und prozentuale Aufteilung		
			Gesamtvarianz	Varianz innerhalb der Schulen	Varianz zwischen den Schulen
Australien	1	16	110	72	28
Belgien	4	12	123	50	50
Chile	1	16	76	57	43
Dänemark	1	16	79	84	16
Deutschland	4	10	109	47	53
Estland	1	15	77	83	17
Finnland	1	16	83	93	7
Frankreich	3	15	m	m	m
Griechenland	2	15	90	68	32
Irland	4	15	84	82	18
Island	1	16	99	90	10
Israel	2	15	129	57	43
Italien	4	14	100	49	51
Japan	2	15	103	47	53
Kanada	1	16	93	81	19
Korea	3	14	114	61	39
Luxemburg	4	13	130	59	41
Mexiko	3	15	65	65	35
Neuseeland	1	16	118	76	24
Niederlande	7	12	99	34	66
Norwegen	1	16	95	87	13
Österreich	4	10	99	52	48
Polen	1	16	95	79	21
Portugal	3	15	104	70	30
Schweden	2	16	98	88	12
Schweiz	4	12	106	64	36
Slowakische Republik	5	11	118	50	50
Slowenien	3	14	98	41	59
Spanien	1	16	91	81	19
Tschechische Republik	6	11	104	49	51
Türkei	3	11	98	38	62
Ungarn	3	11	102	38	62
Vereinigte Staaten	1	16	95	76	24
Vereinigtes Königreich	1	16	105	72	28
OECD-Durchschnitt	3	14	100	63	37

Anmerkung: m: fehlende Angaben.

keren Unterschiedlichkeit von Schulen innerhalb des Landes einherzugehen; hierbei liegt Deutschland zwischen den Kennwerten der Nachbarstaaten.

Varianz innerhalb und zwischen Schulen in Deutschland

In Deutschland splittet sich die Varianz der Schülerkompetenz in Mathematik jeweils in etwa zur Hälfte auf Unterschiede *innerhalb von Schulen* (47 Prozent) und *zwischen Schulen* (53 Prozent) auf. In der Varianz zwischen den Schulen sind einerseits die Unterschiede zwischen Schularten, andererseits die Unterschiede zwischen Einzelschulen innerhalb der Schularten kombiniert. Berechnet man nun den Varianzanteil zwischen den Schulen *innerhalb* der einzelnen Schularten, so sollte dieser deutlich niedriger ausfallen. Abbildung 5.1 zeigt die Varianzanteile mathematischer Kompetenz zwischen den Schulen bei den Sekundarschularten in Deutschland.

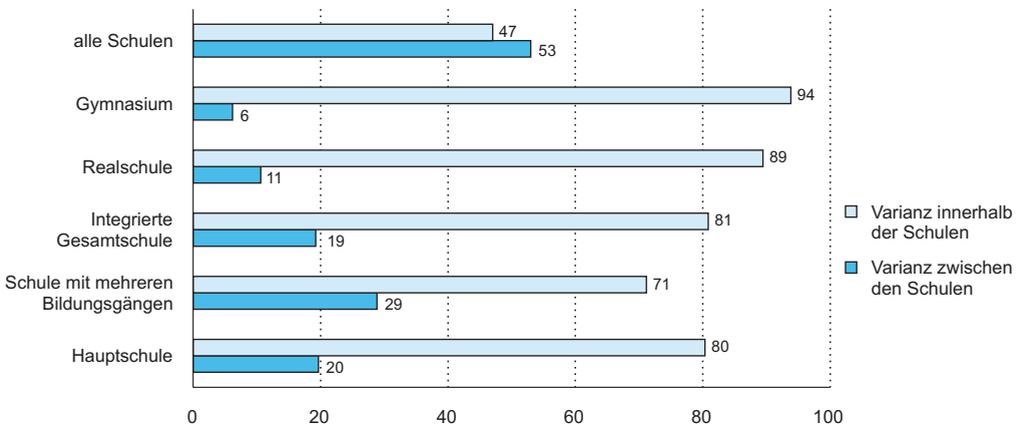


Abbildung 5.1: Varianzanteile der Schülerkompetenz (Mathematik) zwischen den Schulen, nach Schulart

Die Varianzanteile zwischen den Schulen variieren in Abhängigkeit von der Schulart beträchtlich. Als homogen erweisen sich die Gymnasien mit einer sehr großen Varianz *innerhalb der Schulen* (94 Prozent) und einer sehr geringen Varianz (6 Prozent) *zwischen den Schulen*. Dieser Befund unterstreicht, dass die Gymnasien auch über die Ländergrenzen in Deutschland hinweg (zumindest bezogen auf die mathematische Kompetenz) eine relativ hohe Ähnlichkeit aufweisen. Nennenswerte Unterschiede *zwischen den Schulen* (19 Prozent) finden sich bei den Integrierten Gesamtschulen, den Hauptschulen (20 Prozent) und insbesondere den Schulen mit mehreren Bildungsgängen (29 Prozent). In Anbetracht der Vielfalt an Varianten für Schulen mit mehreren Bildungsgängen muss die große Varianz zwischen diesen Schulen nicht überraschen.

Ein Grund für die relativ geringen Unterschiede zwischen Gymnasien dürften die deutschlandweiten Übereinstimmungen in den Lehrplänen und einheitlichen Prüfungsanforderungen sein (Pant et al., 2013), was als Ausdruck eines gemeinsamen Verständnisses für ihren Bildungsauftrag verstanden werden kann; zugleich ermöglicht die leistungsgesteuerte Auswahl der Schülerschaft – sowohl beim Übergang ins Gymnasium wie auch bei Schulartwechseln – eine stärkere Homogenisierung der Schülerschaft. Die anderen, *nicht gymnasialen* Schularten unterscheiden sich in ihren curricularen Konzeptionen stärker voneinander, was sich auch in zahlreichen verschiedenen Bezeichnungen für diese Schularten in den 16 Ländern zeigt.

5.1.2 Alter und Klassenstufe der PISA-Kohorte

Mit der Grundentscheidung von PISA, die fünfzehnjährigen Schülerinnen und Schüler in den Fokus des internationalen Vergleichs zu stellen, wird der Umgang mit der Lebenszeit in Schulsystemen zu einem Thema. Der Zeitpunkt der Einschulung, Wahlmöglichkeiten dieses Zeitpunkts oder die Möglichkeiten der Wiederholung einer Jahrgangsstufe beeinflussen die Verteilung der Fünfzehnjährigen auf die Klassenstufen. Wegen der Möglichkeit der Wiederholung oder der verzögerten Einschulung sprechen wir im Folgenden von „Klassenstufen“ statt „Jahrgangsstufen“.

Seit PISA 2000 führt dieser Erhebungsansatz gelegentlich zu Missverständnissen, etwa bezüglich der Fairness des Tests. In einem Schulsystem, das früher einschult oder auf Klassenwiederholungen verzichtet, befinden sich die Fünfzehnjährigen auf höheren Klassenstufen und hätten dadurch den Vorteil, mehr aufbauenden Unterricht genossen zu haben. Auf der anderen Seite könnten flexible Einschulungen und Klassenwiederholungen eine bessere Passung von Unterrichtsangeboten an Schülervoraussetzungen bedeuten und zur Sicherung eines soliden Grundverständnisses beitragen. Der Erhebungsansatz von PISA ermöglicht jedenfalls, einen Überblick über unterschiedliche Varianten des Umgangs mit Zeit in Bildungssystemen zu erhalten.

Tabelle 5.2 gibt eine Übersicht darüber, auf welche Klassenstufen die PISA-Kohorte der fünfzehnjährigen Schülerinnen und Schüler in den OECD-Staaten bei PISA 2012 verteilt war. Als Modalklasse ist diejenige Klassenstufe bezeichnet, die im jeweiligen Staat von der Mehrheit der Fünfzehnjährigen besucht wird.

Tabelle 5.2: Klassenstufe der PISA-Kohorte in den OECD-Staaten

Staat	Modal- klasse	Varianz in der besuchten Klassenstufe		Prozentualer Anteil der Schülerinnen und Schüler					
		SD	(SE)	Klassenstufen unter der Modalklasse		Modalklasse		Klassenstufen über der Modalklasse	
				%	(SE)	%	(SE)	%	(SE)
Australien	10	0.5	(0.0)	10.9	(0.5)	70.0	(0.6)	19.1	(0.4)
Belgien	10	0.7	(0.0)	37.4	(0.6)	59.5	(0.6)	3.0	(0.3)
Chile	10	0.7	(0.0)	27.1	(1.2)	66.1	(1.2)	6.7	(0.3)
Dänemark	9	0.4	(0.0)	18.3	(0.9)	80.6	(0.8)	1.0	(0.2)
Deutschland	9	0.7	(0.0)	10.6	(0.6)	51.9	(0.8)	37.5	(0.9)
Estland	9	0.5	(0.0)	22.7	(0.7)	75.4	(0.7)	1.9	(0.3)
Finnland	9	0.4	(0.0)	14.9	(0.4)	85.0	(0.4)	0.1	(0.1)
Frankreich	10	0.6	(0.0)	29.8	(0.7)	66.6	(0.7)	3.6	(0.3)
Griechenland	10	0.3	(0.0)	5.5	(1.0)	94.5	(1.0)	0.0	n/a
Irland	9	0.7	(0.0)	1.9	(0.2)	60.5	(0.8)	37.6	(0.8)
Island	10	0.0	(0.0)	0.0	(0.0)	100.0	(0.0)	0.0	(0.0)
Israel	10	0.4	(0.0)	17.5	(0.9)	81.7	(0.9)	0.8	(0.3)
Italien	10	0.5	(0.0)	18.9	(0.6)	78.5	(0.7)	2.6	(0.2)
Japan	10	0.0	(0.0)	0.0	(0.0)	100.0	(0.0)	0.0	(0.0)
Kanada	10	0.4	(0.0)	14.4	(0.6)	84.6	(0.6)	1.1	(0.1)
Korea	10	0.2	(0.0)	5.9	(0.8)	93.8	(0.8)	0.2	(0.1)
Luxemburg	9	0.7	(0.0)	10.9	(0.2)	50.7	(0.1)	38.5	(0.1)
Mexiko	10	0.7	(0.0)	37.0	(1.1)	60.8	(1.1)	2.2	(0.3)
Neuseeland	11	0.3	(0.0)	6.3	(0.4)	88.3	(0.5)	5.4	(0.4)
Niederlande	10	0.6	(0.0)	50.3	(1.1)	49.2	(1.1)	0.5	(0.1)
Norwegen	10	0.1	(0.0)	0.4	(0.1)	99.4	(0.1)	0.2	(0.0)
Österreich	10	0.6	(0.0)	49.0	(1.0)	51.0	(1.0)	0.1	(0.0)
Polen	9	0.3	(0.0)	4.6	(0.4)	94.9	(0.4)	0.5	(0.2)
Portugal	10	0.8	(0.0)	35.6	(1.9)	54.9	(2.2)	9.5	(1.4)
Schweden	9	0.2	(0.0)	3.7	(0.3)	94.0	(0.6)	2.2	(0.5)
Schweiz	9	0.6	(0.0)	13.5	(0.8)	60.6	(1.0)	25.9	(1.0)
Slowakische Republik	10	0.7	(0.0)	45.7	(1.4)	52.7	(1.4)	1.6	(0.5)
Slowenien	10	0.3	(0.0)	5.4	(0.8)	90.7	(0.8)	3.9	(0.2)
Spanien	10	0.7	(0.0)	34.0	(0.6)	66.0	(0.6)	0.0	(0.0)
Tschechische Republik	9	0.6	(0.0)	4.9	(0.5)	51.1	(1.2)	44.1	(1.3)
Türkei	10	0.6	(0.0)	30.3	(1.2)	65.5	(1.2)	4.3	(0.3)
Ungarn	9	0.6	(0.0)	11.6	(0.9)	67.8	(0.9)	20.6	(0.6)
Vereinigte Staaten	10	0.5	(0.0)	12.0	(1.1)	71.2	(1.1)	16.8	(0.8)
Vereinigtes Königreich	11	0.2	(0.0)	1.4	(0.3)	95.0	(0.3)	3.6	(0.1)
OECD-Durchschnitt	10	0.5	(0.0)	17.4	(0.1)	73.9	(0.2)	8.7	(0.1)

Anmerkung: Modalklasse: Diejenige Klassenstufe, die in den Teilnehmerstaaten von der Mehrheit der Fünfzehnjährigen besucht wird. Aufgrund von Rundungsdifferenzen weicht die Zeilensumme teilweise minimal von 100 ab.

Betrachtet man zunächst die Werte für die OECD, dann ist in elf Staaten die 9. Klassenstufe die (am häufigsten besuchte) Modalklasse, in 21 Staaten ist es dagegen die 10. Klassenstufe. Im Vereinigten Königreich befinden sich 95 Prozent der Schülerinnen und Schüler auf der 11. Klassenstufe, in Neuseeland 88 Prozent; in diesen beiden Staaten beginnt das erste Schuljahr bereits für Fünfjährige mit einem entsprechend altersangemessenen Curriculum. Daraus resultieren die OECD-Mittelwerte: Knapp 74 Prozent der Schülerinnen und Schüler besuchen die 10. Klassenstufe, 17 Prozent maximal die 9. Klassenstufe und knapp 9 Prozent mindestens die 11. Klassenstufe. Deutschland hat dabei einen vergleichsweise niedrigen Modalwert und eine breite Verteilung der Jugendlichen auf verschiedene Klassenstufen: Knapp 52 Prozent der Fünfzehnjährigen besuchten zum Testzeitpunkt die 9. Klassenstufe, auf der 10. Klassenstufe oder höher befanden sich 37,5 Prozent, und 10,6 Prozent waren höchstens auf der 8. Klassenstufe.

Alter bei der Einschulung und Klassenwiederholung

Im OECD-Durchschnitt starten Kinder im Alter von 6 Jahren ihre Schullaufbahn mit dem Besuch einer Grundschule. Die Varianz dieses Zeitpunktes ist allerdings beträchtlich. So beginnt die Grundschule in Irland bereits im Alter von 4 Jahren und in mehreren Staaten erst mit 7 Jahren (Polen, Finnland, Türkei, Schweden, Ungarn, Dänemark, Schweiz und Norwegen). In einer Reihe von Staaten bestehen Möglichkeiten für eine flexible Einschulung, die dann etwa im Alter von fünf Jahren oder früher beginnen kann. Auch in Deutschland ist es seit einiger Zeit möglich, frühzeitig eingeschult zu werden, Möglichkeiten einer verzögerten Einschulung wurden eher eingeschränkt. Das Durchschnittsalter der PISA-Kohorte 2012 zu Schulbeginn liegt in Deutschland bei 6 Jahren und entspricht damit dem OECD-Durchschnitt.

Worauf lassen sich dann die – vom OECD-Durchschnitt abweichenden – Verteilungen der Fünfzehnjährigen in Deutschland auf die Klassenstufen zurückführen? In Betracht zu ziehen ist hier der Umgang mit Klassenwiederholungen, der in den Staaten unterschiedlich gehandhabt wird.

In einer Reihe von OECD-Staaten gibt es gar keine Klassenwiederholungen, in anderen Staaten besteht diese Möglichkeit auf Wunsch der Schülerinnen und Schüler, und in einer weiteren Gruppe sind Klassenwiederholungen ein mehr oder weniger häufig eingesetztes Instrument (OECD, 2010). Normalerweise sollten Schülerinnen und Schüler mit zunehmenden Alter höhere Klassenstufen besuchen, so dass mit dem Alter auch die Klassenstufe ansteigt. Wird während der Schullaufbahn eine Klassenstufe (oder auch mehrere) wiederholt, so verzögert sich der kumulative schulische Werdegang.

Tabelle 5.3 bildet ab, wie viele Schülerinnen und Schüler bei PISA 2012 in den OECD-Staaten angaben, bereits mindestens einmal eine Klasse wiederholt zu haben und differenziert dabei wiederholte Schuljahre in der Primarstufe sowie in der Sekundarstufe I.

Tabelle 5.3: Prozentualer Anteil der Schülerinnen und Schüler, die mindestens eine Klasse wiederholt haben

Staat	Prozentualer Anteil der Schülerinnen und Schüler, die mindestens eine Klasse wiederholt haben								Klassenwiederholungen in der bisherigen Schullaufbahn	
	Grundschule				Sekundarstufe I					
	einmal		zweimal und mehr		einmal		zweimal und mehr		%	(SE)
	%	(SE)	%	(SE)	%	(SE)	%	(SE)		
Australien	6.4	(0.2)	0.3	(0.1)	1.2	(0.1)	0.1	(0.0)	7.5	(0.3)
Belgien	17.8	(0.6)	2.7	(0.2)	15.5	(0.6)	1.2	(0.1)	36.1	(0.6)
Chile	9.9	(0.6)	2.7	(0.5)	5.9	(0.6)	1.2	(0.3)	25.2	(1.2)
Dänemark	3.9	(0.4)	0.1	(0.0)	1.0	(0.2)	0.0	(0.0)	4.7	(0.4)
Deutschland	9.6	(0.6)	0.7	(0.1)	12.3	(0.6)	0.5	(0.1)	20.3	(0.8)
Estland	1.7	(0.2)	0.2	(0.1)	1.6	(0.3)	0.2	(0.1)	3.5	(0.4)
Finnland	3.1	(0.1)	0.1	(0.1)	0.7	(0.2)	0.0	(0.0)	3.8	(0.4)
Frankreich	16.5	(0.7)	0.5	(0.1)	13.9	(0.7)	0.5	(0.1)	28.4	(0.8)
Griechenland	0.9	(0.2)	0.7	(0.1)	2.8	(0.5)	1.2	(0.3)	4.5	(0.7)
Irland	7.7	(0.4)	0.1	(0.1)	1.0	(0.1)	0.1	(0.0)	8.6	(0.4)
Island	0.5	(0.1)	0.2	(0.1)	0.6	(0.1)	0.2	(0.1)	1.2	(0.2)
Israel	1.2	(0.2)	0.0	(0.0)	0.7	(0.2)	0.0	(0.0)	1.9	(0.3)
Italien	0.9	(0.1)	0.1	(0.0)	6.1	(0.3)	1.4	(0.2)	17.1	(0.5)
Japan	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Kanada	3.9	(0.2)	0.3	(0.1)	3.8	(0.2)	0.7	(0.1)	8.0	(0.3)
Korea	2.4	(0.2)	0.8	(0.1)	2.2	(0.2)	0.9	(0.1)	3.6	(0.3)
Luxemburg	19.3	(0.5)	2.2	(0.2)	18.5	(0.6)	0.8	(0.1)	34.5	(0.5)
Mexiko	11.2	(0.4)	1.4	(0.1)	3.1	(0.3)	0.3	(0.0)	15.5	(0.6)
Neuseeland	3.7	(0.3)	0.3	(0.1)	1.5	(0.2)	0.3	(0.1)	5.4	(0.3)
Niederlande	20.2	(1.0)	0.7	(0.1)	7.8	(0.6)	0.1	(0.0)	27.6	(0.9)
Norwegen	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Österreich	5.0	(0.4)	0.1	(0.0)	4.6	(0.4)	0.3	(0.1)	11.9	(0.7)
Polen	1.3	(0.2)	0.2	(0.1)	2.9	(0.3)	0.2	(0.1)	4.2	(0.4)
Portugal	17.9	(1.2)	5.4	(0.6)	17.5	(1.4)	2.4	(0.3)	34.3	(1.9)
Schweden	3.1	(0.3)	0.2	(0.1)	1.1	(0.2)	0.2	(0.1)	4.0	(0.4)
Schweiz	12.7	(0.7)	0.5	(0.1)	7.9	(0.5)	0.2	(0.0)	19.9	(0.9)
Slowakische Republik	3.5	(0.5)	1.4	(0.2)	2.9	(0.4)	0.5	(0.1)	7.6	(0.6)
Slowenien	m	m	m	m	2.5	(0.4)	0.4	(0.1)	3.4	(0.4)
Spanien	12.9	(0.4)	0.8	(0.1)	25.0	(0.6)	2.7	(0.2)	32.9	(0.6)
Tschechische Republik	1.9	(0.4)	0.3	(0.1)	3.0	(0.4)	0.3	(0.1)	4.9	(0.6)
Türkei	2.3	(0.3)	0.1	(0.0)	m	m	m	m	14.2	(0.9)
Ungarn	4.2	(0.5)	0.7	(0.2)	4.2	(0.5)	1.5	(0.4)	10.8	(0.9)
Vereinigte Staaten	10.7	(0.9)	0.4	(0.1)	4.0	(0.3)	0.1	(0.0)	13.3	(1.0)
Vereinigtes Königreich	1.8	(0.2)	0.1	(0.0)	0.7	(0.1)	0.1	(0.0)	2.7	(0.3)
OECD-Durchschnitt	7.0	(0.1)	0.7	(0.0)	5.7	(0.1)	0.6	(0.0)	12.4	(0.1)

Anmerkung: m: fehlende Angaben.

Im OECD-Durchschnitt haben 12.4 Prozent der Schülerinnen und Schüler vor dem PISA-Test bereits mindestens einmal eine Klasse wiederholt. 7.7 Prozent berichten von einer oder mehreren Wiederholungen in der Primarstufe, 6.3 Prozent in der Sekundarstufe I (und, in der Tabelle nicht einzeln ausgewiesen, 2.1 Prozent in der Sekundarstufe II).

In Japan und Norwegen gibt es die Möglichkeit der Klassenwiederholung nicht. In den nordischen Staaten besteht ansonsten nur die Möglichkeit zur Klassenwiederholung, wenn diese von den Schülerinnen und Schülern und deren Eltern gewünscht wird. Entsprechend gering (unter 5 Prozent) sind dort die Wiederholungsquoten bis zum Alter von fünfzehn Jahren. Auch im Vereinigten Königreich sind Klassenwiederholungen nur in Ausnahmefällen möglich (2.7 Prozent). In Korea existierte die Praxis der Klassenwiederholung ebenfalls bis vor kurzem nicht; mittlerweile geben hier jedoch auch kleine Gruppen der Schülerinnen und Schüler an, ein Schuljahr zweimal besucht zu haben. Zieht man wiederum die in Kapitel 2 berichteten Leistungsergebnisse mit heran, so sieht man, dass Staaten, die keine Wiederholungen vorsehen oder nur in Ausnahmefällen zulassen, durchaus sehr starke Leistungen in den Mathematiktests vorweisen können.

Die Wiederholungsquote in Deutschland fällt mit 20.3 Prozent höher aus als im OECD-Durchschnitt und ist damit ähnlich hoch wie in Italien oder der Schweiz. Nochmals deutlich höher, zwischen 25 und 39 Prozent, liegt die Klassenwiederholungsrate in Portugal und Spanien, in Frankreich, Belgien, Luxemburg und den Niederlanden sowie in Chile. Mit Ausnahme der Niederlande liegen die Mittelwerte der mathematischen Kompetenz in diesen Staaten im Bereich des OECD-Mittelwerts oder darunter.

Insgesamt weist der Forschungsstand zum Thema Klassenwiederholungen darauf hin, dass deren lernförderliche Wirkungen gering sind, in Einzelfällen greifen können, und insgesamt mit erheblichen Nebenwirkungen verbunden sind (Boller, Möller & Palowski, 2013; Ehmke, Drechsel & Carstensen, 2008; Hong & Raudenbush, 2005; Krohne, Meier & Tillmann, 2011; OECD, 2011). Die mit Klassenwiederholungen verbundenen Kosten werden immer wieder an den durch Klassenwiederholungen gewonnenen Erträgen gemessen und allzu oft resultiert dabei, dass sich die aufgewendeten Ressourcen nicht auszahlen (Alexander, Entwisle & Dauber, 2003; Klemm, 2009). Diese Befundlage rechtfertigt es jedenfalls nicht, eine verordnete Klassenwiederholung als wirksame pädagogische Maßnahme zu propagieren.

Die in PISA 2012 berichteten Wiederholungsquoten werden in Tabelle 5.4 auf die verschiedenen Schularten in Deutschland und Klassenstufen aufgeschlüsselt. Darüber hinaus werden die prozentualen Anteile der Schülerinnen und Schüler abgebildet, die angaben, schon ein- oder mehrmals ein Schuljahr zweimal besucht zu haben (differenziert nach Grundschule und Sekundarstufe I).

Tabelle 5.4: Anteile der Fünfzehnjährigen in Prozent, Verteilung auf Klassenstufen und Klassenwiederholungen

	Hauptschule	Schule mit mehreren Bildungsgängen	Integrierte Gesamtschule	Realschule	Gymnasium	alle Schulen
Klassenstufe						
Klasse 7 oder 8	26.9	14.8	8.7	10.8	3.0	10.5
Klasse 9	59.5	58.3	55.9	53.6	47.4	52.1
Klasse 10	13.6	25.8	35.4	35.8	47.9	36.6
Klasse 11		1.1			1.6	0.7
Wiederholer						
Grundschule	30.4	14.6	14.9	8.8	1.2	9.6
Sekundarstufe I	26.7	14.6	7.2	19.8	6.6	12.3

Der Tabelle kann entnommen werden, dass sich in den Hauptschulen der größte prozentuale Anteil von Fünfzehnjährigen befindet, die in ihrer bisherigen Schullaufbahn eine Klasse wiederholt haben. Dies hängt nicht zuletzt mit Schulartwechseln zusammen: Wer beispielsweise im Gymnasium nicht versetzt wird, wechselt möglicherweise auf eine andere Schulart. Die Klassenwiederholungen sind also häufig nicht „hausgemacht“, sondern betreffen zugleich einen Wechsel der Schulart und werden der aufnehmenden Schulart zugerechnet.

Insgesamt ist die Klassenwiederholungsquote in Deutschland seit PISA 2003 von 23.1 Prozent (Drechsel & Senkbeil, 2004) auf 20.3 Prozent bei PISA 2012 zurückgegangen. Dieser Trend könnte in Zukunft verstärkt werden, denn in einzelnen Ländern Deutschlands wird seit kurzem auf die Wiederholung von Klassen verzichtet; teilweise begrenzt auf einzelne Schulformen oder als Modellversuch, der nach und nach erweitert werden kann. Aus der internationalen Vergleichsperspektive wie aus Forschungssicht dürfte es dabei angezeigt sein, die Möglichkeit einer „Wiederholung auf Wunsch“ von Seiten der Schülerinnen und Schüler sowie ihrer Eltern einzuräumen.

5.1.3 Lernzeit

Die Menge an Zeit, die Schülerinnen und Schüler mit Lernen innerhalb der Schule und bei häuslichen Übungen verbringen, variiert im internationalen Vergleich beträchtlich.

Tabelle 5.5 gibt wieder, wie viele Zeitstunden pro Woche die fünfzehnjährigen Schülerinnen und Schüler mit Unterricht in den Fächern Mathematik, Naturwissenschaften und in ihrer Testsprache (in Deutschland: Deutschunterricht) verbringen. Da die Dauer der Schulstunden international stark schwankt und dadurch nur schwer vergleichbar ist, sind hier als Vergleichsebene Zeitstunden (60 Minuten) angegeben.

Tabelle 5.5: Anzahl der Zeitstunden pro Woche, die in OECD-Staaten mit Unterricht in Mathematik, Naturwissenschaften oder der Testsprache verbracht werden

Staat	Normaler Mathematikunterricht		Normaler Naturwissenschaftsunterricht		Normaler Unterricht in der Testsprache	
	Lernzeit (Zeitstunden)		Lernzeit (Zeitstunden)		Lernzeit (Zeitstunden)	
	<i>M</i>	(<i>SE</i>)	<i>M</i>	(<i>SE</i>)	<i>M</i>	(<i>SE</i>)
Chile	6.6	(0.1)	4.9	(0.1)	6.2	(0.1)
Kanada	5.2	(0.0)	5.1	(0.0)	5.3	(0.0)
Portugal	4.8	(0.1)	4.0	(0.2)	4.0	(0.1)
Israel	4.2	(0.0)	3.3	(0.1)	3.2	(0.0)
Vereinigte Staaten	4.2	(0.1)	4.2	(0.1)	4.3	(0.1)
Mexiko	4.2	(0.0)	4.2	(0.0)	3.9	(0.0)
Island	4.1	(0.0)	2.4	(0.0)	4.0	(0.0)
Neuseeland	4.0	(0.0)	4.1	(0.1)	4.0	(0.0)
Australien	3.9	(0.0)	3.8	(0.0)	3.9	(0.0)
Japan	3.9	(0.1)	2.8	(0.1)	3.4	(0.0)
Italien	3.9	(0.0)	2.3	(0.0)	4.6	(0.0)
Vereinigtes Königreich	3.8	(0.0)	4.9	(0.1)	3.9	(0.0)
Dänemark	3.7	(0.1)	2.9	(0.0)	5.2	(0.1)
Estland	3.7	(0.0)	3.3	(0.0)	3.3	(0.0)
Belgien	3.6	(0.0)	3.2	(0.0)	3.6	(0.0)
Korea	3.6	(0.1)	3.3	(0.1)	3.4	(0.0)
Spanien	3.5	(0.0)	3.1	(0.0)	3.4	(0.0)
Griechenland	3.5	(0.0)	3.8	(0.0)	2.8	(0.0)
Frankreich	3.5	(0.0)	2.9	(0.0)	3.6	(0.0)
Schweiz	3.4	(0.0)	2.7	(0.1)	3.4	(0.1)
Luxemburg	3.4	(0.0)	2.6	(0.0)	3.1	(0.0)
Norwegen	3.3	(0.0)	2.4	(0.0)	3.6	(0.0)
Polen	3.3	(0.0)	2.8	(0.0)	3.7	(0.0)
Deutschland	3.3	(0.0)	4.2	(0.1)	3.2	(0.0)
Irland	3.1	(0.0)	2.4	(0.0)	3.0	(0.0)
Tschechische Republik	3.0	(0.0)	3.6	(0.1)	3.0	(0.0)
Schweden	3.0	(0.0)	3.1	(0.0)	3.0	(0.0)
Slowakische Republik	3.0	(0.0)	2.7	(0.1)	3.0	(0.0)
Finnland	2.9	(0.0)	3.1	(0.0)	2.5	(0.0)
Türkei	2.9	(0.0)	2.8	(0.1)	3.3	(0.0)
Niederlande	2.8	(0.0)	2.7	(0.1)	2.8	(0.0)
Slowenien	2.7	(0.0)	3.1	(0.0)	2.8	(0.0)
Österreich	2.6	(0.0)	3.3	(0.1)	2.4	(0.0)
Ungarn	2.5	(0.0)	3.2	(0.1)	2.7	(0.0)
OECD-Durchschnitt	3.6	(0.0)	3.3	(0.0)	3.6	(0.0)

Im OECD-Durchschnitt werden jeweils etwa dreieinhalb Stunden in der Woche dem Unterricht in den Fächern Mathematik, Naturwissenschaften oder in der Testsprache gewidmet. In Deutschland liegt die Dauer der mit Mathematik- und Deutschunterricht verbrachten Zeit leicht unterhalb des internationalen Durchschnitts, während die Naturwissenschaften mit 4.2 Stunden pro Woche überdurchschnittlich viel Raum erhalten. Vergleicht man Tabelle 5.5 mit den in anderen Kapiteln angegebenen Kompetenzwerten der Schülerinnen und Schüler, so wird deutlich, dass die Sortierung der Staaten nicht ihrer Reihenfolge in der mittleren mathematischen Kompetenz entspricht. Dies ist nicht verwunderlich, weil PISA die kumulativ erworbenen Kompetenzen misst; man müsste also die Lernzeiten bis zum Testzeitpunkt zusammenrechnen (Baumert, Bos & Watermann, 2000). Innerhalb Deutschlands hat sich gegenüber PISA 2003 die Zahl der wöchentlichen Unterrichtseinheiten für Mathematik nicht verändert. Allerdings gibt die PISA-Kohorte 2012 an, sich 15 Minuten pro Woche mehr mit Mathematikunterricht zu beschäftigen als die Fünfzehnjährigen in PISA 2003. Dieser Zuwachs an Lernzeit in Mathematik ist statistisch signifikant.

Auch die Zeit, welche die Schülerinnen und Schüler der OECD-Staaten – neben dem üblichen Curriculum – mit Zusatzunterricht in Mathematik verbringen, hängt auf der Staatenebene *nicht* mit dem Mittelwert mathematischer Kompetenz zusammen. Allerdings wird auf der Schulebene – über die Staaten hinweg – deutlich, dass oft genau diejenigen Schulen zusätzlichen Mathematikunterricht anbieten, deren Schülerschaft im Durchschnitt eher niedrige Werte auf der mathematischen Kompetenzskala erzielt. In Deutschland bieten 63 Prozent der Schulen zusätzlichen Mathematikunterricht an, der über das reguläre Stundenkontingent hinausgeht. Dieser Anteil entspricht dem OECD-Durchschnitt von knapp 66 Prozent. Die Zeit, die von den Fünfzehnjährigen außerhalb der Schule mit Mathematiklernen verbracht wird, kann sehr unterschiedlich gestaltet sein. Es kann sich beispielsweise um Nachhilfeunterricht handeln, der teilweise kostenlos angeboten wird (z. B. in Form von Tutorien in der Schule) oder aber auch um bezahlte, institutionell angebotene Nachhilfe. In den OECD-Staaten entfällt auf solche Lernzeiten in einer normalen Schulwoche im Mittel kaum Zeit. Durchschnittlich verbringen die Fünfzehnjährigen eine Dreiviertelstunde pro Woche in Tutorien oder ähnlichem (Deutschland: eine halbe Stunde) und eine halbe Stunde mit bezahlter Nachhilfe (Deutschland: ebenfalls eine halbe Stunde).

Insgesamt lässt sich in Bezug auf die Lernzeit festhalten, dass es sowohl unter den in Mathematik besonders leistungsstarken als auch unter den schwächer abschneidenden Staaten solche gibt, die vergleichsweise viele Stunden in der Woche für Mathematikunterricht vorsehen und solche, die eher wenig Zeit dafür verwenden. Es bedarf einer genaueren Bestimmung, ob zusätzliche Stunden in Mathematik eher als Stütz- und Förderkurse für schwache Schülerinnen und Schüler eingesetzt werden oder als *Enrichment* für besonders leistungsstarke Jugendliche, und aus methodischer Sicht letztlich eines Längsschnittdesigns, um dem Zusammenhang zwischen der Häufigkeit und Dauer der Beschäftigung mit Mathematik und der mittleren mathematischen Kompetenz genauer nachgehen zu können.

5.2 Die Einzelschule als Ort des Lernens

Neben den in den vorherigen Abschnitten beschriebenen Merkmalen der organisatorischen und institutionellen Voraussetzungen schulischen Lernens ist für Politik und Öffentlichkeit stets auch von Interesse, wie sich die einzelnen Schulen in Deutschland und anderen Staaten voneinander unterscheiden und welche Merkmale sie aufweisen. In den nächsten Abschnitten geht es um die *Klassengröße*, um *Schulentwicklungsmaßnahmen* und um das *Schulklima* – Merkmale, die für einzelne Schulen einen wesentlichen Faktor für die Gestaltung der Lernumwelt darstellen.

5.2.1 Klassengröße

Ein Aspekt, der vielfach mit den Bedingungen für erfolgreiches Lernen in Verbindung gebracht wird, ist die Größe von Schulklassen. In Deutschland unterscheidet sich die Klassengröße in Abhängigkeit von der Schulart, der Schulstufe und der einzelnen Schule mit ihren spezifischen Gegebenheiten. Die Klassengröße – und damit die Größe einer Lerngruppe – kann Lernen und das Unterrichtsgeschehen auf vielfältige Weise beeinflussen (Yeh, 2009). Intuitiv scheinen kleinere Klassen besser dazu geeignet sein, das Lernen individuell gezielt zu fördern und anzuleiten. In großen Klassen sollte es schwerer fallen, Störungen oder Unterbrechungen des Unterrichts zu vermeiden als in kleineren Klassen. Diese verbreiteten Annahmen werden allerdings durch die vorliegende empirische Evidenz nicht systematisch gestützt (Arnhold, 2005; Hattie, 2008). Insbesondere eignen sich die Befunde aus *Large Scale Assessments* nicht zur Unterstützung der Auffassung, dass kleinere Klassen höheren Lernerfolg bedeuten (Bromme, Prenzel & Jäger, 2011). Ein Grund dafür liegt darin, dass die Unterrichtsmuster in großen oder kleinen Klassen sich zumeist nicht unterscheiden, die Lehrkräfte also die pädagogischen Möglichkeiten kleinerer Klassen nicht wahrnehmen und ausschöpfen (Lankes & Carstensen, 2010). Wo dies jedoch der Fall ist, wie etwa im Englischunterricht, kann die Klassengröße durchaus einen Effekt haben (Helmke et al., 2008).

Wenn man über Schulklassen oder Klassengrößen spricht, wird es international teilweise schwierig, Vergleichbarkeit herzustellen. In Deutschland ist eine Schulklasse in der Regel eine Gruppe von Schülerinnen und Schülern, die für mehrere Schuljahre einen gemeinsamen Stundenplan hat, von einer Klassenleitung betreut und möglicherweise von den gleichen Lehrkräften unterrichtet wird. Sobald in Schulsystemen oder Schularten dazu übergegangen wird, innerhalb der Schule differenzierte Lerngruppen in verschiedenen Fächern zu unterrichten, ist das Konzept der Schulklasse ein anderes. Deshalb ist ein Vergleich allein anhand der Zahl von Schülerinnen und Schülern, die in den PISA-Befragungen als Klasse bezeichnet werden, mit Vorsicht zu interpretieren. Tabelle 5.6 gibt einen Überblick über die durchschnittlichen Klassengrößen in der Sekundarstufe I in allen OECD-Staaten.

Tabelle 5.6: Durchschnittliche Klassengröße in der Sekundarstufe I

Staat	Durchschnittliche Klassengröße in der Sekundarstufe I	
	<i>M</i>	(<i>SE</i>)
Australien	23	(0.2)
Belgien	16	(0.4)
Chile	30	(1.5)
Dänemark	20	(0.2)
Deutschland	24	(0.2)
Estland	21	(0.2)
Finnland	18	(0.2)
Frankreich	23	(0.3)
Griechenland	21	(0.7)
Irland	24	(0.2)
Island	19	(0.1)
Israel	29	(0.4)
Italien	19	(0.5)
Japan	m	m
Kanada	24	(0.3)
Korea	31	(1.0)
Luxemburg	20	(0.1)
Mexiko	28	(0.5)
Neuseeland	24	(0.6)
Niederlande	23	(0.3)
Norwegen	23	(0.2)
Österreich	23	(0.2)
Polen	23	(0.2)
Portugal	20	(0.3)
Schweden	21	(0.3)
Schweiz	19	(0.2)
Slowakische Republik	19	(0.4)
Slowenien	14	(1.0)
Spanien	22	(0.1)
Tschechische Republik	21	(0.3)
Türkei	27	(1.2)
Ungarn	20	(0.9)
Vereinigte Staaten	22	(0.7)
Vereinigtes Königreich	m	m
OECD-Durchschnitt	24	(0.1)

Anmerkung: m: fehlende Angaben.

Im OECD-Durchschnitt umfasst die Klassengröße 24 Schülerinnen und Schüler. Diesem Durchschnitt entspricht der Mittelwert für Deutschland exakt. Die durchschnittliche Klassengröße hat sich dementsprechend seit PISA 2003 nicht verändert (Drechsel & Senkbeil, 2004). Allerdings ist die Streubreite der Klassengrößen bei den OECD-Staaten in PISA 2012 beachtlich. Betrachtet man Nachbarstaaten Deutschlands in Europa, dann findet man in den Niederlanden, Österreich und Polen ähnlich durchschnittliche Klassengrößen wie in Deutschland. Auch in Mathematik leistungsstarke Staaten in Übersee (Australien, Neuseeland, Kanada) unterrichten typischerweise Klassengrößen wie Deutschland. Aber es finden sich auch in der Nachbarschaft zu Deutschland deutlich andere Konstellationen: Die durchschnittlichen Klassengrößen in Belgien liegen bei 16 und in der Schweiz bei 19 Schülerinnen und Schülern.

5.2.2 Schulentwicklung und Qualitätssicherung

PISA 2012 untersucht auf der Basis von Angaben der Schulleitungen den Einsatz von Instrumenten zur Qualitätssicherung und Schulentwicklung. Zu den Maßnahmen der Qualitätssicherung gehört unter anderem die systematische Erfassung von Daten, wie etwa zur Anwesenheit von Schülerinnen und Schülern und Lehrkräften, zu Prüfungsergebnissen oder zum Besuch von Fortbildungsmaßnahmen. Ein wichtiger Punkt ist auch die Nutzung interner Evaluation beziehungsweise Selbstevaluation sowie externer Evaluation. Die Möglichkeit eines Mentorats für Lehrkräfte, Gespräche mit Experten für Schulentwicklung und die Umsetzung schulinterner standardisierter Vereinbarungen für den Mathematikunterricht (z. B. in Form eines Schulcurriculums) sind weitere Möglichkeiten, wie Schulen an ihrer Entwicklung und Qualität arbeiten können. In Tabelle 5.7 ist für diese sechs Instrumente abgebildet, wie hoch der prozentuale Anteil an Schülerinnen und Schülern ist, deren Schulleitung angab, sie einzusetzen.

Tabelle 5.7: Instrumente der Qualitätssicherung und -entwicklung an Schulen

Staat	Prozentualer Anteil von Schülerinnen und Schülern, deren Schulleitung angab, Instrumente zur Qualitätssicherung und -entwicklung einzusetzen											
	Systematische Erfassung von Daten, z. B. Anwesenheit, Prüfungsergebnisse		Interne Evaluation/Selbstevaluation		Externe Evaluation		Mentorat für Lehrkräfte		Experten-gespräche zur Schul-entwicklung		Umsetzung standardisierter Vereinbarun-gen für den Mathematik-unterricht	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)
Australien	98.1	(0.4)	94.5	(0.9)	69.9	(1.7)	92.5	(1.0)	72.2	(1.8)	76.6	(1.7)
Belgien	76.8	(2.3)	79.5	(2.5)	69.2	(2.8)	72.2	(2.5)	40.1	(3.3)	42.0	(2.6)
Chile	86.8	(2.4)	89.9	(2.6)	55.3	(3.8)	20.9	(3.2)	40.3	(3.9)	49.5	(3.7)
Dänemark	80.2	(3.2)	87.6	(2.4)	58.3	(3.7)	51.7	(3.5)	49.7	(3.2)	23.9	(2.8)
Deutschland	76.8	(3.0)	73.9	(3.0)	60.0	(3.4)	32.9	(3.4)	19.2	(2.6)	55.1	(3.8)
Estland	95.5	(1.2)	99.4	(0.1)	77.1	(2.3)	79.9	(2.4)	39.2	(2.9)	88.0	(1.9)
Finnland	74.0	(2.9)	95.9	(1.1)	51.4	(3.0)	55.2	(3.5)	10.3	(2.0)	63.2	(2.6)
Frankreich	74.9	(2.8)	60.8	(3.7)	51.9	(3.9)	17.2	(2.6)	20.7	(3.1)	43.9	(3.4)
Griechenland	68.5	(3.4)	32.5	(3.9)	5.7	(1.9)	87.0	(2.3)	76.7	(3.2)	69.9	(3.7)
Irland	89.4	(2.5)	82.9	(3.0)	81.8	(3.1)	64.3	(3.8)	52.9	(4.4)	81.4	(3.3)
Island	95.0	(0.1)	99.3	(0.1)	79.4	(0.2)	19.3	(0.2)	46.1	(0.2)	46.6	(0.2)
Israel	95.8	(1.2)	81.8	(3.3)	60.0	(3.4)	94.1	(1.7)	54.0	(3.8)	86.7	(2.7)
Italien	52.2	(2.0)	76.1	(2.0)	34.0	(2.2)	77.5	(1.8)	23.0	(1.7)	56.5	(1.9)
Japan	53.7	(3.8)	96.2	(1.5)	77.3	(3.1)	87.9	(2.4)	4.8	(1.5)	38.1	(3.3)
Kanada	89.8	(1.1)	80.9	(1.7)	62.0	(2.1)	86.0	(1.5)	68.8	(1.5)	80.1	(1.9)
Korea	93.7	(1.9)	97.3	(1.4)	78.6	(3.0)	87.8	(2.9)	59.3	(3.8)	65.0	(4.1)
Luxemburg	70.9	(0.1)	75.5	(0.1)	40.4	(0.1)	64.8	(0.1)	41.7	(0.1)	59.9	(0.1)
Mexiko	94.3	(0.8)	93.9	(0.8)	74.7	(1.7)	53.9	(1.9)	52.3	(1.4)	67.9	(1.5)
Neuseeland	98.1	(0.7)	99.7	(0.3)	89.0	(2.2)	97.2	(1.2)	63.4	(3.8)	80.8	(2.6)
Niederlande	99.1	(0.8)	91.4	(2.2)	81.2	(3.3)	97.5	(1.2)	46.7	(4.9)	46.8	(4.5)
Norwegen	83.7	(2.8)	61.1	(3.7)	52.5	(3.9)	69.7	(3.7)	33.2	(3.4)	28.9	(3.4)
Österreich	75.1	(3.5)	86.5	(2.7)	20.3	(2.9)	88.5	(2.8)	54.9	(4.2)	61.5	(3.3)
Polen	99.2	(0.3)	97.4	(1.2)	78.6	(3.4)	86.6	(2.2)	39.4	(4.0)	81.8	(3.2)
Portugal	96.5	(1.0)	97.6	(1.3)	85.5	(2.8)	77.8	(3.7)	28.9	(3.8)	74.6	(3.7)
Schweden	95.2	(1.6)	89.9	(2.6)	65.1	(3.6)	68.2	(3.5)	31.8	(3.3)	29.5	(3.1)
Schweiz	63.0	(3.1)	84.3	(2.4)	62.8	(2.2)	71.0	(3.2)	27.5	(3.2)	53.6	(2.7)
Slowakische Republik	93.4	(1.8)	94.5	(1.5)	37.7	(3.4)	87.9	(2.9)	53.7	(3.8)	61.2	(3.8)
Slowenien	86.4	(0.3)	92.2	(0.6)	32.4	(0.8)	67.2	(0.7)	41.0	(0.8)	67.1	(0.6)
Spanien	92.0	(1.4)	82.2	(1.7)	78.5	(2.1)	26.1	(1.9)	27.2	(2.4)	38.2	(3.0)
Tschechische Republik	84.7	(2.7)	97.9	(1.1)	62.9	(3.8)	95.9	(0.8)	27.3	(3.0)	90.2	(2.4)
Türkei	96.3	(1.8)	98.6	(1.3)	79.5	(3.9)	86.3	(2.2)	59.7	(3.6)	74.4	(3.2)
Ungarn	79.9	(3.5)	96.9	(1.3)	57.4	(3.8)	71.5	(3.5)	17.3	(3.2)	69.4	(4.0)
Vereinigte Staaten	98.1	(1.1)	92.5	(2.3)	86.1	(3.2)	98.4	(1.0)	73.5	(3.8)	88.1	(3.0)
Vereinigtes Königreich	99.6	(0.2)	100.0	(0.0)	91.4	(2.0)	96.4	(0.9)	80.2	(2.3)	74.3	(3.1)
OECD-Durchschnitt	85.5	(0.4)	87.1	(0.4)	63.2	(0.5)	71.5	(0.4)	43.4	(0.5)	62.2	(0.5)

Aus der Tabelle 5.7 wird deutlich, dass nur eines dieser Instrumente in Deutschland ähnlich oft eingesetzt wird wie im OECD-Durchschnitt: die externe Evaluierung (etwa 60 Prozent). Hierin spiegelt sich die Tatsache, dass Deutschland im vergangenen Jahrzehnt Schulinspektionen und Vergleichsarbeiten eingeführt hat, die in vielen anderen Staaten bereits eine längere Tradition haben (vgl. Altrichter & Maag Merki, 2010). Deutlich häufiger wird OECD-weit von *interner* Evaluierung der Schule berichtet (87 Prozent). In den angelsächsischen Staaten und Ostasien (Japan, Korea) ist diese praktisch flächendeckend eingeführt. In Deutschland hingegen ist interne Evaluation lediglich zu 74 Prozent üblich, aber auch hier kommt interne Evaluation häufiger als externe Evaluation zum Einsatz. Eine deutsche Zusatzstudie zu PISA 2009 kam zu dem Ergebnis, dass Schulen, die schon an PISA 2000 teilgenommen und zwischendurch in eigener Regie, nach selbst ausgewählten Zielen, eine Evaluation vorgenommen hatten, neun Jahre später nicht nur ihr Leistungsniveau (mittlere Lesekompetenz), sondern auch das Schulklima (Qualität der Schüler-Lehrer-Beziehungen) weiter entwickelt hatten als Schulen ohne interne Evaluierung (Bischof et al., 2013). Selbstevaluierung ist häufig verbunden mit der Sichtung statistischer Daten, beispielsweise zu Tests, Prüfungsergebnissen oder Fehltagen von Schülerinnen und Schülern. Der Schulbefragung bei PISA 2012 zufolge findet eine solche systematische Nutzung von Daten – wie interne Evaluation überhaupt – an deutschen Schulen seltener statt als im OECD-Durchschnitt.

Noch größer ist die Diskrepanz zu etablierten Praktiken in anderen OECD-Staaten, wenn es um die aktive Beratung für die Qualitätsentwicklung an Schulen geht. PISA 2012 unterscheidet hier zwischen Gesprächen mit externen Beratern über Fragen der Schulentwicklung über mindestens ein halbes Jahr hinweg sowie Beratungsangeboten (Mentorat durch erfahrene Lehrkräfte) für Lehrerinnen und Lehrer. Beides kommt in Deutschland nur etwa halb so häufig vor wie im OECD-Durchschnitt. Das Mentorat hat unter anderem in Ostasien (Japan, Korea) und den angelsächsischen Staaten große Bedeutung. Angelsächsische Schulen berichten zudem mehrheitlich von Expertengesprächen zur Schulentwicklung. In Deutschland besucht nur jede beziehungsweise jeder fünfte Jugendliche eine Schule, welche Schulentwicklungsberatung nutzt. In unseren Nachbarstaaten, mit Ausnahme Frankreichs, ist der Anteil höher.

Ein fachspezifischer Indikator ergibt sich aus der Frage, ob die Schule eine eigene Strategie für die schulinterne Qualitätsentwicklung des Mathematikunterrichts entwickelt hat. Der PISA-Schulfragebogen präzisiert, dass dazu beispielsweise ein Schulcurriculum, der Austausch von Unterrichtsmaterialien sowie begleitende Weiterbildungsmaßnahmen für Fachlehrkräfte gehören. Solche Maßnahmen wurden für Deutschland im Zuge der Implementation von Bildungsstandards vorgeschlagen (Klieme et al., 2003). Im Jahr 2012 lagen die deutschsprachigen Staaten bei der Umsetzung im OECD-Durchschnitt, die angelsächsischen Staaten darüber. In Frankreich, Belgien und den Niederlanden waren derartige schulinterne Vereinbarungen seltener anzutreffen, in dänischen, schwedischen und norwegischen Schulen sogar besonders selten.

Als Fazit lässt sich festhalten, dass Deutschland bei externer Qualitätssicherung und fachspezifischer Qualitätsentwicklung im international üblichen Bereich liegt, nicht

jedoch bei interner Evaluierung und bei Beratungsleistungen für Schulen und Lehrkräfte.

Im Rahmen der Qualitätsentwicklung spielt das Führungshandeln der Schulleitungen eine wichtige Rolle. Auch diesbezüglich wurde bei PISA 2012 befragt, allerdings – in Ermangelung eines internationalen Lehrerfragebogens – nur im Schulfragebogen, also aus Sicht der Schulleitungen selbst. Auf der Basis dieser Perspektive lassen sich Profile des Führungshandelns skizzieren, die unter anderem als Grundlage für vertiefende Untersuchungen dienen könnten. Beispielsweise wird unter den OECD-Staaten recht deutlich, dass es große Unterschiede in der Ausprägung *instruktionaler* Führung gibt. Instruktionale Führung bedeutet, dass die Schulleitung sich um Leitideen für den Unterricht und die Qualität des Lehrerhandelns kümmert und verantwortlich dafür ist, beispielsweise eine positive Lernkultur zu gestalten oder ein transparentes Regelsystem zu etablieren. An Schulen in Deutschland wird diese Art der Führung nach Angaben der Schulleitungen häufiger genutzt als im OECD-Durchschnitt, jedoch deutlich seltener als etwa in den angelsächsischen Staaten (Vereinigte Staaten, Vereinigtes Königreich, Australien und Kanada). Deutlich seltener als im OECD-Mittel nutzen unsere Nachbarstaaten Frankreich und die Schweiz diesen Führungsstil und mit Abstand am seltensten wird instruktionale Führung in Japan gewählt.

5.2.3 Schülerverhalten und Schulklima

Im institutionellen Rahmen verfügen Schulen neben den Aspekten der Qualitätssicherung auch über Gestaltungsspielräume, um ihrem Bildungsauftrag nachzukommen und sich als Lern- und Lebensort zu profilieren (Fend, 1988; Fend, 2006). Das Schulklima wird als Kultur des gemeinsamen Umgehens der verschiedenen Mitglieder einer Schule geprägt. Ein Schulklima, das als angenehm wahrgenommen wird, hängt mit einer ganzen Reihe von wünschenswerten psychosozialen Entwicklungen und häufig auch mit dem Lernerfolg zusammen (Grob, 2007; Jäger, 2012). Eine besondere Herausforderung für Schulen sind daher Beeinträchtigungen des Schulklimas und damit schlussendlich auch des Lernens.

Beeinträchtigungen des Lernens

In PISA 2012 wurden die Schulleitungen aller teilnehmenden Schulen in einem Fragebogen befragt, inwieweit sie an ihrer Schule das Lernen durch unerwünschtes Schülerverhalten beeinträchtigt sehen. Dazu gehören beispielsweise unentschuldigtes Fehlen im Unterricht (während des kompletten Schultages), Schwänzen einzelner Stunden, Zu spät-Kommen, mangelnder Respekt gegenüber Lehrkräften, Unterrichtsstörungen, Alkohol- und Drogenkonsum oder auch das Einschüchtern oder Schikanieren von Mitschülerinnen und Mitschülern.

Insgesamt scheinen in den PISA-Teilnehmerstaaten Schulschwänzen und Unterrichtsstörungen aus der Sicht der Schulleitungen größere Probleme für das Lernen mit sich zu bringen als der Konsum von Alkohol, Drogen oder Einschüchterung und mangelnder Respekt gegenüber Lehrkräften. Generell sind Befragungen bezüglich Themen, die dem Befragten möglicherweise unangenehm sein könnten, nicht unproblematisch. Es ist davon auszugehen, dass in den PISA-Staaten große Unterschiede im Antwortverhalten und auch in der Einschätzung von Belastungen des Schulklimas existieren. Der Punkt, ab dem ein Schulleiter oder eine Schulleiterin angibt, dass durch Dinge wie Unterrichtsstörungen oder fehlenden Respekt das Lernen an ihrer Schule beeinträchtigt wird, ist sicher sowohl individuell als auch kulturell gesehen unterschiedlich. Insofern sind die Vergleiche bei dieser Thematik mit Vorsicht anzustellen.

Was sich über die OECD-Staaten hierzu berichten lässt, ist, dass aus der Sicht der Schulleitungen in der überwiegenden Mehrheit der Schulen das Lernen „nicht“ oder „nur wenig“ unter den genannten Faktoren leidet. In den OECD-Ländern besuchen 94 Prozent der Schülerinnen und Schüler eine Schule, bei der die Schulleitung keine oder sehr wenige Probleme durch Alkohol und Drogen berichtet.

An neun von zehn Schulen in den OECD-Staaten sieht die Schulleitung keine oder kaum Probleme durch das Einschüchtern oder Schikanieren von Mitschülerinnen und Mitschülern. Es gibt jedoch auch einige Staaten, in denen die Schulleitungen in einem größeren Ausmaß (mehr als 20 Prozent) von solchen Beeinträchtigungen berichten, und zwar Staaten, die durchaus als erfolgreich im internationalen Leistungsvergleich gelten: Finnland, die Niederlande und Korea sowie unter den Partnerstaaten Shanghai (China).

Unterrichtsstörungen sind an gut zwei Dritteln der Schulen in OECD-Staaten kein oder nur ein marginales Problem aus der Sicht der Schulleitungen. In Finnland und Portugal jedoch besucht mehr als die Hälfte der Schülerinnen und Schüler Schulen, bei denen nach Auskunft der Schulleitungen das Lernen „bis zu einem gewissen Grad“ oder gar „sehr“ durch Unterrichtsstörungen beeinträchtigt wird.

Schulversäumnisse

Schule findet statt, wenn die Beteiligten anwesend sind. Vom Unterricht können nur diejenigen Schülerinnen und Schüler profitieren, die ihn auch besuchen. Die Schulpflicht soll sicherstellen, dass alle Kinder und Jugendlichen vergleichbare Lernbedingungen vorfinden und damit die Chance haben, als relevant betrachtete Bildungsziele zu erreichen. Wenn Schülerinnen und Schüler dem Unterricht fernbleiben, verpassen sie die Gelegenheit, sich mit solchen Themen und mit ihrer schulischen Umwelt auseinanderzusetzen und zu lernen.

Schulversäumnisse, sei es in einzelnen Fächern oder Schulstunden oder auch für ganze Schultage, gehören zu den Ereignissen, die über die OECD-Staaten hinweg von den Schulleitungen als besonders problematisch für gelingendes Lernen angesehen werden. Etwa jeder dritte Jugendliche besucht eine Schule, deren Schulleitung angibt, dass das Lernen an ihrer Schule „bis zu einem gewissen Grad“ oder sogar „sehr“ durch das

Schwänzen einzelner Schulstunden oder durch zu spät zum Unterricht kommende Schülerinnen und Schüler beeinträchtigt wird. In Slowenien, Kanada, der Slowakischen Republik und der Türkei gilt dies für mindestens die Hälfte der Jugendlichen. Beim Zu-spät-Kommen zeichnet sich dasselbe für Finnland und Chile ab sowie wiederum für Kanada. Auch das unentschuldigte Fehlen ganzer Schultage ist OECD-weit in etwa einem Drittel der Schulen (gewichtet nach Anzahl der Schüler) ein Problem, unter dem das Lernen zumindest bis zu einem gewissen Grad leidet.

In Tabelle 5.8 ist zusammengefasst, inwieweit Schulleiterinnen und Schulleiter in den OECD-Staaten das Lernen an ihrer Schule durch Schülerabsenzen (sowohl einzelne Stunden als auch ganze Tage) und zu spät zum Unterricht kommende Schülerinnen und Schüler beeinträchtigt sehen.

In Deutschland besucht jeder sechste Jugendliche eine Schule, an der nach Auskunft der Schulleitung das Lernen aufgrund von Schülerabsenzen zumindest ein Stück weit behindert wird. Der OECD-Durchschnitt zu dieser Problematik liegt bei mehr als 30 Prozent. Aus der Sicht der Schulleitungen in Deutschland sind Schülerabsenzen also ein vergleichsweise geringes Problem. Ein ähnlich geringes oder noch geringeres Problem durch schwänzende Jugendliche sehen Schulleitungen im Vereinigten Königreich, der Schweiz, in Luxemburg, Korea, in Japan, Island und Irland.

Über die OECD-Staaten hinweg scheinen Schülerinnen und Schüler, die zu spät im Unterricht erscheinen, ein Problem für schulisches Lernen darzustellen, das in seinem Ausmaß dem der Schülerabsenzen entspricht. In Deutschland allerdings wird das Zu-spät-Kommen im Vergleich zu kompletten Absenzen von sehr viel mehr Schulleitungen als relevante Beeinträchtigung bewertet. Dieser Störfaktor gilt in etwa einem Drittel unseres Schulsystems als problematisch für das schulische Lernen. In Chile, Finnland und Kanada gilt das sogar für etwa die Hälfte der Schulen.

Neben der Sichtweise der Schulleiterinnen und Schulleiter wurde in PISA 2012 auch die Perspektive der Schülerinnen und Schüler auf Schulversäumnisse erfasst. Dazu gaben die Fünfzehnjährigen an, ob und wie oft sie in den letzten beiden vollständigen Schulwochen vor dem PISA-Test einzelne Unterrichtsstunden geschwänzt, ganze Schultage unentschuldigt gefehlt haben oder zu spät zum Unterricht erschienen sind. Seit PISA 2000 wird den Schülerinnen und Schülern diese Frage gestellt.

Bei der Erhebung von Schulversäumnissen hängen die berichteten Quoten davon ab, wonach genau gefragt wird und auf welchen Zeitraum sich die Angaben der Schülerinnen und Schüler beziehen sollen. Einerseits wissen ausschließlich die Jugendlichen selbst, wann sie geschwänzt haben und wann sie tatsächlich einen gesetzlich vorgesehenen Grund für ein Unterrichtsversäumnis hatten. Andererseits besteht bei Schüler-selbstberichten auch die Gefahr einer Über- oder Untertreibung von erfragten Verhaltensweisen. Im Vergleich zu anderen Studien über Schulversäumnisse in Deutschland (etwa Ricking, Schulze & Wittrock, 2009 oder Weiß, 2007) ist der erfragte Zeitraum von zwei vollen Schulwochen stark eingegrenzt. Fragt man die Schülerinnen und Schüler nach unentschuldigten Absenzen im vergangenen Schuljahr oder im aktuellen Halbjahr, ist der Zeitraum, in dem möglicherweise geschwänzt wurde, sehr viel größer und

Tabelle 5.8: Beeinträchtigungen des Lernens durch Schülerabsenzen oder Zu-spät-Kommen

Staat	Beeinträchtigung des Lernens durch Schülerabsenzen				Beeinträchtigung des Lernens durch zu spät kommende Schülerinnen und Schüler			
	Nicht oder wenig		Bis zu einem gewissen Grad oder sehr		Nicht oder wenig		Bis zu einem gewissen Grad oder sehr	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)
Australien	75.4	(3.7)	24.6	(2.2)	65.5	(2.6)	34.5	(2.1)
Belgien	79.9	(5.9)	20.1	(3.1)	68.4	(5.1)	31.6	(3.7)
Chile	79.3	(6.4)	20.7	(4.4)	49.0	(5.9)	51.0	(6.4)
Dänemark	78.6	(5.5)	21.4	(3.7)	73.8	(6.3)	26.2	(4.0)
Deutschland	83.5	(5.5)	16.5	(3.0)	69.1	(4.8)	30.9	(3.8)
Estland	62.8	(4.5)	37.2	(3.9)	70.0	(5.1)	30.0	(3.2)
Finnland	65.0	(4.9)	35.0	(3.6)	49.0	(4.3)	51.0	(4.0)
Frankreich	71.7	(5.9)	28.3	(4.0)	73.2	(5.5)	26.8	(4.0)
Griechenland	77.9	(6.0)	22.1	(4.5)	72.9	(5.8)	27.1	(5.2)
Irland	85.4	(6.7)	14.6	(3.8)	75.2	(6.1)	24.8	(4.4)
Island	91.6	(0.5)	8.4	(0.2)	83.5	(0.5)	16.5	(0.2)
Israel	57.9	(5.9)	42.1	(5.8)	62.6	(6.2)	37.4	(5.7)
Italien	63.4	(3.5)	36.6	(3.0)	61.4	(3.1)	38.6	(2.9)
Japan	90.0	(7.3)	10.0	(2.5)	62.7	(4.9)	37.3	(4.6)
Kanada	43.2	(3.3)	56.8	(4.0)	46.6	(2.6)	53.4	(3.6)
Korea	85.3	(7.9)	14.7	(4.1)	74.5	(7.4)	25.5	(4.8)
Luxemburg	88.4	(0.1)	11.6	(0.1)	72.5	(0.1)	27.5	(0.1)
Mexiko	67.3	(3.0)	32.7	(2.9)	67.0	(3.0)	33.0	(2.6)
Neuseeland	67.1	(5.5)	32.9	(5.3)	68.9	(6.0)	31.1	(5.5)
Niederlande	70.6	(4.7)	29.4	(4.6)	60.9	(5.5)	39.1	(5.3)
Norwegen	70.2	(5.8)	29.8	(4.3)	74.9	(5.3)	25.1	(3.5)
Österreich	59.2	(6.2)	40.8	(6.2)	60.0	(6.4)	40.0	(5.9)
Polen	59.8	(5.7)	40.2	(4.8)	73.5	(6.2)	26.5	(4.5)
Portugal	58.7	(6.3)	41.3	(6.5)	71.2	(7.9)	28.8	(5.4)
Schweden	59.9	(5.0)	40.1	(4.5)	69.5	(5.0)	30.5	(4.2)
Schweiz	82.5	(5.1)	17.5	(3.1)	80.2	(5.5)	19.8	(3.2)
Slowakische Republik	28.2	(4.0)	71.8	(5.8)	69.1	(6.6)	30.9	(5.0)
Slowenien	33.9	(0.8)	66.1	(0.9)	64.6	(0.8)	35.4	(0.7)
Spanien	74.6	(5.5)	25.4	(3.2)	84.0	(4.9)	16.0	(2.3)
Tschechische Republik	60.2	(6.5)	39.8	(5.2)	89.9	(7.3)	10.1	(2.6)
Türkei	45.9	(5.5)	54.1	(6.5)	57.0	(6.8)	43.0	(6.7)
Ungarn	78.5	(6.3)	21.5	(4.6)	67.2	(4.9)	32.8	(4.9)
Vereinigte Staaten	69.0	(6.4)	31.0	(4.8)	65.9	(6.3)	34.1	(5.5)
Vereinigtes Königreich	94.5	(6.3)	5.5	(1.6)	84.8	(5.1)	15.2	(2.3)
OECD-Durchschnitt	69.4	(0.9)	30.6	(0.7)	68.8	(0.9)	31.2	(0.7)

Anmerkung: Die in der Tabelle angegebenen Prozentwerte beziehen sich auf Schülerinnen und Schüler, deren Schulleitung angab, die genannten Beeinträchtigungen an der eigenen Schule vorzufinden.

damit auch die Wahrscheinlichkeit, dass die Jugendlichen zumindest einmal unerlaubt dem Unterricht fern geblieben sind. Allerdings besteht in diesem Fall auch die Gefahr von Erinnerungslücken.

Vor diesem Hintergrund sind die Anteile der Schülerinnen und Schüler, die im Rahmen von PISA 2012 angaben, in den letzten beiden vollen Schulwochen vor dem PISA-Test einzelne Stunden oder ganze Schultage unentschuldig gefehlt zu haben, relativ hoch: 14 Prozent der Fünfzehnjährigen in den OECD-Staaten gaben an, in diesem Zeitraum mindestens einen Schultag unentschuldig versäumt zu haben und 18 Prozent berichteten, dass sie einzelne Unterrichtsstunden geschwänzt haben. Absentismus ist also in den OECD-Staaten durchaus ein Thema. In Deutschland liegen diese prozentualen Anteile wesentlich niedriger. 5.2 Prozent der Fünfzehnjährigen haben mindestens einen ganzen Schultag lang unentschuldig gefehlt und 9.7 Prozent einzelne Stunden. Damit liegt Deutschland deutlich unter dem OECD-Durchschnitt.

Auch in Bezug auf das Zu-spät-Kommen zum Unterricht wurde die Perspektive der Schülerinnen und Schüler in PISA 2012 erfasst. Im OECD-Durchschnitt sind 35 Prozent der Schülerinnen und Schüler in den zwei letzten vollen Schulwochen vor dem PISA-Test mindestens einmal zu spät zum Unterricht gekommen. In Deutschland betraf dies 22.7 Prozent der Schülerinnen und Schüler, also deutlich weniger als im OECD-Mittel. Da nicht nach Gründen gefragt wurde, kann allerdings nicht ausgeschlossen werden, dass hierbei auch Verkehrsverhältnisse und Erreichbarkeit der Schulen eine Rolle spielen.

Gefühl der Zugehörigkeit zur eigenen Schule

Ein weiteres wesentliches Element des Schulklimas ist das Gefühl der Zugehörigkeit, das die Schülerinnen und Schüler gegenüber ihrer eigenen Schule empfinden. Dieses Gefühl der Zugehörigkeit ist ein Indikator für das emotionale und soziale Wohlbefinden und drückt beispielsweise aus, wie gerne sie jeden Morgen dort hingehen. In diesem Sinne spricht ein stark ausgeprägtes Zugehörigkeitsgefühl dafür, dass es einer Schule gelingt, für ihre Schülerinnen und Schüler ein Ort zu sein, an dem sie sich gerne aufhalten. Im Einzelnen erfasst der Index unter anderem, ob sich Schülerinnen und Schüler in ihrer Schule als Außenseiter fühlen, ob sie dort leicht Freunde finden, glücklich oder auch mit ihrer Schule insgesamt zufrieden sind. Tabelle 5.9 stellt eine Übersicht über die Ausprägung des Gefühls der Zugehörigkeit zur eigenen Schule in den OECD-Staaten dar.

Tabelle 5.9: Gefühl der Zugehörigkeit zur eigenen Schule in den OECD-Staaten

Staat	Gefühl der Zugehörigkeit zur Schule		Mathematische Kompetenz, in Quartile nach dem Gefühl der Zugehörigkeit unterteilt							
			Unteres Quartil		Zweites Quartil		Drittes Quartil		Oberes Quartil	
	M	(SE)	M	(SE)	M	(SE)	M	(SE)	M	(SE)
Australien	-0.15	(0.02)	483	(2.9)	503	(3.4)	513	(3.3)	519	(2.8)
Belgien	-0.05	(0.02)	501	(4.2)	526	(3.5)	532	(3.5)	529	(3.3)
Chile	0.14	(0.02)	420	(4.2)	417	(5.0)	427	(4.1)	428	(4.2)
Dänemark	-0.05	(0.02)	491	(3.6)	507	(3.9)	515	(3.7)	508	(3.5)
Deutschland	0.27	(0.02)	511	(5.0)	520	(4.7)	533	(5.1)	526	(4.9)
Estland	-0.33	(0.02)	510	(3.6)	519	(3.8)	523	(4.1)	530	(4.3)
Finnland	-0.22	(0.02)	510	(3.4)	528	(3.4)	526	(3.3)	527	(2.7)
Frankreich	-0.11	(0.02)	467	(4.6)	504	(4.2)	508	(4.2)	512	(5.2)
Griechenland	-0.14	(0.02)	444	(4.4)	460	(4.0)	459	(4.4)	459	(3.7)
Irland	-0.03	(0.02)	501	(4.1)	496	(3.5)	504	(4.1)	504	(3.5)
Island	0.36	(0.02)	476	(3.6)	500	(5.0)	500	(3.8)	509	(4.7)
Israel	0.41	(0.03)	460	(6.7)	487	(6.6)	478	(6.2)	469	(6.3)
Italien	-0.21	(0.01)	477	(2.6)	493	(2.9)	497	(2.9)	482	(2.9)
Japan	-0.16	(0.02)	530	(4.7)	541	(4.7)	540	(4.7)	538	(4.8)
Kanada	-0.09	(0.01)	510	(3.4)	519	(3.0)	528	(2.7)	525	(2.9)
Korea	-0.32	(0.02)	532	(5.9)	552	(5.4)	561	(5.1)	571	(7.1)
Luxemburg	0.20	(0.02)	466	(3.4)	488	(3.5)	499	(3.6)	506	(3.3)
Mexiko	0.10	(0.01)	406	(1.9)	410	(1.9)	420	(2.0)	423	(1.8)
Neuseeland	-0.14	(0.02)	490	(4.5)	496	(5.4)	511	(4.2)	502	(4.3)
Niederlande	-0.03	(0.02)	507	(5.7)	533	(5.2)	538	(4.5)	535	(5.2)
Norwegen	0.08	(0.02)	472	(5.3)	496	(4.7)	499	(4.3)	495	(4.1)
Österreich	0.55	(0.03)	488	(5.4)	510	(4.3)	518	(3.5)	517	(4.3)
Polen	-0.32	(0.02)	519	(4.7)	521	(4.8)	515	(4.7)	521	(6.0)
Portugal	0.03	(0.02)	468	(5.5)	488	(5.1)	501	(4.8)	502	(4.7)
Schweden	-0.04	(0.02)	465	(4.6)	485	(4.3)	491	(3.5)	489	(4.3)
Schweiz	0.43	(0.02)	510	(4.1)	533	(4.4)	541	(4.3)	542	(4.3)
Slowakische Republik	-0.31	(0.02)	467	(6.3)	486	(5.1)	489	(5.8)	496	(4.8)
Slowenien	-0.01	(0.02)	489	(3.3)	505	(4.3)	513	(4.2)	511	(4.7)
Spanien	0.41	(0.02)	472	(3.5)	492	(2.6)	495	(3.3)	486	(3.0)
Tschechische Republik	-0.36	(0.02)	485	(4.1)	507	(4.6)	512	(3.6)	513	(4.9)
Türkei	0.12	(0.02)	439	(6.5)	449	(4.8)	457	(5.9)	452	(6.3)
Ungarn	0.11	(0.02)	457	(4.8)	475	(4.8)	487	(4.6)	494	(5.0)
Vereinigte Staaten	-0.05	(0.02)	474	(5.2)	480	(4.6)	491	(4.9)	490	(4.5)
Vereinigtes Königreich	-0.02	(0.02)	484	(4.4)	501	(3.7)	499	(5.3)	505	(4.8)
OECD-Durchschnitt	0.0	0.0	482	(4.4)	498	(4.3)	504	(4.2)	503	(4.3)

In Tabelle 5.9 ist in den beiden Spalten ganz links die Ausprägung des individuellen Zugehörigkeitsgefühls zur eigenen Schule für die OECD-Staaten im Überblick dargestellt (*Sense of Belonging*). Dabei sind die Skalenwerte in der Tabelle so normiert, dass der OECD-Mittelwert den Wert 0 und die Standardabweichung den Wert 1 annimmt (z-Standardisierung). Die Angaben in der Tabelle sind folglich so zu lesen, dass ein positiver Wert eine Ausprägung des Zugehörigkeitsgefühls bedeutet, die oberhalb des OECD-Durchschnitts liegt, und negative Werte entsprechend eine Ausprägung, die niedriger ist als der OECD-Durchschnitt. Weiter rechts in der Tabelle ist dann angegeben, wie hoch die mathematische Kompetenz ist, wenn man die Schülerinnen und Schüler in einem Staat so in vier gleich große Gruppen (Quartile) einteilt, dass der Grad des Zugehörigkeitsgefühls in der eigenen Schule sehr niedrig (unterstes Quartil), niedrig (zweites Quartil), hoch (drittes Quartil) oder sehr hoch (oberstes Quartil) ist. Auf diese Weise lässt sich ein Eindruck gewinnen, ob und inwieweit Schülerinnen und Schüler, die sich in ihrer Schule emotional wohl und sozial integriert fühlen, höhere Kompetenzwerte in Mathematik erreichen als andere oder anders betrachtet: ob höhere Kompetenzwerte dazu führen können, Schule als angenehmer zu erleben.

Die Angaben der Schülerinnen und Schüler sind sicherlich durch bestimmte Antworttendenzen beeinflusst, wie sie aus der kulturvergleichenden Bildungsforschung bekannt sind (vgl. etwa Vieluf et al., 2013). Beschränkt man sich bei der Betrachtung des Gefühls der Zugehörigkeit daher auf Deutschland und seine Nachbarstaaten, so fällt auf, dass in den meisten dieser Staaten das Zugehörigkeitsgefühl signifikant geringer ausgeprägt ist als in Deutschland. Ausnahmen sind Österreich und die Schweiz, wo die Werte nochmals deutlich höher liegen als in Deutschland. Insgesamt scheinen sich Jugendliche im deutschen Sprach- und Kulturraum an ihren Schulen subjektiv wohler zu fühlen als Gleichaltrige in den meisten anderen Staaten, einschließlich der übrigen Nachbarstaaten, Skandinaviens sowie der ostasiatischen und der angelsächsischen Staaten.

Betrachtet man das von den Schülerinnen und Schülern angegebene Zugehörigkeitsgefühl über die Zeit, so zeigt sich zwischen PISA 2003 (als die Mathematik zum ersten Mal Schwerpunkt war) und PISA 2012 weder im Durchschnitt der OECD-Staaten noch in Deutschland ein Unterschied. Beide PISA-Kohorten geben also ein ähnlich hohes Zugehörigkeitsgefühl zu ihrer Schule an. In einigen Staaten ist jedoch eine signifikante Veränderung zu verzeichnen im Sinne eines signifikant höheren (z. B. Japan, Österreich, Türkei) beziehungsweise niedrigeren (z. B. Schweden, Australien) Zugehörigkeitsgefühls als in PISA 2003.

Die Befunde zum Gefühl der Zugehörigkeit zur eigenen Schule zeigen exemplarisch, dass PISA den Bildungs- und Entwicklungsstand von Jugendlichen sehr viel umfassender erschließt als dies eine Messung durch Leistungstests allein erlauben würde. Schulen müssen sowohl kognitive Kompetenzen vermitteln, wie dies in PISA als Lesekompetenz, mathematische und naturwissenschaftliche Kompetenz erfasst wird, als auch die psychosoziale Entwicklung der Kinder und Jugendlichen unterstützen.

Beide Ziele sind – jedenfalls in unserem Kulturkreis – nicht unabhängig voneinander: In den meisten Staaten geht ein höherer Grad an Zugehörigkeit, wie Tabelle 5.9

zeigt, mit besseren (mathematischen) Leistungen einher. Der Zusammenhang ist aber in der Regel nicht linear. So bleiben in den meisten Staaten Jugendliche, die ein sehr niedriges Zugehörigkeitsgefühl berichten (unteres Quartil), mit ihren Mathematikleistungen durchschnittlich 15 Punkte oder mehr unter den Leistungen von Gleichaltrigen, die mittlere Werte des Zugehörigkeitsgefühls berichten (zweites und drittes Quartil); diese Unterschiede sind statistisch signifikant. Dies kann darauf hinweisen, dass Leistungsprobleme und psychosoziale Probleme überzufällig häufig zusammen auftreten. Mittlere und höhere Leistungen variieren hingegen meist unabhängig von der emotionalen und sozialen Situation. Dieses Muster gilt unter den OECD-Staaten für fast alle europäischen Staaten (Ausnahmen: Estland, Irland, Polen), während in den außereuropäischen Staaten Leistung und Wohlbefinden generell kaum zusammenhängen (Ausnahmen: Australien, Korea).

5.3 Zusammenfassung und Diskussion

In diesem Kapitel wurde das Thema Schule aus zwei Perspektiven betrachtet: Aus der Sicht des institutionellen Systems und aus der Sicht der Einzelschule, die bestimmte Merkmale aufweist und Herausforderungen meistert. Dabei wurde zunächst ein Blick auf die Ebene der Gliederung verschiedener Schulsysteme in der OECD geworfen und festgestellt, dass mehrgliedrige Schulsysteme nicht automatisch eine niedrigere mathematische Kompetenz mit sich bringen, auch wenn dies in der Diskussion über Befunde aus PISA immer wieder behauptet wurde. Um die Schulstruktur und ihre Wirkungen angemessen abzubilden, sollten Schulen in mehrgliedrigen Schulsystemen auch innerhalb einer Schulart verglichen werden. In Deutschland bilden die Gymnasien eine Schulart mit relativ wenig Varianz zwischen den Schulen.

Weiterhin wurde ersichtlich, dass die fünfzehnjährigen Schülerinnen und Schüler in Deutschland auf niedrigere Klassenstufen verteilt sind als in zahlreichen anderen OECD-Staaten. Ein Grund dafür ist in der im internationalen Vergleich relativ hohen Häufigkeit von verzögerten Schullaufbahnen zu sehen. Diese Verzögerungen ergeben sich vor allem aus dem überdurchschnittlich häufigen Wiederholen von Klassenstufen.

Im internationalen Vergleich bemerkenswert sind die unterschiedlichen Zeitanteile, die für den Mathematikunterricht vorgesehen sind. Dieser Zeitanteil liegt in Deutschland mit 3.3 Zeitstunden etwas unter dem OECD-Durchschnitt von 3.6 Stunden. Dementsprechend verdienen die im OECD-Vergleich überdurchschnittlichen Mathematikleistungen Anerkennung.

Besondere Herausforderungen für Schulen sind Verhaltensweisen, die generell das Lernen beeinträchtigen können. In PISA 2012 wurden Probleme wie der Konsum von Alkohol und Drogen, Schulversäumnisse oder das Einschüchtern und Schikanieren von Mitschülerinnen und Mitschülern untersucht. Generell sehen die Schulleiterinnen und Schulleiter in den OECD-Staaten das Lernen an ihren Schulen nur wenig durch die erfassten Merkmale eingeschränkt. Das stärkste Problem aus ihrer Sicht sind Beeinträch-

tigungen durch Schulversäumnisse und Zu-spät-Kommen in den Unterricht, wie jeweils von etwa 30 Prozent der Schulleitungen angegeben wurde. 16.5 Prozent der Schulen in Deutschland müssen sich mit Zu-spät-Kommen und 30.9 Prozent mit unentschuldigtem Fehlen auseinandersetzen, was trotz des ähnlich hohen OECD-Durchschnitts beträchtlich ist – zumal sich der Befragungszeitraum lediglich auf zwei Unterrichtswochen bezieht. Wichtig ist, dass PISA mit solchen Faktoren, aber auch mit der Frage nach dem Gefühl der Zugehörigkeit zur Schule Erkenntnisse liefert, die das Verhalten und die psychosoziale Entwicklung von Jugendlichen berücksichtigen. Damit werden Aussagen zu Qualität und erzieherischen Ergebnissen der Schulen gemacht, die über eine bloße Messung von Leistungen hinausgehen.

Insgesamt erscheinen die Schulen in Deutschland im OECD-Vergleich als wenig auffällig und letztlich auch als wenig problematisch. Die Klassen sind durchschnittlich groß, unsere Schulen leiden nicht übermäßig an Beeinträchtigungen des Lernens, auch Absentismus hält sich in Grenzen. Deutschland ist einer der 21 OECD-Staaten, die Schülerinnen und Schülern im Alter von fünfzehn Jahren ein differenziertes Schulangebot unterbreiten. Zusammen mit Österreich erfolgt allerdings in Deutschland die Differenzierung sehr früh, am Ende der vierten Klassenstufe (mit Ausnahme der Länder Berlin und Brandenburg). Diese Strukturmerkmale wurden in der öffentlichen Diskussion um die Ergebnisse von PISA 2000 in Deutschland besonders kritisch betrachtet, weil sie mit dem damaligen schlechten Abschneiden und den starken sozialen Disparitäten im deutschen Schulsystem assoziiert wurden. Die sozialen Unterschiede in Bildungswegen und Bildungserfolg sind in der Tat – so der Erkenntnisstand der PISA-Studien – in stark gegliederten Systemen größer. Allerdings zeigen vertiefende Analysen zu einzelnen PISA-Erhebungsrunden und die Entwicklung bis PISA 2012, dass weder das mittlere Leistungsniveau der Jugendlichen eines Staates noch der nationale Leistungstrend durch die Strukturmerkmale erklärt werden kann. Für die Entwicklung von Schulsystemen ist möglicherweise weniger deren Gliederung relevant als eine kluge Politik mit Maßnahmen, welche beispielsweise helfen, Klassenwiederholungen zu vermeiden oder die Einzelschulen darin bestärken, ihre Qualität kritisch zu prüfen und in kollegialer Zusammenarbeit weiterzuentwickeln.

Literatur

- Alexander, K., Entwisle, D. & Dauber, S. (2003). *On the success of failure: a reassessment of the effects of retention in the early grades*. Cambridge: Cambridge University Press.
- Altrichter, H. & Maag Merki, K. (Hrsg.). (2010). *Handbuch neue Steuerung im Schulsystem* (Educational Governance, Bd. 7, 1. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Arnhold, G. (2005). *Kleine Klassen – große Klasse? Eine empirische Studie zur Bedeutung der Klassengröße für Schule und Unterricht*. Bad Heilbrunn: Klinkhardt.
- Baumert, J., Bos, W. & Watermann, R. (2000). Mathematische und naturwissenschaftliche Grundbildung im internationalen Vergleich. In J. Baumert, W. Bos & R. Lehmann

- (Hrsg.), *TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (Bd. 1, S. 135–197). Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W. et al. (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J. & Stanat, P. (2010). Internationale Schulleistungsvergleiche. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (4. Aufl., S. 324–335). Weinheim: Beltz.
- Bischof, L. M., Hochweber, J., Hartig, J. & Klieme, E. (2013). Schulentwicklung im Verlauf eines Jahrzehnts. Erste Ergebnisse des PISA-Schulpanels. In N. Jude & E. Klieme (Hrsg.), *PISA 2009 – Impulse für die Schul- und Unterrichtsforschung*. 59. Beiheft der Zeitschrift für Pädagogik (S. 172–199). Weinheim: Beltz.
- Boller, S., Möller, M. & Palowski, M. (2013). Wiederholen in der gymnasialen Oberstufe – wissenschaftliche Befunde und pädagogische Unterstützungsmöglichkeiten. In D. Bosse, F. Eberle & B. Schneider-Taylor (Hrsg.), *Standardisierung in der gymnasialen Oberstufe* (S. 175–188). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bromme, R., Prenzel, M. & Jäger, M. (2011). *Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. Eine Analyse von Anforderungen an die Darstellung, Interpretation und Rezeption empirischer Befunde. Gutachten für das BMBF*. Münster: Universität Münster.
- Diederich, J. & Tenorth, H.-E. (1997). *Theorie der Schule. Ein Studienbuch zu Geschichte, Funktionen und Gestaltung*. Berlin: Cornelsen Scriptor.
- Drechsel, B. & Senkbeil, M. (2004). Institutionelle und organisatorische Rahmenbedingungen von Schule und Unterricht. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs* (S. 284–291). Münster: Waxmann.
- Dreeben, R. & Barr, R. (1988). Classroom composition and the design of instruction. *Sociology of Education*, 61, 129–142.
- Duru-Bellat, M. & Suchaut, B. (2005). Organization and context, efficiency and equity of educational systems: What PISA tells us. *European Educational Research Journal*, 4 (3), 181–194.
- Ehmke, T., Drechsel, B. & Carstensen, C. H. (2008). Klassenwiederholen in PISA-I-Plus: Was lernen Sitzbleiber in Mathematik dazu? *Zeitschrift für Erziehungswissenschaft*, 11 (3), 368–387.
- Fend, H. (1988). Schulqualität. Die Wiederentdeckung der Schule als pädagogische Gestaltungsebene. *Neue Sammlung*, 28 (4), 537–547. Zugriff am 04.10.2013. Verfügbar unter <http://www.pedocs.de/volltexte/2009/1629>
- Fend, H. (2006). *Neue Theorie der Schule. Einführung in das Verstehen von Bildungssystemen* (Lehrbuch, 1. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gorard, S. & Smith, E. (2004). An international comparison of equity in education systems. *Comparative Education*, 40 (1), 15–28.
- Grob, U. (2007). Schulklima und politische Sozialisation. Hat das perzipierte soziale Klima einen nachhaltigen Einfluss auf die Entwicklung von politischem Interesse und Toleranz? *Zeitschrift für Pädagogik*, 53 (6), S. 774–799.
- Hattie, J. (2008). *Visible learning. A synthesis of meta-analyses relating to achievement*. London: Routledge.
- Helmke, A., Helmke, T., Schrader, F.-W., Wagner, W., Klieme, E., Nold, G. et al. (2008). Wirksamkeit des Englischunterrichts. In E. Klieme, W. Eichler, A. Helmke, R. Lehmann,

- G. Nold, H.-G. Rolff et al. (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 382–397). Weinheim: Beltz.
- Hong, G. & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, 27 (3), 205–224.
- Jäger, D. J. (2012). Schulklima, Selbstwirksamkeit und Arbeitszufriedenheit aus Sicht der Lehrpersonen und Schüler/-innen in Hessen und Bremen. In K. Maag Merki (Hrsg.), *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland* (S. 65–93). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Klemm, K. (2009). *Klassenwiederholungen – teuer und unwirksam*. Gütersloh: Bertelsmann Stiftung.
- Klieme, E. (2013). The role of large-scale assessments in research on educational effectiveness and school development. In M. von Davier, E. Gonzalez, I. Kirsch & K. Yamamoto (Hrsg.), *The role of international large-scale assessments: perspectives from technology, economy, and educational research* (S. 115–148). Dordrecht: Springer.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards. Expertise*. Bonn: BMBF.
- Klieme, E. & Steinert, B. (2008). Schulentwicklung im Längsschnitt. Ein Forschungsprogramm und erste explorative Analysen. In M. Prenzel & J. Baumert (Hrsg.), *Vertiefende Analysen zu PISA 2006. 10. Beiheft der Zeitschrift für Erziehungswissenschaft* (S. 221–238). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Klieme, E. & Vieluf, S. (2013). Schulische Bildung im internationalen Vergleich. In N. Jude & E. Klieme (Hrsg.), *PISA 2009 – Impulse für die Schul- und Unterrichtsforschung. Zeitschrift für Pädagogik*, 59, 229–246.
- Krohne, J. A., Meier, U. & Tillmann, K.-J. (2011). Sitzenbleiben, Geschlecht und Migration – Klassenwiederholungen im Spiegel der PISA-Daten. *Zeitschrift für Pädagogik*, 50 (3), 373–391.
- Lankes, E.-M. & Carstensen, C. H. (2010). Kann man große Klassen erfolgreich unterrichten? In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes et al. (Hrsg.), *IGLU 2006 – Die Grundschule auf dem Prüfstand. Vertiefende Analysen zu Rahmenbedingungen schulischen Lernens* (S. 121–142). Münster: Waxmann.
- Meyer, H.-D. & Rowan, B. (2006). *The new institutionalism in education*. Albany: State University of New York Press.
- OECD. (2009). *PISA data analysis manual* (2. Aufl.). Paris: OECD.
- OECD. (2010). *What Makes a School Successful?* (PISA 2009 Results). Paris: OECD.
- OECD. (2011). *When students repeat grades or are transferred out of school: What does it mean for education systems?* (PISA in Focus Nr. 2011/6). Paris: OECD.
- OECD. (2013). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. et al. (Hrsg.). (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.

- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M. et al. (Hrsg.). (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M. et al. (Hrsg.). (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.
- Ricking, H., Schulze, G. C. & Wittrock, M. (2009). *Schulabsentismus und Dropout. Erscheinungsformen – Erklärungsansätze – Intervention* (UTB Pädagogik, Bd. 3213). Paderborn: Schöningh.
- Scheerens, J. (2004). *Review of school and instructional effectiveness research. Paper commissioned for the EFA Global Monitoring Report 2005, The Quality Imperative*. Paris: Unesco.
- Scheerens, J. & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Seidel, T. & Prenzel, M. (2008). Large scale assessment. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of competencies in educational contexts* (S. 279–304). Göttingen: Hogrefe & Huber.
- Vieluf, S., Kunter, M. & van de Vijver, F. J. R. (2013). Teacher self-efficacy in cross-national perspective. *Teaching and Teacher Education*, 35, 92–103.
- Weiß, B. (2007). Wer schwänzt wie häufig die Schule? Eine vergleichende Sekundäranalyse auf Grundlage von 12 deutschen Studien. In M. Wagner (Hrsg.), *Schulabsentismus: Soziologische Analysen zum Einfluss von Familie, Schule und Freundeskreis* (S. 37–56). München: Juventa.
- Wößmann, L. (2009). International evidence on school tracking: a review. *CESifo DICE Report*, 7 (1), 26–34.
- Yeh, S. S. (2009). Class size reduction or rapid formative assessment? A comparison of cost effectiveness. *Educational Research Review*, 4, 7–15.

6

Naturwissenschaftliche Kompetenz in PISA 2012: Ergebnisse und Herausforderungen

Anja Schiepe-Tiska, Katrin Schöps, Silke Rönnebeck,
Olaf Köller und Manfred Prenzel

Naturwissenschaften und Technik prägen unser Leben und unsere Umwelt, auch wenn dies im Alltag oft unbemerkt bleibt. Erkenntnisfortschritte und Entwicklungsarbeiten in den Naturwissenschaften, die entscheidend zur Verbesserung der Lebensqualität oder zur Sicherstellung von Lebensgrundlagen beitragen, werden von vielen Menschen als selbstverständlich betrachtet. Technische Geräte werden genutzt, ohne zu wissen oder zu fragen, wie diese aufgebaut sind, und wie sie funktionieren beziehungsweise auf welchen generellen Prinzipien ihre Wirkungen beruhen. Ähnlich selbstverständlich werden Naturphänomene akzeptiert und nicht weiter hinterfragt. Das beginnt bei Wetterphänomenen und endet bei der Entstehungsgeschichte des Universums. Nun kann man sein Leben auch einigermaßen unbeschadet führen, ohne verstanden zu haben, wie ein Handy funktioniert oder was dunkle Materie ist. Wenn das Handy allerdings nicht funktioniert, oder wenn jemand vor Handys wegen der Strahlengefahr warnt, beginnt naturwissenschaftliches Wissen hilfreich zu werden. Der Stellenwert von naturwissenschaftlichem Verständnis steigt sprunghaft an, sobald der Anspruch vertreten wird, Entscheidungen im Alltag, im Beruf oder in der Politik möglichst „vernünftig“ beziehungsweise wissenschaftsbasiert zu treffen, also begründet durch Evidenzen und unter Berücksichtigung von Folgen und Nebenwirkungen. Denkt man an aktuelle Kontexte solcher Entscheidungen (wie z. B. Gesundheit, Ernährung, Nutzung von Ressourcen oder Bewertung moderner Technologien), dann wird die Relevanz einer naturwissenschaftlichen Grundbildung für die gesellschaftliche Teilhabe jedes Einzelnen ersichtlich. Naturwissenschaftliche Kompetenz erlaubt es, sich an öffentlichen Diskussionen zu beteiligen und fördert die Entwicklung zu selbständigen, verantwortungsvollen Bürgerinnen und Bürgern (Nationales MINT-Forum, im Druck). Daher gilt es auch in Zukunft, den naturwissenschaftlichen Erkenntnisstand weiter auszubauen und Wissen kreativ und verantwortungsbewusst für die Entwicklung neuer Technologien zu nutzen. Deshalb muss Nachwuchs für die Naturwissenschaften und Technik gewonnen werden, der sich durch ein besonders stark entwickeltes naturwissenschaftliches Verständnis auszeichnet.

Vor diesem Hintergrund gaben die Ergebnisse früherer internationaler Schulleistungstudien, an denen Deutschland teilnahm, Anlass zu größter Sorge. Die naturwissenschaftliche Kompetenz von Schülerinnen und Schülern lag in der *Third International Mathematics and Science Study* (TIMSS) im Vergleich zu anderen Staaten nur im

mittleren Leistungsbereich (Baumert et al., 1997; Beaton et al., 1997) und in der ersten Runde des *Programme for International Student Assessment* (PISA) im Jahr 2000 signifikant unter dem Durchschnitt der OECD-Staaten (Prenzel, Rost, Senkbeil, Häußler & Klopp, 2001).

Bereits nach TIMSS wurden in Deutschland vielfältige und zum Teil breit angelegte Maßnahmen zur Verbesserung der Kompetenzentwicklung in den Naturwissenschaften eingeleitet. Eine wichtige Maßnahme war das unter der Bezeichnung SINUS bekannte Programm zur Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, 1997; Prenzel, Friedrich & Stadler, 2009), das ab 1998 in einem Qualitätsentwicklungsansatz Problembereiche des naturwissenschaftlichen und mathematischen Unterrichts bearbeitete (z. B. die Rolle von Experimenten, der Umgang mit Fehlern, die Förderung von Mädchen und Jungen). Im Rahmen des SINUS-Programms sowie des Nachfolgeprogramms SINUS Transfer wurden zahlreiche Maßnahmen zur Verbesserung des mathematisch-naturwissenschaftlichen Unterrichts und seiner Aufgabenkultur an mehr als 1800 Sekundarschulen in Deutschland entwickelt (Ostermeier, Prenzel & Duit, 2010; Prenzel, Carstensen, Senkbeil, Ostermeier & Seidel, 2005). Im Jahr 2004 setzte mit SINUS Transfer Grundschule auch im Primarbereich eine verstärkte Förderung der mathematischen und naturwissenschaftlichen Kompetenz ein (Fischer, Dedekind, Rieck, Trepke & Prenzel, 2010). Dieses Engagement wird seit 2009 mit dem Programm SINUS an Grundschulen weiter verfolgt (Fischer & Prenzel, 2009; Fischer et al., 2013). Ab den 1990er Jahren wurden außerdem die sogenannten Kontextprojekte ins Leben gerufen, die auf eine Weiterentwicklung des Chemie-, Physik- beziehungsweise Biologieunterrichts abzielten (Bayrhuber et al., 2007; Demuth, Gräsel, Parchmann & Ralle, 2008; Mikelskis-Seifert & Duit, 2010). In diesen Projekten dienen alltägliche Anwendungen als Ausgangspunkt für einen systematischen Aufbau fachlicher Konzepte. Darüber hinaus rücken all diese Projekte auch das Forschen und das prozessbezogene Methodenlernen, die sozialen Fähigkeiten und die Selbstorganisation der Schülerinnen und Schüler in den Vordergrund.

Eine weitere wichtige Konsequenz des unbefriedigenden Abschneidens bei den internationalen Schulleistungsstudien war die Entwicklung und Einführung bundesweit verbindlicher Bildungsstandards. Im Bereich der Naturwissenschaften wurden die Bildungsstandards in Biologie, Chemie und Physik (KMK, 2005a, 2005b, 2005c) zu Beginn des Schuljahres 2005/2006 von den Bundesländern als Grundlagen der fachspezifischen Anforderungen für den Mittleren Schulabschluss übernommen. Inzwischen liegen dazu auch empirisch validierte Kompetenzstrukturmodelle für die Bereiche Fachwissen und Erkenntnisgewinnung vor (Pant, Stanat, Schroeders et al., 2013). Darüber hinaus wird naturwissenschaftliche Kompetenz über die Lebensspanne hinweg im Rahmen des Nationalen Bildungspanels (NEPS) untersucht (Hahn et al., 2013).

Bereits die Befunde von PISA 2003 deuteten an, dass diese Maßnahmen Wirkung zeigen könnten. Die naturwissenschaftliche Kompetenz der Jugendlichen in Deutschland lag nun im Durchschnittsbereich der OECD-Staaten (Rost, Walter, Carstensen, Senkbeil

& Prenzel, 2004). Diese positive Entwicklung setzte sich in PISA 2006 fort (Prenzel et al., 2007): Die Jugendlichen in Deutschland schnitten erstmals signifikant besser ab als der Durchschnitt der OECD-Staaten. Damit hatte sich die naturwissenschaftliche Kompetenz der Schülerinnen und Schüler in Deutschland schneller und stärker verbessert, als es in den Kompetenzbereichen Lesen und Mathematik der Fall war. Bei PISA 2009 stabilisierten sich die Ergebnisse auf diesem hohen Niveau (Rönnebeck, Schöps, Prenzel, Mildner & Hochweber, 2010). Kein anderer OECD-Staat konnte bisher eine vergleichbar positive Entwicklung der naturwissenschaftlichen Kompetenz verzeichnen wie Deutschland. Ob es den Schülerinnen und Schülern erneut gelingt, diese positive Entwicklung fortzuführen, ist eine zentrale Fragestellung dieses Kapitels.

In der aktuellen PISA-Erhebung 2012 wird naturwissenschaftliche Kompetenz als Nebendomäne erfasst. Im folgenden Abschnitt wird zunächst kurz die Rahmenkonzeption naturwissenschaftlicher Kompetenz vorgestellt und ihre Umsetzung im Naturwissenschaftstest erläutert. Wie die 15-jährigen Schülerinnen und Schüler in Deutschland im Jahr 2012 im internationalen Vergleich abgeschnitten haben, wird im darauffolgenden Abschnitt beschrieben. Einen vertiefenden Einblick zur Lage in Deutschland bietet schließlich der dritte Abschnitt, der die naturwissenschaftliche Kompetenz in den unterschiedlichen Schularten vergleicht und zeigt, wie sich die naturwissenschaftliche Kompetenz seit PISA 2006 verändert hat.

6.1 Naturwissenschaftliche Kompetenz in PISA 2012

Die Rahmenkonzeption des Naturwissenschaftstests wurde von PISA 2006 übernommen, als die Naturwissenschaften zum ersten Mal Hauptdomäne waren. Bei PISA stehen in den Naturwissenschaften drei Fragen im Mittelpunkt: (1) Was wissen und können Jugendliche, wenn sie mit naturwissenschaftlichen Fragen und Problemen konfrontiert werden, (2) inwiefern sind sie in der Lage, ihr Wissen flexibel in unterschiedlichen Situationen anzuwenden, und (3) welche Bedeutung messen sie Naturwissenschaften und Technik generell bei (OECD, 2013). Damit befasst sich PISA nicht nur mit naturwissenschaftlichem Wissen, wie es in der Schule erworben wird, sondern untersucht, inwieweit Jugendliche dieses Wissen in alltagsnahen persönlichen, sozialen und globalen Kontexten anwenden können (Bybee & McCrae, 2011). Dieses Verständnis berücksichtigt die internationalen Diskussionen über Ziele naturwissenschaftlicher Grundbildung (Osborne, 2007; Osborne & Dillon, 2008; Roberts, 2007) und beruht auf einer differenzierten Vorstellung naturwissenschaftlicher Grundbildung für alle (*Scientific Literacy*), die Menschen dazu befähigt, die wachsenden Anforderungen einer durch Naturwissenschaften und Technik geprägten Welt erfolgreich zu bewältigen (vgl. OECD, 2013; Prenzel et al., 2007). Dementsprechend ist naturwissenschaftliche Grundbildung ein Fundament für lebenslanges Lernen (Prenzel et al., 2001). Sie beschreibt die Fähigkeit einer Person,

- naturwissenschaftliches Wissen anzuwenden, um Fragestellungen zu erkennen, sich neues Wissen anzueignen, naturwissenschaftliche Phänomene zu beschreiben und aus Belegen Schlussfolgerungen zu ziehen,
- die charakteristischen Eigenschaften der Naturwissenschaften als eine Form menschlichen Wissens und Forschens zu verstehen,
- zu erkennen und sich darüber bewusst zu sein, wie Naturwissenschaften und Technik unsere materielle, intellektuelle und kulturelle Umwelt formen,
- sowie die Bereitschaft, sich mit naturwissenschaftlichen Ideen und Themen zu beschäftigen und sich reflektierend mit ihnen auseinanderzusetzen (OECD, 2007, 2013; Prenzel et al., 2007).

Die Rahmenkonzeption der Domäne Naturwissenschaften

Die naturwissenschaftliche Grundbildung ist bei PISA in ein theoretisches Kompetenzmodell eingebettet. Diese Rahmenkonzeption unterscheidet drei Teilkompetenzen (vgl. Abbildung 6.1): Das *Erkennen und Formulieren von Fragestellungen*, das *Beschreiben und Erklären naturwissenschaftlicher Phänomene* sowie das *Interpretieren naturwissenschaftlicher Evidenz*, um Entscheidungen treffen zu können (OECD, 2013). Diese Teilkompetenzen sind in verschiedene, alltagsnahe Kontexte eingebettet, um möglichst unterschiedliche Interessen von Jugendlichen berücksichtigen zu können. Die Kontexte beziehen sich auf *persönliche* Lebenssituationen (z. B. Bewegung und Ernährung), *soziale und gemeinschaftliche* Situationen (z. B. Entsorgung von Abfällen) sowie *globale* Situationen (z. B. Klimaveränderung und das Aussterben von Arten).

Zwei Wissensaspekte bilden die Grundlage für die drei Teilkompetenzen. Der erste Aspekt ist das eher objektorientierte, fachbezogene *naturwissenschaftliche Wissen*, welches den Bereichen physikalische Systeme, lebende Systeme, Erd- und Weltraumsysteme sowie technologische Systeme zugeordnet werden kann. Der zweite Aspekt ist das eher prozessorientierte, fächerübergreifende (Meta-) *Wissen über die Naturwissenschaften*, welches in die Bereiche naturwissenschaftliches Forschen und naturwissenschaftliche Erklärungen unterteilt wird. Naturwissenschaftliches Forschen stellt einen zentralen Prozess der Naturwissenschaften dar, während naturwissenschaftliche Erklärungen sich auf die Ergebnisse des naturwissenschaftlichen Forschens, deren Darstellung und Besonderheiten sowie das Verständnis von Regeln des naturwissenschaftlichen Arbeitens beziehen (Prenzel et al., 2007).

Die Entwicklung der Teilkompetenzen basiert darüber hinaus auf den motivationalen Orientierungen einer Person, die eine wichtige Facette der naturwissenschaftlichen Grundbildung und Bestandteil eines umfassenden Kompetenzbegriffs sind (Weinert, 1999). Damit sind Einstellungen und Überzeugungen in Bezug auf Naturwissenschaften wie *Interesse an Naturwissenschaften*, *Wertschätzung naturwissenschaftlicher Forschung* und *Verantwortungsbewusstsein gegenüber der Umwelt und natürlichen Ressourcen* gemeint.

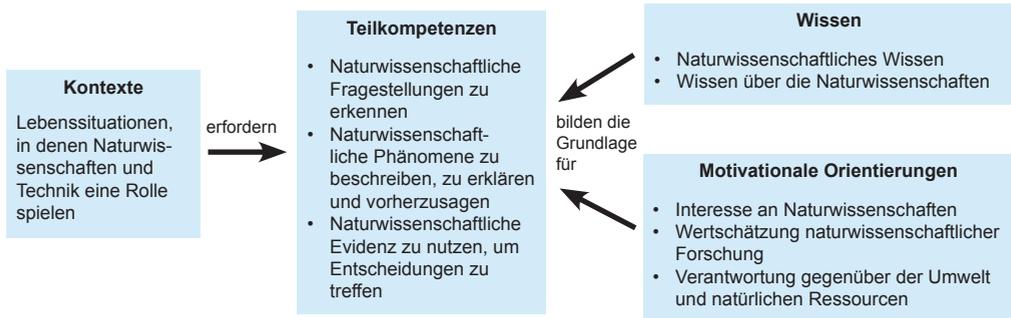


Abbildung 6.1: Die PISA-Rahmenkonzeption naturwissenschaftlicher Grundbildung

Der Naturwissenschaftstest in PISA 2012

Da die Naturwissenschaften bei PISA 2012 als Nebendomäne untersucht werden, umfasst der Test nur Aufgaben, die aus früheren Erhebungsrounden stammen und bereits 2009 zum Einsatz kamen. Diese Anker- oder Link-Aufgaben werden über die Erhebungsrounden geheim gehalten. Jede Aufgabe beinhaltet Stimulusmaterial und bis zu vier Items zu einem Kontext (vgl. Abbildung 6.2). Es wurden bei PISA 2012 insgesamt 18 Aufgaben eingesetzt, die sich aus 53 Items zusammensetzten. Diese Items lassen sich einerseits den Teilkompetenzen und andererseits den Wissensaspekten zuordnen (vgl. Tabelle 6.1). Die Zuordnung erfolgte so, dass jedes Item sowohl eine der Teilkompetenzen als auch eines der Systeme naturwissenschaftlichen Wissens oder einen Aspekt des Wissens über die Naturwissenschaften erfasst. Motivationale Orientierungen wurden aufgrund des Status von Naturwissenschaften als Nebendomäne bei PISA 2012 nicht erhoben.

Bei der Bearbeitung der Aufgaben mussten drei unterschiedliche Antwortformate bewältigt werden. Wie bereits bei PISA 2009 beinhalteten die Aufgaben 35 Mehrfachauswahl-Items (Multiple-Choice-Items), bei denen die Schülerinnen und Schüler aus mehreren vorgegebenen Antworten eine oder mehrere richtige Antworten auswählen mussten. Einige dieser Items hatten ein sogenanntes komplexes Mehrfachwahl-Format, bei dem für mehrere vorgegebene Aussagen jeweils entschieden werden musste, ob die Aussage richtig oder falsch ist. Eine Teilmenge von 17 Items besaß ein offenes Antwortformat, bei dem die Schülerinnen und Schüler selbständig eine Antwort formulierten und diese aufschrieben. Ihre Antworten wurden im Nachhinein von geschulten Kodierinnen und Kodierern den Kategorien richtig, teilweise richtig oder falsch zugeordnet (vgl. Kapitel 10). Zusätzlich gab es ein Item mit einem geschlossen Antwortformat, bei dem die Fünfzehnjährigen als Antwort ein „+“ oder „-“ angaben.

Tabelle 6.1: Verteilung der Items des Naturwissenschaftstests auf die drei Teilkompetenzen und zwei Wissensaspekte

Bereich der naturwissenschaftlichen Kompetenz	Anzahl der Items
<i>Teilkompetenzen</i>	53
Naturwissenschaftliche Fragestellungen erkennen	13
Naturwissenschaftliche Phänomene erklären	22
Naturwissenschaftliche Evidenz nutzen	18
<i>Naturwissenschaftliches Wissen</i>	26
Physikalische Systeme	6
Lebende Systeme	9
Erd- und Weltraumsysteme	7
Technologische Systeme	4
<i>Wissen über Naturwissenschaften</i>	27
Naturwissenschaftliches Forschen	14
Naturwissenschaftliche Erklärung	13

Die Kompetenzstufen

Wie bei PISA 2009 wurde aufgrund ihres Status als Nebendomäne nur eine Gesamtskala für die naturwissenschaftliche Kompetenz gebildet. Die Überprüfung einzelner Teilkompetenzen wird erst bei PISA 2015 wieder möglich sein, wenn die Naturwissenschaften zum zweiten Mal Hauptdomäne sein werden. Die Gesamtskala Naturwissenschaften mit einem ursprünglichen OECD-Mittelwert von 500 und einer Standardabweichung von 100 wurde in der PISA-Erhebung 2006 definiert, als die Naturwissenschaften das erste Mal die Hauptdomäne waren. Diese Skala bildet wiederum den Bezugspunkt für die Ergebnisse von PISA 2012.

Wie bereits in den vorangegangenen PISA-Zyklen wurden psychometrische Modelle der Item-Response-Theorie verwendet, um das Antwortverhalten der Schülerinnen und Schüler beim Naturwissenschaftstest zu skalieren (vgl. Kapitel 10; Adams & Wu, 2007; Rönnebeck et al. 2010). Dadurch können die Fähigkeiten (Kompetenzen) einer Person und die Schwierigkeiten von Items auf einer gemeinsamen Skala abgebildet werden. Diese Skala wird in Abschnitte unterteilt, die Items mit vergleichbaren Schwierigkeiten zusammenfassen. Die Abschnitte werden bei PISA als Kompetenzstufen bezeichnet. Für die naturwissenschaftliche Kompetenz werden sechs Kompetenzstufen unterschieden, die inhaltlich definiert sind (vgl. Tabelle 6.2; Prenzel et al., 2007).

Die Anforderungen, die die Items an die Fähigkeiten der Schülerinnen und Schüler stellen, werden mit ansteigender Kompetenzstufe komplexer. Items, die dem unteren Bereich der Skala zuzuordnen sind, bilden einfache und den Jugendlichen bekannte Kontexte ab. Zur Lösung dieser Items wenden sie ihr naturwissenschaftliches Wissen

Tabelle 6.2: Stufen naturwissenschaftlicher Kompetenz in PISA 2012

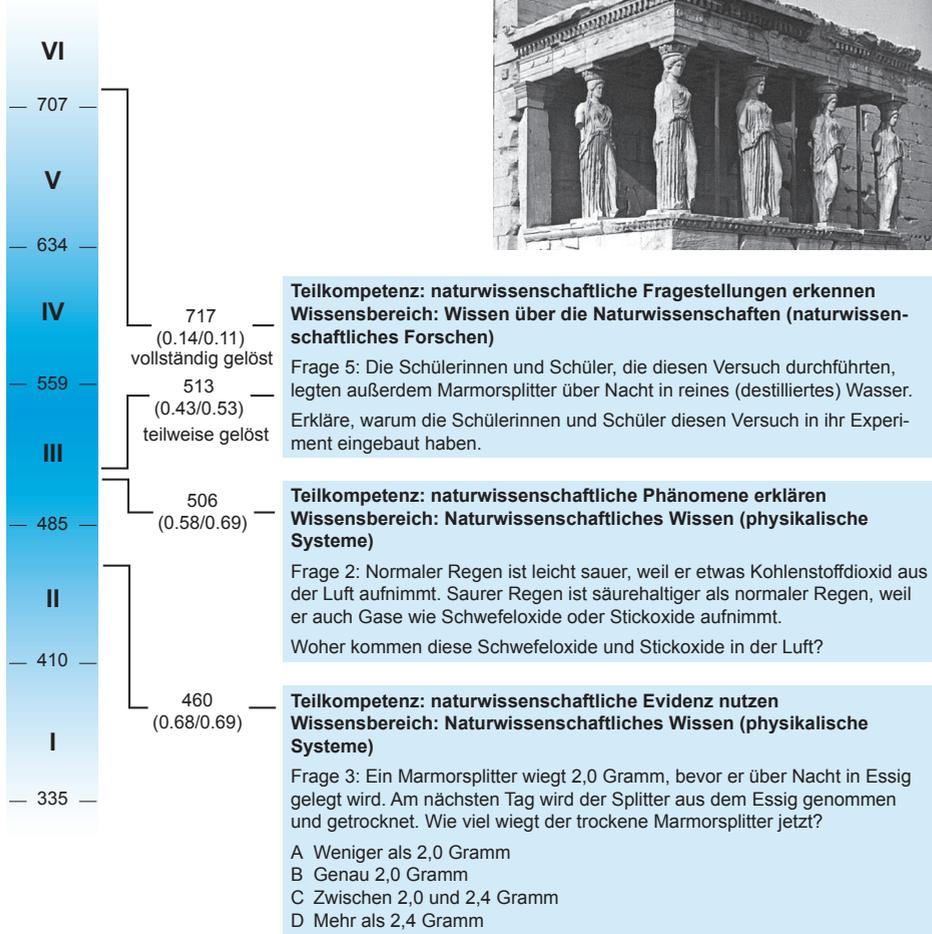
Kompetenzstufe	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
VI > 707 Punkte	Auf Stufe VI können Schülerinnen und Schüler in konsistenter Weise naturwissenschaftliches Wissen und Wissen über die Naturwissenschaften in einer Vielzahl komplexer Lebenssituationen erkennen, erklären und anwenden. Sie können verschiedene Informationsquellen und Erklärungen zueinander in Beziehung setzen und die Beweise, die aus diesen Quellen folgen, nutzen, um Entscheidungen zu begründen. Sie demonstrieren ein weit entwickeltes naturwissenschaftliches und logisches Denkvermögen und sind bereit, ihr naturwissenschaftliches Verständnis einzusetzen, um Lösungen für unbekannte naturwissenschaftliche oder technologische Probleme zu finden. Schülerinnen und Schüler auf dieser Stufe können naturwissenschaftliches Wissen anwenden und Argumente entwickeln, um Empfehlungen auszusprechen beziehungsweise Entscheidungen zu treffen, die von persönlicher, sozialer oder globaler Bedeutung sind.
V 634–707 Punkte	Auf Stufe V können Schülerinnen und Schüler die Bedeutung der Naturwissenschaften in komplexen Lebenssituationen erkennen. Sie können sowohl ihr konzeptuelles Wissen als auch ihr Wissen über die Naturwissenschaften auf diese Situationen anwenden und naturwissenschaftliche Beweise vergleichen, auswählen und bewerten, um in angemessener Weise auf diese Situationen zu reagieren. Schülerinnen und Schüler auf dieser Stufe besitzen ein gut entwickeltes Verständnis naturwissenschaftlicher Untersuchungen, können ihr Wissen verknüpfen und Situationen kritisch bewerten. Sie können auf Basis ihrer kritischen Analysen auf Beweisen beruhende Erklärungen entwickeln und Aussagen treffen.
IV 559–633 Punkte	Schülerinnen und Schüler auf Stufe IV können mit Situationen und Fragestellungen umgehen, die es erfordern, Rückschlüsse über die Rolle von Naturwissenschaften und Technik zu ziehen. Sie können Erklärungen aus den verschiedenen naturwissenschaftlichen beziehungsweise technologischen Disziplinen auswählen und zueinander in Beziehung setzen, um diese Erklärungen direkt auf alltägliche Lebenssituationen anzuwenden. Schülerinnen und Schüler auf dieser Stufe sind in der Lage, ihre Handlungen zu reflektieren und ihre Entscheidungen auf Basis ihres naturwissenschaftlichen Wissens und der vorhandenen Evidenz zu kommunizieren.
III 485–558 Punkte	Auf Stufe III können Schülerinnen und Schüler eindeutig beschriebene naturwissenschaftliche Fragestellungen in verschiedenen Kontexten erkennen. Sie können Fakten und Wissen auswählen, um naturwissenschaftliche Phänomene zu beschreiben und sind in der Lage, einfache Modelle oder Auswertungsverfahren anzuwenden. Schülerinnen und Schüler auf dieser Stufe können Konzepte aus den verschiedenen naturwissenschaftlichen Disziplinen interpretieren und anwenden. Sie sind in der Lage, Fakten zu verwenden, um kurze Berichte zu verfassen und Entscheidungen zu treffen, die auf naturwissenschaftlichem Wissen beruhen.
II 410–484 Punkte	Schülerinnen und Schüler auf Stufe II besitzen ein angemessenes naturwissenschaftliches Wissen, um in bekannten Kontexten mögliche Erklärungen zu liefern oder aus einfachen Experimenten Schlussfolgerungen zu ziehen. Sie sind in der Lage, in direkten Zusammenhängen zu denken und die Ergebnisse naturwissenschaftlicher Untersuchungen oder technologischer Problemlöseverfahren in ihrer Umgangssprache wiederzugeben.
I 335–409 Punkte	Auf Stufe I ist das naturwissenschaftliche Wissen der Schülerinnen und Schüler derart eingeschränkt, dass es nur auf einige wenige, bekannte Situationen angewendet werden kann. Die Schülerinnen und Schüler auf dieser Stufe sind in der Lage, offensichtliche naturwissenschaftliche Erklärungen zu liefern, die direkt vorliegenden Beweisen zu entnehmen sind.

Saurer Regen

Das Foto unten zeigt Statuen, die so genannten Kariatiden, die vor mehr als 2500 Jahren auf der Akropolis in Athen aufgestellt wurden. Die Statuen bestehen aus Marmor (einer Gesteinsart). Marmor besteht aus Calciumcarbonat.

1980 wurden die Originalstatuen in das Innere des Museums der Akropolis gebracht und durch Kopien ersetzt. Die Originale waren vom sauren Regen zerfressen worden.

Die Wirkung von saurem Regen auf Marmor kann simuliert werden, indem man Marmorsplitter über Nacht in Essig legt. Essig und saurer Regen haben in etwa denselben Säuregehalt. Wenn man ein Stück Marmor in Essig legt, bilden sich Gasblasen. Das Gewicht der trockenen Marmorsplitter kann vor und nach dem Versuch bestimmt werden.



Anmerkung. Die Werte an den Verbindungslinien zwischen den Beispielen und der Kompetenzsäule geben das für eine 62 %-Lösungswahrscheinlichkeit erforderliche Kompetenzniveau und die Werte in Klammern die relativen internationalen (OECD-Durchschnitt) und deutschen Lösungshäufigkeiten an. Hierbei handelt es sich um beobachtete Häufigkeiten.

Abbildung 6.2: Beispielaufgabe „Saurer Regen“

direkt an und zeigen, dass sie einfache, allgemein bekannte naturwissenschaftliche Konzepte verstehen. Zur Lösung der Items, die im oberen Bereich der Skala liegen, müssen dagegen komplexe, unbekannte Daten interpretiert sowie naturwissenschaftliche Konzepte und Prozesse auf unbekannte Situationen und Fragestellungen übertragen und angewendet werden. Abbildung 6.2 zeigt eine Beispielaufgabe und wie die Items den Kompetenzstufen zugeordnet werden können. Inhaltlich lassen sich die einzelnen Items darüber hinaus den verschiedenen Teilkompetenzen und Wissensbereichen zuordnen.

Die Kompetenzstufe II wird bei PISA als Mindestniveau definiert, das alle fünfzehnjährigen Jugendlichen erreichen sollten. Jugendlichen mit Werten auf Kompetenzstufe I und darunter fehlen grundlegende naturwissenschaftliche Kompetenzen, um einfachste Aufgaben zu lösen. Ihnen wird es im Alltag schwerfallen, einfache naturwissenschaftliche Probleme oder Zusammenhänge zu verstehen oder aus Informationen Schlussfolgerungen zu ziehen. Hierdurch sind ihre Voraussetzungen für Ausbildungen oder Berufslaufbahnen, die ein gewisses naturwissenschaftliches Verständnis erfordern, als vergleichsweise schlecht einzuschätzen. Schülerinnen und Schüler mit Werten auf den Kompetenzstufen V und VI dagegen zeichnen sich durch eine sehr stark ausgeprägte naturwissenschaftliche Kompetenz aus. Sie sind in der Lage, Aufgaben zu lösen, die besonders hohe fachliche Anforderungen stellen. Damit bringen sie hervorragende Voraussetzungen für Studiengänge im natur- oder ingenieurwissenschaftlichen Bereich mit.

6.2 Naturwissenschaftliche Kompetenz im internationalen Vergleich

Einen ersten Eindruck der Ergebnisse vermitteln die Testmittelwerte und Streuungsmaße der Staaten, die an PISA 2012 teilgenommen haben. Die Verteilung der Schülerinnen und Schüler auf die unterschiedlichen Kompetenzstufen verdeutlicht im Anschluss, welche konkreten Leistungen die Fünfzehnjährigen in den einzelnen Ländern erreichen.

Naturwissenschaftliche Kompetenz: Mittelwerte im Vergleich

Die Mittelwerte, Standardabweichungen und Perzentilbänder der naturwissenschaftlichen Kompetenz in den OECD-Staaten sind in Abbildung 6.3 dargestellt (vgl. auch Tabelle 6 im Anhang). Bei PISA 2012 beträgt die durchschnittliche naturwissenschaftliche Kompetenz über alle OECD-Staaten 501 Punkte und bleibt damit gegenüber PISA 2009 unverändert. Die naturwissenschaftliche Kompetenz der Fünfzehnjährigen in Deutschland beträgt 524 Punkte. Deutschland gehört damit zur Gruppe der Staaten, die signifikant oberhalb des OECD-Durchschnitts liegen. Insgesamt erreichen die Schülerinnen und Schüler in 16 Staaten mittlere Kompetenzwerte, die den OECD-Durchschnitt signifikant übertreffen. An der Spitze der OECD-Staaten stehen Japan mit 547 Punkten, Finnland mit 545 Punkten, Estland mit 541 Punkten und Korea mit 538 Punkten.

Die Kompetenzwerte dieser Spitzengruppe von Staaten liegen bis zu einer halben Standardabweichung über dem OECD-Durchschnitt und können als Beispiel dafür dienen, welches naturwissenschaftliche Kompetenzniveau bereits bis zu einem Alter von fünfzehn Jahren entwickelt werden kann. Diese vier Staaten liegen auch signifikant oberhalb des mittleren Kompetenzwerts von Deutschland, wohingegen sich Polen mit 526 Punkten und Kanada mit 525 Punkten nicht signifikant von Deutschland unterscheiden. Die niedrigsten mittleren Kompetenzwerte erreicht Mexiko mit 415 Punkten. Die Schülerinnen und Schüler in Mexiko zeigen damit eine mittlere naturwissenschaftliche Kompetenz, die fast eine Standardabweichung unter dem OECD-Durchschnitt und eine Drittel Standardabweichung unter der des zweitschwächsten Staates, Chile (445 Punkte), liegt.

Die Größenordnung der Unterschiede in den Kompetenzwerten kann mit Hilfe von Daten aus einer Längsschnittstudie in Deutschland veranschaulicht werden. Bei der Follow-up-Erhebung von PISA 2003 betrug der mittlere Kompetenzzuwachs in den Naturwissenschaften von der 9. zur 10. Klassenstufe 21 Punkte (Walter, Senkbeil, Rost, Carstensen & Prenzel, 2006). Betrachtet man dies als Maßstab zur Interpretation der beobachteten Kompetenzunterschiede zwischen den Staaten, zeigen sich beträchtliche Leistungsunterschiede. So beträgt der Leistungsabstand in den Naturwissenschaften zwischen Japan und Belgien beispielsweise ca. zwei Schuljahre, zwischen Japan und Mexiko mehr als sechs Schuljahre.

Werden auch die OECD-Partnerstaaten, die sich im Naturwissenschaftstest durch überdurchschnittliche Leistungen auszeichnen, in die Betrachtung einbezogen, dann ist festzustellen, dass insbesondere die Schülerinnen und Schüler in Shanghai (China) mit 580 Punkten nochmals deutlich höhere mittlere Kompetenzwerte erreichen als die Fünfzehnjährigen in Japan, das unter den OECD-Staaten am besten abgeschnitten hat (vgl. Tabelle 6 im Anhang). Der Unterschied von 33 Punkten entspricht in etwa einer Drittel Standardabweichung. Ausgezeichnete naturwissenschaftliche Kompetenzen werden außerdem von den OECD-Partnerstaaten Hongkong (China) mit 555 Punkten und Singapur mit 551 Punkten erzielt. Gemäß obiger Betrachtung liegen jedoch auch diese Partnerstaaten mehr als ein Schuljahr hinter Shanghai (China) zurück. Sie unterscheiden sich in den mittleren Kompetenzwerten ihrer Jugendlichen nicht signifikant von Japan.

Die Streuung der naturwissenschaftlichen Kompetenz

Die durchschnittliche Streuung der naturwissenschaftlichen Kompetenz wird im Kennwert der Standardabweichung sichtbar (vgl. Abbildung 6.3). An der Standardabweichung lässt sich ablesen, wie homogen die Kompetenzverteilungen in den einzelnen Staaten sind. Im Durchschnitt aller OECD-Staaten beträgt die Streuung 93 Punkte; die für Deutschland gemessene Standardabweichung von 95 Punkten liegt im Bereich des OECD-Mittelwerts.

Eine im internationalen Vergleich besonders heterogene Verteilung der naturwissenschaftlichen Kompetenz findet sich in Israel ($SD = 108$), Neuseeland ($SD = 105$) und Luxemburg ($SD = 103$). Aber auch in Australien, Belgien, dem Vereinigten König-

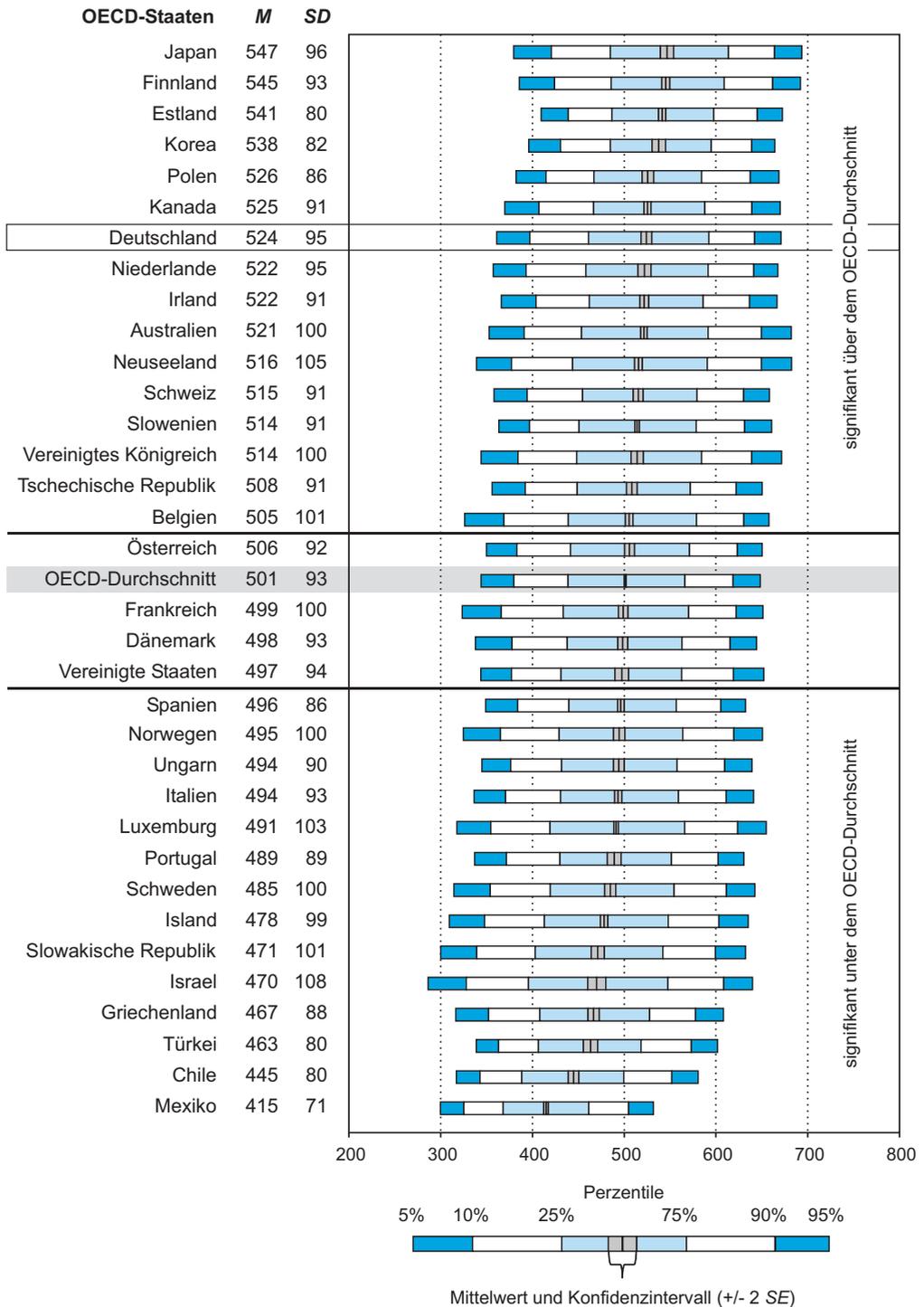


Abbildung 6.3: Perzentilbänder naturwissenschaftlicher Kompetenz in den OECD-Staaten

reich, Frankreich, Norwegen, Schweden, der Slowakischen Republik und Island streut die naturwissenschaftliche Kompetenz mit Standardabweichungen zwischen 99 und 101 Punkten signifikant stärker als im OECD-Durchschnitt. Eine besonders homogene Verteilung der naturwissenschaftlichen Kompetenz zeigt sich im Bereich der leistungsstarken Staaten insbesondere in Estland ($SD = 80$), Korea ($SD = 82$) und Polen ($SD = 86$). Diesen Staaten gelingt es offensichtlich sehr gut, die naturwissenschaftliche Kompetenz ihrer Jugendlichen im Sinne von *Naturwissenschaften für alle* (vgl. Fensham, 1985) in der Breite zu fördern, ohne die Spitzengruppe der hochkompetenten Schülerinnen und Schüler zu vernachlässigen. Eine geringe Streuung findet sich ebenso im Bereich der leistungsschwachen Staaten in Mexiko, Chile und der Türkei mit Standardabweichungen zwischen 71 und 80 Punkten.

Grafisch lässt sich die Streuung durch sogenannte Perzentilbänder darstellen (Abbildung 6.3; Kennwerte können der Tabelle 6 im Anhang entnommen werden). So geben beispielsweise die Kennwerte für das 5. und 95. Perzentil Auskunft über die Kompetenzniveaus in den Randbereichen der Kompetenzverteilung. Ein erneuter Blick auf die Staaten in der Spitzengruppe des internationalen Vergleichs zeigt in diesen Randbereichen deutliche Unterschiede. Während beispielsweise in Japan und Finnland die Spitzengruppe der kompetenzstärksten 5 Prozent der Schülerinnen und Schüler bei 693 und 692 Punkten beginnt, liegt diese Grenze für Estland bei 672 Punkten und in Deutschland bei 671 Punkten. Die Spitzengruppe der Jugendlichen in Japan und Finnland erreicht damit ein höheres Niveau naturwissenschaftlicher Kompetenz als die Spitzengruppe in Estland und Deutschland.

Am unteren Rand der Leistungsverteilung zeigen die Perzentilwerte an, auf welchem Niveau zum Beispiel die schwächsten 5 oder 10 Prozent der Schülerinnen und Schüler eines Landes liegen. Das 5-Prozent-Perzentil gibt etwa den Testwert an, der von 95 Prozent der Altersgruppe erreicht wird. Dieser Perzentilwert liegt für Estland bei 409 und für Korea bei 396 Punkten. Für Deutschland liegt das 5-Prozent-Perzentil bei 361 Punkten. Das bedeutet, dass die schwächsten Schülerinnen und Schüler in Estland und Korea (aber auch in Japan und Finnland) ein deutlich höheres Niveau naturwissenschaftlicher Kompetenz erreichen als die Jugendlichen in Deutschland. So beträgt der Kennwert für das 10-Prozent-Perzentil in Deutschland 397 Punkte – und entspricht dem Wert für das 5-Prozent-Perzentil in Korea (396 Punkte). Diese Vergleiche weisen darauf hin, dass bezogen auf das Ziel *Naturwissenschaften für alle* in vielen Staaten, und so auch in Deutschland, noch weitere Anstrengungen nötig sind.

Verteilung auf die Stufen der naturwissenschaftlichen Kompetenz

Kompetenzstufen erlauben es, die von den Schülerinnen und Schülern im Naturwissenschaftstest erreichten Kompetenzwerte kriteriumsorientiert (also anhand inhaltlicher Bezugsnormen) zu interpretieren. Insbesondere die Randbereiche der Kompetenzstufenverteilungen sind dabei von Interesse. Wie schon bei der Diskussion der Perzentilwerte deutlich wurde, zeigen sie an, inwieweit es in einem Bildungssystem gelingt, einerseits die Anteile schwacher Schülerinnen und Schüler gering zu halten und andererseits eine Spitzengruppe zu entwickeln, die beste Voraussetzungen für eine vertiefte Beschäftigung mit den Naturwissenschaften und verwandten Fachrichtungen mitbringt.

Die Abbildung 6.4 zeigt für alle OECD-Staaten die Anteile der Schülerinnen und Schüler auf der Kompetenzstufe I und darunter beziehungsweise den Anteil auf den Kompetenzstufen V und VI. Im OECD-Durchschnitt befinden sich 17.8 Prozent der Jugendlichen auf Kompetenzstufe I (13.0 Prozent) und darunter (4.8 Prozent). Diese Anteile umfassen also Jugendliche mit einem marginalen naturwissenschaftlichen Verständnis, das für viele Ausbildungen unzureichend ist und kaum begründete Entscheidungen (z. B. zu Fragen der Ernährung und Gesundheit oder auch der Umwelt) zulässt. Der entsprechende Anteil der Jugendlichen in Deutschland fällt mit 12.2 Prozent (9.3 Prozent auf Stufe I und 2.9 Prozent darunter) signifikant geringer aus als der OECD-Durchschnitt. Hervorzuheben sind wieder die sehr geringen Anteile von Jugendlichen mit schwacher naturwissenschaftlicher Kompetenz in anderen OECD-Staaten der Spitzengruppe, die etwa für Estland 5.0 Prozent, Korea 6.6 Prozent, Finnland 7.7 Prozent, Japan 8.5 Prozent und Polen 9.0 Prozent betragen. Extrem niedrige Anteile auf und unter der ersten Stufe naturwissenschaftlicher Kompetenz finden sich auch in den leistungsstarken OECD-Partnerstaaten Shanghai (China) mit 2.7 Prozent, Hongkong (China) mit 5.6 Prozent und Singapur mit 9.6 Prozent.

Im Bereich der herausragenden Leistungen, auf den Kompetenzstufen V und VI also, befinden sich im OECD-Durchschnitt 8.4 Prozent der Schülerinnen und Schüler (davon 1.2 Prozent auf Kompetenzstufe VI). In Deutschland beträgt dieser Anteil 12.2 Prozent (davon 1.6 Prozent auf Stufe VI) und weist damit auf eine überdurchschnittlich große Spitzengruppe in den Naturwissenschaften hin. Allerdings lässt die Abbildung 6.4 eine Anzahl von Staaten mit noch deutlich größeren Anteilen auf den beiden höchsten Stufen naturwissenschaftlicher Kompetenz erkennen. In der Vergleichsgruppe der OECD-Staaten findet man die größten Anteile kompetenzstarker Jugendlicher in Japan (18.2 Prozent) und in Finnland (17.1 Prozent). Auch in Australien (13.6 Prozent), Neuseeland (13.4 Prozent) und Estland (12.8 Prozent) erreichen im Vergleich zu Deutschland etwas mehr Jugendliche die Kompetenzstufen V und VI. Werden auch die OECD-Partnerstaaten berücksichtigt, erreicht vor allem in den Spitzenländern Shanghai (China, 27.2 Prozent), Singapur (22.7 Prozent) und Hongkong (China, 16.7 Prozent) ein vergleichsweise hoher Anteil von Schülerinnen und Schülern die Kompetenzstufen V und VI. Insgesamt zeigt sich bei der Betrachtung der Kompetenzstufenverteilungen innerhalb der OECD-Staaten wie auch 2009, dass es insbesondere in Japan und Finnland in den Naturwissen-

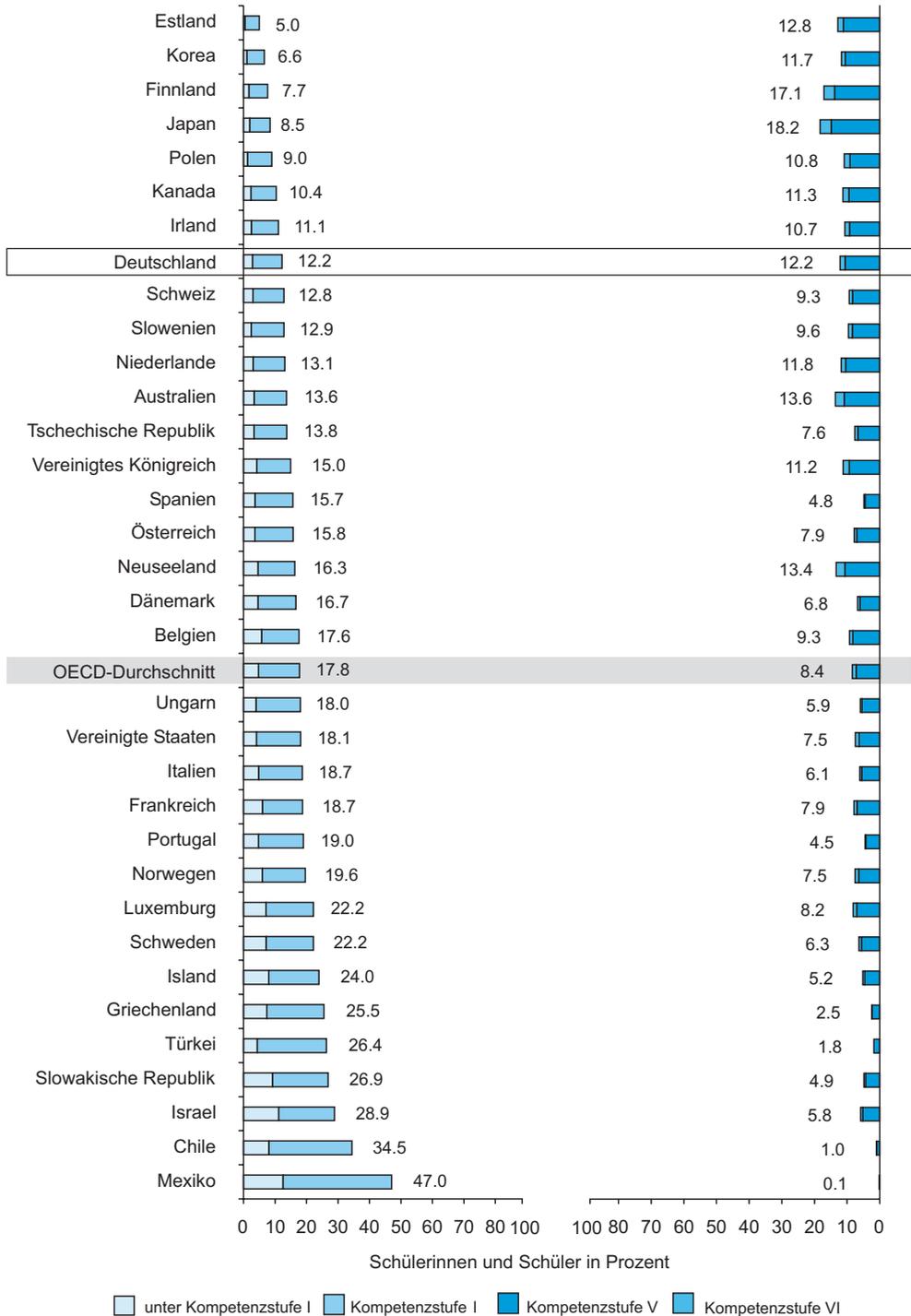


Abbildung 6.4: Prozentuale Anteile von Schülerinnen und Schülern auf Kompetenzstufe I oder darunter beziehungsweise auf Kompetenzstufe V oder VI

schaften gelingt, kompetenzschwache und kompetenzstarke Jugendliche gleichermaßen erfolgreich zu fördern.

Unterschiede zwischen Mädchen und Jungen in der naturwissenschaftlichen Kompetenz

Im Durchschnitt der OECD-Staaten ist keine statistisch signifikante Differenz in der naturwissenschaftlichen Kompetenz von Mädchen und Jungen zu beobachten (vgl. Abbildung 6.5). Dieses Ergebnis schließt an den Befund von PISA 2009 an (vgl. Rönnebeck et al., 2010) und zeigt, dass in den Naturwissenschaften weltweit kaum mehr Disparitäten nach Geschlecht festzustellen sind. Statistisch signifikante Differenzen von mehr als 10 Punkten finden sich lediglich in Dänemark, Japan, dem Vereinigten Königreich und Luxemburg zugunsten der Jungen sowie in Finnland, Griechenland und der Türkei zugunsten der Mädchen. Anzumerken bleibt, dass hinter den ähnlichen Mittelwerten für die Leistungen im Naturwissenschaftstest doch Unterschiede stecken könnten, die sich bei pauschaler Betrachtung ausgleichen. In dem umfassenderen Naturwissenschaftstest bei PISA 2006 zeichneten sich relative Stärken der Mädchen bei lebenswissenschaftlichen Aufgaben und bei den Jungen bei physikalischen Themen ab (vgl. Prenzel et al., 2007). In der nächsten Erhebungsrunde 2015, wenn die Naturwissenschaften als Hauptdomäne erneut umfassend getestet werden, kann dieser Frage differenzierter nachgegangen werden.

In Deutschland gibt es – wie im Durchschnitt der OECD-Staaten – keinen signifikanten Unterschied zwischen Mädchen und Jungen in der naturwissenschaftlichen Kompetenz. Bereits bei PISA 2006 und PISA 2009 waren die Differenzwerte statistisch nicht mehr gegen den Zufall abzusichern. Betrachtet man, wie sich die Mädchen und Jungen in Deutschland auf die unterschiedlichen Kompetenzstufen verteilen, zeigt sich, dass sich auf den Kompetenzstufen I und darunter etwas mehr Jungen (12,9 Prozent) als Mädchen (11,5 Prozent) befinden. Allerdings erreichen auch etwas mehr Jungen (12,9 Prozent) als Mädchen (11,4 Prozent) die Kompetenzstufen V und VI.

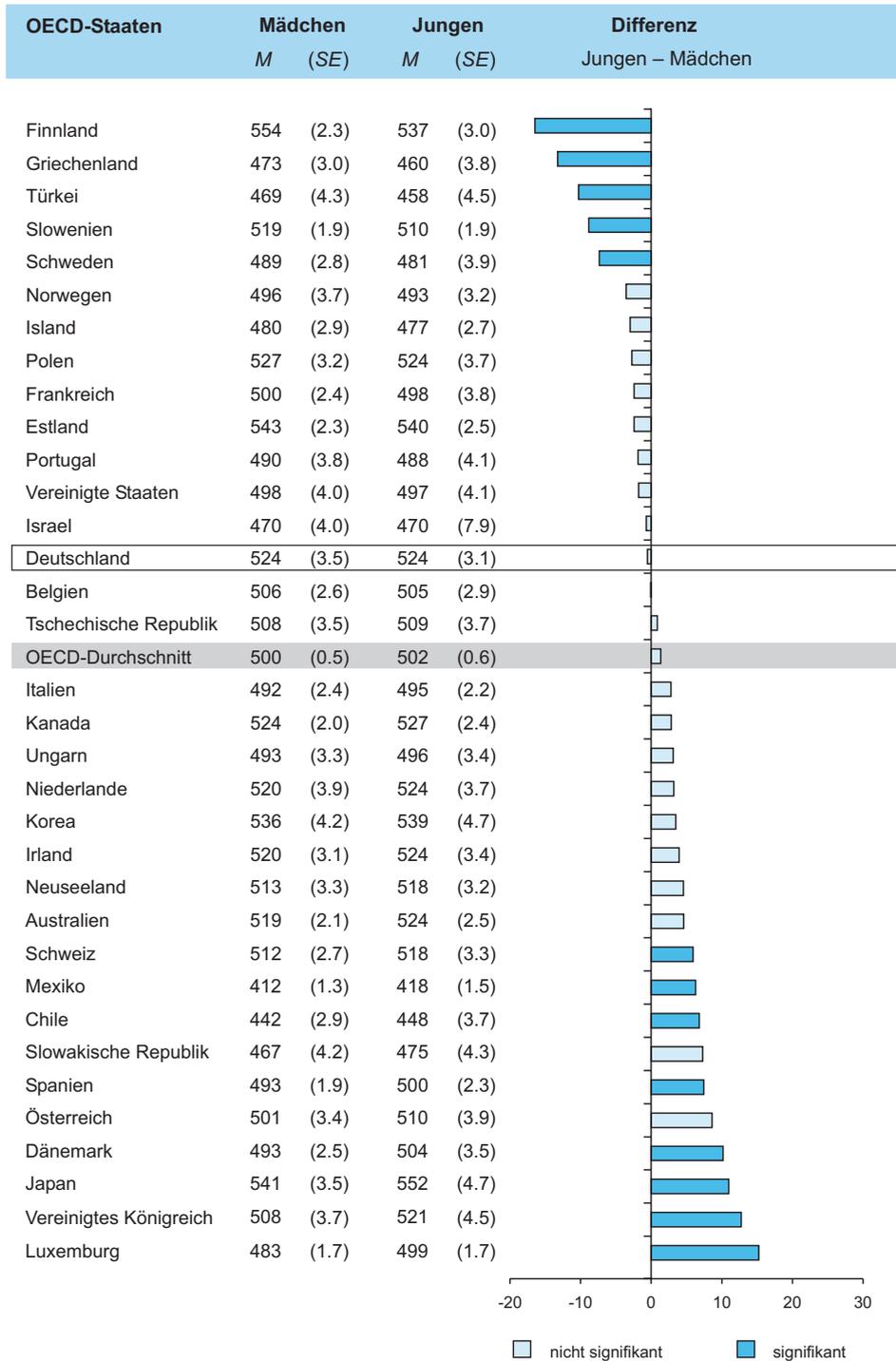


Abbildung 6.5: Mittelwerte naturwissenschaftlicher Kompetenz von Mädchen und Jungen in den OECD-Staaten

6.3 Vertiefende Analysen der naturwissenschaftlichen Kompetenz in Deutschland

Die folgenden Abschnitte präsentieren differenziertere Befunde zur naturwissenschaftlichen Kompetenz fünfzehnjähriger Schülerinnen und Schüler in Deutschland. Es werden zunächst Unterschiede zwischen den Schularten in Deutschland dargestellt. Im Anschluss wird gezeigt, wie sich die naturwissenschaftliche Kompetenz in Deutschland seit PISA 2006, als die Naturwissenschaften die Hauptdomäne bildeten, verändert hat.

6.3.1 Unterschiede zwischen Schularten

In Deutschland finden in den Ländern zurzeit große Umstrukturierungen in der Schullandschaft statt. Zahlreiche Schulen mit einer Sekundarstufe I – mit Ausnahme der Gymnasien – wurden durch die Kombination von Bildungsgängen in neue oder neu bezeichnete Schularten überführt (Pant, Stanat, Pöhlmann et al., 2013). Dieser Prozess ist noch nicht abgeschlossen. Damit verbunden sind Veränderungen in der Zusammensetzung der Schülerschaft in bestimmten Schularten. Im vorliegenden Bericht werden zwar sieben Schularten unterschieden: Gymnasium, Realschule, Hauptschule, Schule mit mehreren Bildungsgängen, Integrierte Gesamtschule, Berufsschule und Förderschule; differenzierte Aussagen für nicht gymnasiale, allgemeinbildende Schulen können jedoch aus dem oben genannten Grund nicht getroffen werden. Darüber hinaus werden keine gesonderten Ergebnisse für die Berufs- und Förderschulen präsentiert, weil die dort getesteten Stichproben keine repräsentativen Aussagen über Ergebnisse dieser Schularten erlauben (vgl. Kapitel 1). Im Folgenden werden daher nur die Ergebnisse für die Schularten Gymnasium, Realschule, Hauptschule, Schule mit mehreren Bildungsgängen und Integrierte Gesamtschule deskriptiv dargestellt, weiterführende Analysen werden nur für das Gymnasium präsentiert.

Naturwissenschaftliche Kompetenz: Mittelwerte im Vergleich

Die Mittelwerte und Standardabweichungen der naturwissenschaftlichen Kompetenz von Schülerinnen und Schülern der oben genannten Schularten sind in Tabelle 6.3 dargestellt. Die beobachteten mittleren naturwissenschaftlichen Kompetenzen unterscheiden sich zwischen den Schularten beträchtlich. Schülerinnen und Schüler am Gymnasium erzielen erwartungsgemäß die höchsten Kompetenzwerte mit durchschnittlich 595 Punkten. Wie auch bei allen vorangegangenen PISA-Erhebungen erreichen Jugendliche, die eine Hauptschule besuchen, mit Abstand die niedrigsten Werte (437 Punkte) im Naturwissenschaftstest. Dieser Kennwert von Schülerinnen und Schülern, die eine Hauptschule besuchen, liegt mit mindestens 45 Punkten deutlich unter den Mittelwerten der Integrierten Gesamtschulen (492 Punkte) sowie der Schulen mit mehreren

Tabelle 6.3: Mittelwerte und Standardabweichungen der naturwissenschaftlichen Kompetenz in Deutschland nach Schulart

Schulart	<i>M</i>	(<i>SE</i>)	<i>SD</i>	(<i>SE</i>)
Hauptschule	437	(9.1)	74	(3.2)
Schule mit mehreren Bildungsgängen	482	(6.5)	80	(2.9)
Integrierte Gesamtschule	492	(6.1)	75	(2.8)
Realschule	522	(4.1)	71	(1.7)
Gymnasium	595	(2.5)	66	(1.5)
Gesamtstichprobe	524	(3.0)	95	(2.0)

Bildungsgängen (482 Punkte) und insbesondere unter dem Mittelwert der Realschulen (522 Punkte). Der Unterschied zwischen Gymnasien und Hauptschulen beträgt 158 Punkte und entspricht mehr als zwei Kompetenzstufen.

Die Streuung der naturwissenschaftlichen Kompetenz

Auch innerhalb der Schularten gibt es beträchtliche Streuungen der naturwissenschaftlichen Kompetenz. Für Schulen mit mehreren Bildungsgängen (80 Punkte) und Integrierte Gesamtschulen (75 Punkte) ist dies zu erwarten, da dort mehrere Bildungsgänge unterrichtet werden. Doch auch für die anderen Schularten betragen die Streuungen der naturwissenschaftlichen Kompetenz mindestens 66 Punkte, was beträchtliche Kompetenzunterschiede der Schülerinnen und Schüler widerspiegelt. Wie der Ländervergleich zeigt, kann die Streuung innerhalb der Schularten aber auch aufgrund erheblicher durchschnittlicher Leistungsunterschiede in den Naturwissenschaften zwischen den Ländern zustande kommen (vgl. Pant, Stanat, Schroeders et al., 2013).

Abbildung 6.6 zeigt, dass sich die Kompetenzverteilungen der Schularten im PISA-Naturwissenschaftstest überlappen. So erreichen an den Realschulen etwas mehr als 10 Prozent der Schülerinnen und Schüler ein Kompetenzniveau, das über dem Mittelwert der Gymnasien liegt. Ebenso zeigen etwas mehr als 10 Prozent der Schülerinnen und Schüler, die Hauptschulen besuchen, im Test ein Leistungsniveau, das über dem Mittelwert der Realschulen liegt. Umgekehrt müssen etwa 20 Prozent der Schülerinnen und Schüler an Realschulen unterhalb des Mittelwerts der Hauptschulen eingeordnet werden.

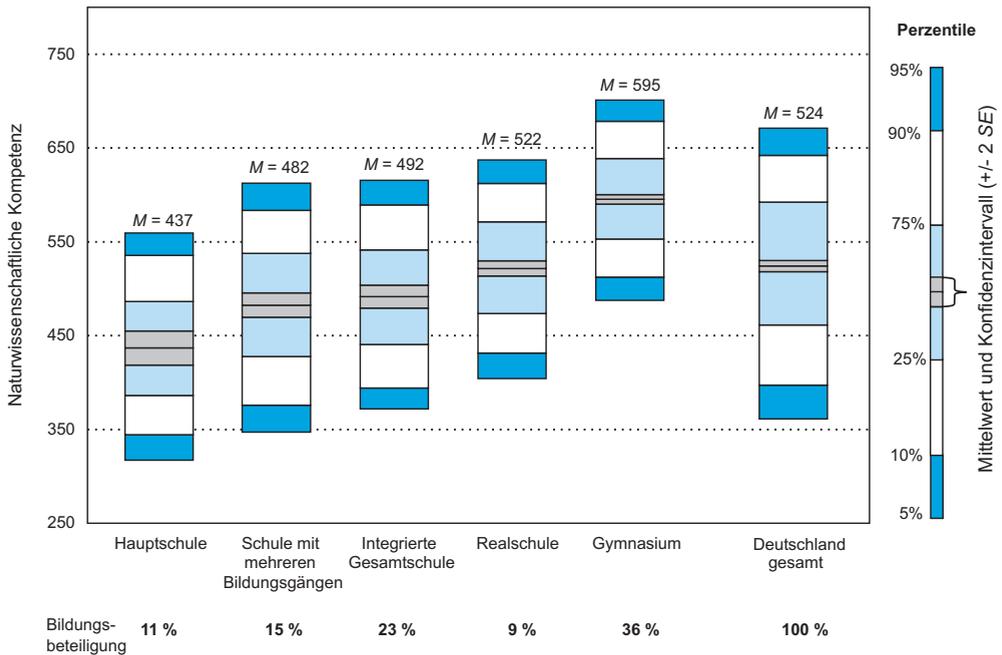


Abbildung 6.6: Perzentilbänder naturwissenschaftlicher Kompetenz in Deutschland nach Schulart

Verteilung auf die Stufen der naturwissenschaftlichen Kompetenz

Noch besser zu fassen sind die Unterschiede zwischen den Schularten, wenn man die Verteilungen der Schülerinnen und Schüler auf die Kompetenzstufen betrachtet (vgl. Abbildung 6.7). Fast 30 Prozent der Jugendlichen am Gymnasium erreichen die Kompetenzstufen V oder VI, während an der Hauptschule nur gut 5 Prozent der Jugendlichen überhaupt die Kompetenzstufe IV oder besser erreichen. An den Realschulen findet man ein breit gefächertes Kompetenzspektrum, dessen Schwerpunkt auf den Kompetenzstufen III und IV liegt. Jedoch erreichen auch knapp 6 Prozent der Schülerinnen und Schüler an den Realschulen die Kompetenzstufen V oder VI. Der Anteil der Jugendlichen, deren Kompetenz auf oder unter Kompetenzstufe I eingeordnet wurde, beträgt an der Realschule nicht einmal 6 Prozent und ist damit so gering wie in keiner anderen Schulart außer am Gymnasium. Beunruhigend ist, dass die Kompetenz von über 36 Prozent der Hauptschülerinnen und Hauptschüler auf Kompetenzstufe I oder darunter verortet ist. Diese Jugendlichen sind demnach nicht in der Lage, einfache naturwissenschaftliche Probleme oder Zusammenhänge zu verstehen oder aus Informationen Schlussfolgerungen zu ziehen.

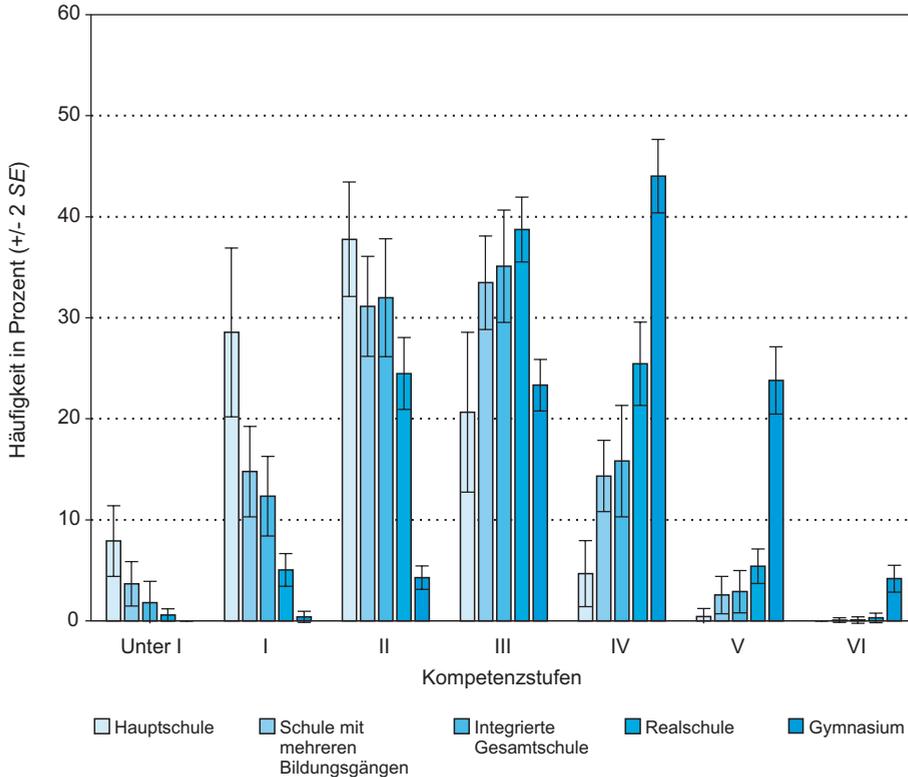


Abbildung 6.7: Prozentuale Anteile von Schülerinnen und Schülern auf den Stufen der naturwissenschaftlichen Kompetenz in Deutschland nach Schulart

6.3.2 Veränderung der naturwissenschaftlichen Kompetenz seit PISA 2006

Der Naturwissenschaftstest beruht in allen PISA-Erhebungsrounden auf einem übergreifenden theoretischen Ansatz, einer Rahmenkonzeption. Diese wurde 2006, als die Naturwissenschaften das erste Mal Hauptdomäne eines PISA-Zyklus war, im Vergleich zu 2000 und 2003 elaboriert und weiter ausdifferenziert (OECD, 2006). Die OECD hat diesen Zeitpunkt auch als Grundlage zur Normierung der Naturwissenschaftsskala herangezogen und den OECD-Mittelwert analog zu Lesen und Mathematik auf 500 Punkte sowie die Standardabweichung auf 100 Punkte festgelegt. Die Entwicklung der naturwissenschaftlichen Kompetenz lässt sich daher erst seit 2006 international vergleichen.

Deutschland hatte die Erhebung 2006 zusätzlich genutzt, um mit Hilfe einer nationalen Skalierung der Link-Aufgaben auch Vergleiche zu den beiden vorangegangenen PISA-Zyklen 2000 und 2003 zu ermöglichen (Carstensen, Prenzel & Baumert, 2008). Dabei zeigte sich ein leichter Anstieg der naturwissenschaftlichen Kompetenz zwischen den Jahren 2000 und 2003 mit einer Effektstärke von $d = 0.06$, der sich jedoch nicht sta-

tistisch signifikant absichern ließ (Prenzel et al., 2007). Im Vergleich von 2003 zu 2006 hingegen stieg die naturwissenschaftliche Kompetenz signifikant mit einer Effektstärke von $d = 0.12$ an. Damit konnten bereits für den Zeitraum von 2000 bis 2006 beträchtliche Fortschritte in der Entwicklung der naturwissenschaftlichen Kompetenz an Schulen in Deutschland belegt werden.

Die Ergebnisse für Deutschland im Naturwissenschaftstest seit PISA 2006 sind in Abbildung 6.8 dargestellt. Bei PISA 2006 erreichte Deutschland 516 Punkte und lag damit erstmals signifikant oberhalb des OECD-Durchschnitts. Drei Jahre später (2009) stieg die naturwissenschaftliche Kompetenz um vier Punkte an und lag erneut signifikant oberhalb des OECD-Durchschnitts, jedoch ließ sich der weitere Anstieg nicht als statistisch signifikant absichern. Bei PISA 2012 haben sich die Schülerinnen und Schüler in Deutschland im Vergleich zu 2009 zusätzlich um vier Punkte verbessert und liegen abermals signifikant über dem OECD-Durchschnitt. Zusammengefasst zeigt sich seit 2006 eine Zunahme der naturwissenschaftlichen Kompetenz von insgesamt 9 Punkten. Diese Verbesserung seit 2006 ist jedoch nicht groß genug, um sie statistisch als signifikanten Zuwachs absichern zu können. Die Veränderungen in den Testleistungen seit PISA 2006 können aber auf jeden Fall als eine Stabilisierung der naturwissenschaftlichen Kompetenz auf einem hohen, gegenüber dem OECD-Mittelwert überdurchschnittlichen Niveau bezeichnet werden. Ebenso positiv zu vermerken ist die Abnahme in der Leistungsstreuung (beginnend mit PISA 2000) und die Abnahme in den Anteilen der Schülerinnen und Schüler unterhalb der Kompetenzstufe II.

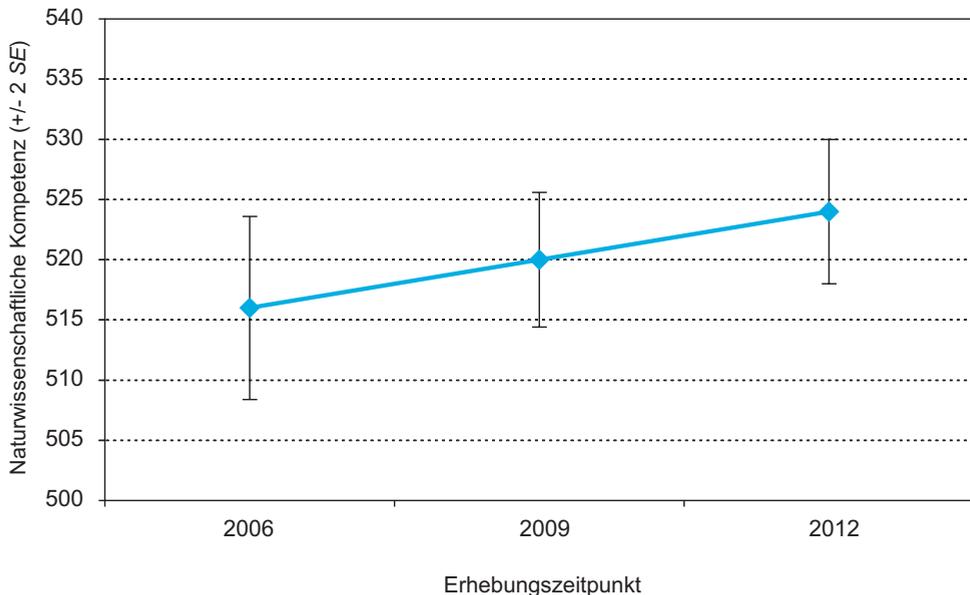


Abbildung 6.8: Naturwissenschaftliche Kompetenz in Deutschland von PISA 2006 bis PISA 2012

Betrachtet man die Veränderungen in den Naturwissenschaftsleistungen seit 2006 in den anderen OECD-Staaten, dann erlebte Polen seit 2009 den stärksten Zuwachs mit einer statistisch bedeutsamen Zunahme der mittleren naturwissenschaftlichen Kompetenz um 18 Punkte. Polen setzt somit eine positive Entwicklung fort, denn bereits zwischen den Jahren 2006 und 2009 verzeichnete das Land einen Anstieg um 28 Punkte und lag damit erstmals im Durchschnitt der OECD-Staaten. Bei PISA 2012 rangiert Polen nun in der Gruppe der Staaten, die statistisch bedeutsam besser abschneiden als der OECD-Durchschnitt. Weitere statistisch bedeutsame Zuwächse von zehn oder mehr Punkten im Vergleich zu 2006 zeigen sich in der Türkei (40 Punkte), in Italien (18 Punkte), in Israel und Korea (jeweils 16 Punkte), in Japan und Portugal (jeweils 15 Punkte) sowie in Irland (14 Punkte) und Estland (10 Punkte). Mindestens ebenso bemerkenswert sind große Abnahmen in den Mittelwerten für den Naturwissenschaftstest seit 2006. Diese belaufen sich in Schweden auf 19 Punkte, in Finnland auf 18 Punkte, in der Slowakischen Republik auf 17 Punkte, in Neuseeland auf 15 Punkte und in Island auf 13 Punkte.

Die Verteilung der fünfzehnjährigen Schülerinnen und Schüler in Deutschland auf die sechs Kompetenzstufen beschreibt, in welchen Bereichen des Leistungsspektrums sich die naturwissenschaftliche Kompetenz verändert hat. Besonders interessant sind dabei Veränderungen im unteren und oberen Bereich der Kompetenzverteilung. Im Vergleich zu 2006 hat sich der Anteil der Jugendlichen, deren naturwissenschaftliche Kompetenz unterhalb der Stufe II liegt, signifikant um 3.2 Prozent verringert. Der Anteil derer, die sich im oberen Leistungsbereich auf den Kompetenzstufen V und VI befinden, ist um 0.4 Prozent gestiegen. Dieser Anstieg kann jedoch nicht als statistisch bedeutsam abgesichert werden. Damit zeigt sich für Deutschland ein ambivalentes Bild. Einerseits ist es gelungen, den Anteil der Schülerinnen und Schülern auf den unteren Kompetenzstufen zu verringern, andererseits konnte der Anteil der Jugendlichen auf den oberen Kompetenzstufen kaum ausgebaut werden.

Hier stellt sich die Frage, inwieweit das Gymnasium in Deutschland die Möglichkeiten eines differenzierten Schulsystems für die Förderung von Talenten und einer Leistungsspitze ausschöpft. Bemerkenswert ist die Tatsache, dass eine ganze Reihe von OECD-Staaten mit nicht differenzierenden Schulsystemen in der Sekundarstufe I größere (oder zumindest statistisch vergleichbare) Anteile von Jugendlichen auf den Kompetenzstufen V und VI nachweisen können (z. B. Japan, Finnland, Australien, Neuseeland; vgl. Abbildung 6.4). Noch dramatischer erscheint die Differenz, wenn Vergleiche zu den Spitzengruppen in Shanghai, Hongkong oder Singapur angestellt werden (vgl. Tabelle 6 im Anhang).

Gegenüber PISA 2006 hat sich der für den Naturwissenschaftstest gemessene Mittelwert der Gymnasien geringfügig von 598 Punkten auf 595 Punkte verändert. Mit einer Vergrößerung der Bildungsbeteiligung von 28.3 Prozent des Altersjahrgangs bei PISA 2006 zu 36.0 Prozent bei PISA 2012 ist es wohl gelungen, das durchschnittliche Leistungsniveau in den Naturwissenschaften zu erhalten. Die Anteile der Schülerinnen und Schüler, die in Deutschland den obersten Kompetenzstufen V und IV zugeordnet werden, sind jedoch seit 2006 nur geringfügig gestiegen (von 11.8 Prozent auf 12.2 Pro-

zent). Auch bei einer Bildungsbeteiligung von 36,0 Prozent befindet sich ein Großteil der leistungsstärksten Schülerinnen und Schüler am Gymnasium. Der Nachweis, dass mit dieser Differenzierung eine bessere Förderung von Spitzengruppen betrieben werden kann, muss im Bereich der naturwissenschaftlichen Bildung noch erbracht werden.

6.4 Diskussion und Ausblick

Insgesamt zeigt sich seit PISA 2000 in Deutschland eine sehr positive Entwicklung im Bereich der naturwissenschaftlichen Grundbildung. Die erste PISA-Erhebung konfrontierte Deutschland mit dem Befund einer im OECD-Vergleich unterdurchschnittlichen Kompetenz der Schülerinnen und Schüler in den Naturwissenschaften. Bereits bei PISA 2006 lag das Leistungsniveau über dem Durchschnitt der OECD-Staaten. Mit PISA 2009 und PISA 2012 setzt sich der positive Entwicklungstrend fort, die Mittelwerte sind seit PISA 2006 um weitere 9 Punkte gestiegen. Damit haben sich die Leistungen in den Naturwissenschaften auf einem im internationalen Vergleich hohen Niveau stabilisiert. Kein anderes OECD-Land hat bisher eine vergleichbar positive Entwicklung im Bereich der naturwissenschaftlichen Kompetenz erlebt wie Deutschland. Dennoch zeigt der internationale Vergleich, dass weitere Verbesserungen in der naturwissenschaftlichen Grundbildung möglich und entsprechend anzustreben sind.

Die Erfolgsgeschichte, die Deutschland mit diesem Bericht für die Entwicklung naturwissenschaftlicher Kompetenz seit TIMSS und PISA 2000 bescheinigt wird, kann als Ermunterung verstanden werden, die weiteren Herausforderungen anzugehen. Wesentlich zu dieser Entwicklung beigetragen haben die vielfältigen Anstrengungen zur Qualitätssicherung und Unterrichtsentwicklung im Bereich der Naturwissenschaften. Die Einführung länderübergreifender Bildungsstandards für den Mittleren Schulabschluss (KMK, 2005a, 2005b, 2005c), die festlegen, über welche naturwissenschaftlichen Kompetenzen Jugendliche am Ende ihrer Pflichtschulzeit verfügen sollen, kann als Meilenstein für eine zielbezogene und in eine Förderstrategie eingebundene Qualitätssicherung gelten (KMK, 2010). Programme wie zum Beispiel SINUS (Ostermeier et al., 2010; Prenzel et al., 2005) oder die sogenannten „Kontext“-Projekte (Bayrhuber et al., 2007; Demuth et al., 2008; Mikelskis-Seifert & Duit, 2010) haben sehr konkret gezeigt, wie in enger Zusammenarbeit von Schulen und mit Unterstützung durch Know-how aus der empirischen Bildungsforschung Unterricht zielorientiert weiterentwickelt werden kann.

Betrachtet man die Module des SINUS-Programms, dann finden sich dort nach wie vor Ansatzpunkte für die weitere Verbesserung der naturwissenschaftlichen Grundbildung, etwa durch fachdidaktisch begründete Formate des Experimentierens (z. B. durch „Inquiry based science teaching“), durch eine noch bessere Abstimmung der Unterrichtsinhalte über die Schul- und Jahrgangsstufen, durch eine gezielte Sicherung des Grundwissens und durch neue Prüfungsformen. Module und Handreichungen alleine, so die Erfahrung aus allen Qualitätsentwicklungsprogrammen, reichen für die Unterrichtsentwicklung nicht aus. Schulnahe Fortbildungen, Ermunterungen und Stärkun-

gen der Fachgruppen zur gemeinsamen Unterrichtsentwicklung, eine weitere Profilierung der Schulen oder die Nutzung von Ganztagsangeboten, um naturwissenschaftliche Erfahrungen außerhalb des regulären Unterrichts zu ermöglichen, sind wichtige Bedingungen für flächendeckende Fortschritte.

Die Befunde, die in diesem Kapitel berichtet wurden, weisen insbesondere auf zwei große Herausforderungen hin. Die erste besteht darin, den Anteil von Schülerinnen und Schülern, die nur über rudimentäre naturwissenschaftliche Kenntnisse verfügen, weiter zu verringern. Zwar sind die Anteile der Jugendlichen unterhalb der zweiten Kompetenzstufe über die PISA-Runden erfreulich gesunken, doch sind sie in Deutschland mit derzeit 12,2 Prozent immer noch zu hoch. Das Motto *Naturwissenschaften für alle* dient nicht nur dazu, Berufsfähigkeit zu gewährleisten. Naturwissenschaftliches Verständnis ist essentiell für eine reflektierte und erfolgreiche Lebensführung im Alltag sowie für eine Beteiligung an gesellschaftlichen Debatten und Entscheidungen. Deshalb lohnt es sich, die Anstrengungen zur breiten Förderung eines grundlegenden naturwissenschaftlichen Verständnisses fortzusetzen und zu verstärken. Dieses Anliegen schließt ein, insbesondere die Motivation und das Interesse an den Naturwissenschaften zu stärken (Krapp & Prenzel, 2011).

Die Berücksichtigung von Motivation und Interesse ist ebenso wichtig für die zweite, sehr große Herausforderung. Sie richtet sich darauf, den Anteil von jungen Menschen zu vergrößern, die sich durch ein besonderes Engagement für die Naturwissenschaften und das Bedürfnis, diese noch besser zu verstehen, auszeichnen. Die Naturwissenschaften spielen eine Schlüsselrolle für die Forschung und die Wirtschaft, für das gesellschaftliche Wohlergehen und für die Sicherung der Lebensgrundlagen für nachfolgende Generationen. Für die Naturwissenschaften brauchen wir „kluge Köpfe“, denn es gilt sehr komplexe Sachverhalte zu durchdringen und zu verstehen. Diese klugen Köpfe sollten möglichst früh die Möglichkeit erhalten, ihre Potentiale zu entfalten. Die Befunde aus PISA 2012 legen, auch im Kontext der Ergebnisse früherer Erhebungen, nahe, dass die Förderung des naturwissenschaftlichen Nachwuchses in Deutschland deutlich verstärkt und verbessert werden muss. Gefordert ist hier insbesondere das Gymnasium.

Literatur

- Adams, R. & Wu, M. (2007). A generalized form of the Rasch model. In M. v. Davier & C. H. Carstensen (Hrsg.), *Multivariate and mixture distribution Rasch models. Extensions and applications* (Statistics for social and behavioral sciences, 1. Aufl., S. 57–76). New York: Springer.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I. et al. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Bayrhuber, H., Bögeholz, S., Elster, D., Hammann, M., Hößle, C., Lücken, M. et al. (2007). *Biologie im Kontext – Ein Programm zur Kompetenzförderung durch Kontextorientie-*

- rung im Biologieunterricht und zur Unterstützung von Lehrerprofessionalisierung. *MNU*, 60, 282–286.
- Beaton, A. E., Mullis, I. V., Martin, M. O., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1997). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (Hrsg.). (1997). *Gutachten zur Vorbereitung des Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“*. Materialien zur Bildungsplanung und zur Forschungsförderung, Heft 60. Bonn: BLK.
- Bybee, R. & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, 33, 7–26.
- Carstensen, C. H., Prenzel, M. & Baumert, J. (2008). Trendanalysen in PISA: Wie haben sich die Kompetenzen in Deutschland zwischen PISA 2000 und PISA 2006 entwickelt? In M. Prenzel & J. Baumert (Hrsg.), *Vertiefende Analysen zu PISA 2006* (Zeitschrift für Erziehungswissenschaft: Sonderheft, Bd. 10.2008, S. 11–34). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Demuth, R. (Hrsg.). (2008). *Chemie im Kontext. Von der Innovation zur nachhaltigen Verbreitung eines Unterrichtskonzepts*. Münster: Waxmann.
- Fensham, J. P. (1985). Science for all. *Journal of Curriculum Studies*, 17, 415–435.
- Fischer, C., Dedekind, B., Rieck, K., Trepke, F. & Prenzel, M. (2010). *Abschlussbericht zum Programm SINUS-Transfer Grundschule*. Kiel: IPN. Zugriff am 20.09.2013. Verfügbar unter http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_SGS/abschlussbericht.pdf.
- Fischer, C., Döring, B., Rieck, K., Dedekind, B., Trepke, F., Kobarg, M. et al. (2013). *Ergebnisse und Erfahrungen des dritten Programmjahres. August 2011 bis Juli 2012. Berichterstattung des Programms SINUS an Grundschulen*. Kiel: Zugriff am 20.09.2013. IPN. Verfügbar unter http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_SGS/Dritter_ZB_SaG_fuer_Web.pdf.
- Fischer, C. & Prenzel, M. (2009). *Skizze für ein länderübergreifendes Anschlussvorhaben SINUS an Grundschulen*. Kiel: IPN. Zugriff am 20.09.2013. Verfügbar unter http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_SGS/Programmskizze_SaG_fuer_Web.pdf.
- Hahn, I., Schöps, K., Rönnebeck, S., Martensen, M., Hansen, S., Saß, S. et al. (2013). Assessing scientific literacy over the lifespan – A description of the NEPS science framework and the test development. *Journal for Educational Research Online*, 5, 110–138.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005a). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den mittleren Bildungsabschluss (Beschluss vom 16. Dezember 2004)*. München: Wolters Kluwer.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005b). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Chemie für den mittleren Bildungsabschluss (Beschluss vom 16. Dezember 2004)*. München: Wolters Kluwer.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005c). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Physik für den mittleren Bildungsabschluss (Beschluss vom 16. Dezember 2004)*. München: Wolters Kluwer.

- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2010). *Förderstrategie für leistungsschwächere Schülerinnen und Schüler* (Beschluss der Kultusministerkonferenz vom 04.03.2010). Zugriff am 20.09.2013. Verfügbar unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2010/2010_03_04-Foerderstrategie-Leistungsschwaechere.pdf
- Krapp, A. & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education*, 33, 27–50.
- Mikelskis-Seifert, S. & Duit, R. (2010). *Physik im Kontext – Konzepte, Ideen, Materialien für effizienten Physikunterricht*. Seelze: Friedrich Verlag.
- OECD. (2006). *Assessing scientific, reading and mathematical literacy. A framework for PISA 2006*. Paris: OECD.
- OECD. (2007). *PISA 2006. Science competencies for tomorrow's world*. Paris: OECD.
- OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
- Osborne, J. (2007). Science education for the twenty first century. *Eurasia Journal of Mathematics, Science & Technology Education*, 3, 173–184.
- Osborne, J. & Dillon, J. (2008). *Science education in Europe: Critical reflections. A report to the Nuffield Foundation*. London: King's College.
- Ostermeier, C., Prenzel, M. & Duit, R. (2010). Improving science and mathematics instruction: The SINUS project as an example for reform as teacher professional development. *International Journal of Science Education*, 32, 303–327.
- Pant, H. A., Stanat, P., Pöhlmann, C., Hecht, M., Jansen, M., Kampa, N. et al. (2013). Der Blick in die Länder. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 159–248). Münster: Waxmann.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Prenzel, M., Carstensen, C. H., Senkbeil, M., Ostermeier, C. & Seidel, T. (2005). Wie schneiden SINUS-Schulen bei PISA ab? Ergebnisse der Evaluation eines Modellversuchsprogramms. *Zeitschrift Für Erziehungswissenschaft*, 8, 540–562.
- Prenzel, M., Friedrich, A. & Stadler, M. A. (2009). *Von SINUS lernen. Wie Unterrichtsentwicklung gelingt* (Sinus-Transfer, 1. Aufl.). Seelze-Velber: Klett/Kallmeyer.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P. & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 192–250). Opladen: Leske + Budrich.
- Prenzel, M., Schöps, K., Rönnebeck, S., Senkbeil, M., Walter, O., Carstensen, C. et al. (2007). Naturwissenschaftliche Kompetenz im internationalen Vergleich. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme et al. (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 61–124). Münster: Waxmann.
- Roberts, D. (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 PISA 2009. In E. Klieme, C. Artelt, J. Har-

- tig, N. Jude, O. Köller, M. Prenzel et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 177–198). Münster: Waxmann.
- Rost, J., Walter, O., Carstensen, C., Senkbeil, M. & Prenzel, M. (2004). Naturwissenschaftliche Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs* (S. 111–146). Münster: Waxmann.
- Walter, O., Senkbeil, M., Rost, J., Carstensen, C. & Prenzel, M. (2006). Die Entwicklung der naturwissenschaftlichen Kompetenz von der neunten zur zehnten Klassenstufe: Deskriptive Befunde. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 88–118). Münster: Waxmann.
- Weinert, F. E. (1999). *Concepts of competence (Contribution within the OECD project definition and selection of competencies: Theoretical and conceptual foundations (DeSeCo))*. Neuchatel: DeSeCo. Zugriff am 20.09.2013. Verfügbar unter <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.111.1152&rep=rep1&type=pdf>.

7 Lesekompetenz in PISA 2012: Veränderungen und Perspektiven

Katharina Hohn, Anja Schiepe-Tiska, Christine Sälzer
und Cordula Artelt

Lesen ist eine Schlüsselkompetenz in der Wissensgesellschaft. Der größte Teil des Wissens wird in Texten unterschiedlicher Art – mündlich, schriftlich, in Dokumenten oder im Internet – gespeichert und transportiert. Über Lesen kann die in Texten enthaltene Information erschlossen und in Wissen umgesetzt werden. Lesekompetenz ist somit auch eine notwendige Voraussetzung für Lernen und für die gesellschaftliche wie kulturelle Teilhabe. Als solche muss sie unterschiedlichen Textsorten und Leseanlässen gerecht werden. So sollen zum Beispiel Zeitungsartikel, amtliche Schreiben, persönliche Briefe oder Verträge genauso erfolgreich erfasst werden wie Fahrpläne, Schulbuchtexte oder Einkaufslisten. Beim Lesen geht es dabei nicht nur um das Entziffern von Texten, sondern auch um das Verstehen und Erkennen der wesentlichen Aussagen sowie das Interpretieren und kritische Reflektieren möglicher Botschaften. Entsprechend reicht die Lesekompetenz, so wie sie bei PISA erhoben wird, weit über eine einfache, technische Lesefertigkeit hinaus. Ein Verdienst der PISA-Rahmenkonzeption zur Lesekompetenz besteht unter anderem darin, auf verschiedene Facetten von Lesekompetenz hinzuweisen, die für eine Teilhabe am gesellschaftlichen Leben und für Bildungsprozesse von Bedeutung sind. Vor diesem Hintergrund waren die ernüchternden Ergebnisse, die in Deutschland bei PISA 2000 im Test zur Lesekompetenz erzielt wurden, besonders aufrüttelnd. Im Unterschied zu den Testbereichen Mathematik und Naturwissenschaften gab es zudem für die Lesekompetenz in Deutschland keine „Warnung“ durch eine vorangegangene Untersuchung wie TIMSS, die auf Schwächen beim Lesen hingewiesen hätte (vgl. Kapitel 2 und 6; vgl. auch Baumert et al., 1997). Umso größer war der „Schock“, als bei PISA 2000 die Leseleistungen der Fünfzehnjährigen in Deutschland deutlich unter dem Durchschnitt der OECD-Staaten lagen (Artelt, Stanat, Schneider & Schiefele, 2001).

Auf PISA 2000 reagierend formulierte die KMK (2002, S. 6) sieben vorrangige Handlungsfelder im Bildungsbereich, die zum Beispiel mit „Maßnahmen zur Verbesserung der Sprachkompetenz“, „durchgängiger Verbesserung der Lesekompetenz“ oder auch „Maßnahmen zur wirksamen Förderung bildungsbenachteiligter Kinder“ auf die Behebung von Leseschwächen zielten. Verschiedene Initiativen – vorwiegend auf der Ebene einzelner Länder – zur Förderung der Lesekompetenz folgten, wie beispielsweise „Bücherwürmer in NRW“, „Lust auf Lesen“ in Thüringen, „Lesen macht stark“ in Schles-

wig-Holstein oder auch „Die BremerLeseLust“, die vorrangig in der Grundschule angesiedelt waren beziehungsweise sind (vgl. Artelt et al., 2007; vgl. auch Redder et al., 2011). Eine ebenfalls notwendige Förderung, die über die Grundschulzeit hinausgeht, verlangte jedoch ein Umdenken, da die Entwicklung einer hinreichenden Lesekompetenz häufig als Aufgabe der Grundschulen angesehen wurde (vgl. Staatsinstitut für Schulqualität und Bildungsforschung, 2005).

Dass auch jenseits der Grundschule eine kontinuierliche und weiterführende Leseförderung erforderlich ist, belegen unter anderem Befunde der internationalen Schulleistungstudie IGLU (*Internationale Grundschul-Lese-Untersuchung*), in der die Lesekompetenz am Ende der vierten Jahrgangsstufe erfasst wird. Hier zeigte sich wiederholt, dass die Leseleistungen der Schülerinnen und Schüler am Ende der Grundschulzeit in Deutschland (statistisch signifikant) über dem mittleren internationalen Leistungsniveau liegen, auch bei Berücksichtigung unterschiedlicher Konstellationen von Vergleichsstaaten (z. B. Bos, Tarelli, Bremerich-Vos & Schwippert, 2012). Damit stellte sich bei der Interpretation der PISA-Ergebnisse immer wieder die Frage, ob die in der Grundschule begonnene Leseförderung in der Sekundarstufe gezielt und wirksam fortgesetzt wird.

Ausgehend von einem Verständnis von Lesekompetenz als fächerübergreifender Fähigkeit, die in allen Unterrichtsfächern gefordert wird und gefördert werden sollte, wurde mit dem KMK-Projekt „ProLesen – Auf dem Weg zur Leseschule“ von 2008 bis 2010 ein deutschlandweiter Versuch gestartet, Konzepte und Materialien zur fächerübergreifenden Förderung der Lesekompetenz zusammenzustellen (Bayerisches Staatsministerium für Unterricht und Kultus und Staatsinstitut für Schulqualität und Bildungsforschung, 2010). Die konkrete Umsetzung von umfassenden Maßnahmen zur Leseförderung, aber auch zur Sprachförderung und Sprachdiagnostik, soll allerdings erst jetzt deutschlandweit durch das von Bund und Ländern gemeinsam getragene Projekt BISS (*Bildung durch Sprache und Schrift*) erfolgen (vgl. Schneider et al., 2013).

Ein weiterer zentraler Ansatzpunkt zur Förderung von Lesekompetenz nach PISA 2000 ist in der Festlegung von Zielsetzungen und Kompetenzanforderungen (hier für das Fach Deutsch) in Form von länderübergreifenden Bildungsstandards zu sehen, die verbindlich eingeführt wurden. In den Standards der Kultusministerkonferenz für den Mittleren Schulabschluss (KMK, 2004, S. 13f.) wird das Lesen als ein zentraler Kompetenzbereich im Fach Deutsch herausgestellt, welcher Fähigkeiten wie „Strategien zum Leseverstehen kennen und anwenden“ oder auch „Texte verstehen und nutzen“ umfasst.

Das bundesweite „Wachrütteln“ durch PISA 2000 und die damit einhergehende Etablierung von Förderansätzen und bundesweiten Bildungsstandards schien erste, wenn auch kleine Fortschritte mit sich zu bringen. So erreichten die Fünfzehnjährigen in den folgenden PISA-Erhebungsrunden (2003, 2006, 2009) Testergebnisse, die nunmehr im mittleren Leistungsbereich der OECD-Staaten lagen (Schaffner, Schiefele, Drechsel & Artelt, 2004; Drechsel & Artelt, 2007; Naumann, Artelt, Schneider & Stanat, 2010). Dennoch verbesserte sich die Lesekompetenz der Jugendlichen in Deutschland im Vergleich zu den anderen Kompetenzbereichen eher schleppend. Anlass zur Sorge gaben insbe-

sondere die immer noch hohen Anteile von Fünfzehnjährigen, deren Lesekompetenz auf den unteren Kompetenzstufen lag.

Dieses Kapitel geht der Frage nach, wie die in PISA 2012 getestete Lesekompetenz der fünfzehnjährigen Schülerinnen und Schüler in Deutschland im internationalen Vergleich ausgeprägt ist, ob Fortschritte in der Entwicklung seit PISA 2000 zu verzeichnen sind und worin weitere Herausforderungen bestehen. Dazu wird zunächst näher erläutert, was unter Lesekompetenz zu verstehen ist und wie diese bei PISA 2012 erfasst wurde. Anschließend werden die Ergebnisse des internationalen Vergleichs vorgestellt, ehe genauer auf die Befunde für die Schülerinnen und Schüler in Deutschland eingegangen wird.

7.1 Lesekompetenz in PISA 2012

Im Mittelpunkt von PISA steht die Frage, inwieweit Jugendliche gegen Ende der Pflichtschulzeit über grundlegende Kompetenzen verfügen, die für ihre weitere Entwicklung und gesellschaftliche Teilhabe relevant sind. Die bei PISA untersuchte Lesekompetenz bezieht sich dabei auf die Fähigkeit einer Person, geschriebene Texte zu verstehen, sie zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, eigenes Wissen und eigene Potentiale zu entwickeln und um aktiv an der Gesellschaft teilzuhaben (OECD, 2013).

Die Entwicklung der Lesekompetenz ist nicht auf die ersten Schuljahre beschränkt. Im Verlauf der Bildungsbiographie steigen die Anforderungen, weil immer wieder neue und anspruchsvollere Textsorten erschlossen werden müssen. Die Lesekompetenz stellt nicht nur eine Grundvoraussetzung für allgemeine Bildungserfolge und lebenslanges Lernen dar, sondern sie ist auch von grundlegender Bedeutung für den beruflichen Erfolg. So konnte zum Beispiel anhand kanadischer Daten im Rahmen des *Youth in Transition Survey* (YITS) gezeigt werden, dass die Lesekompetenz von Fünfzehnjährigen mit deren Bildungserfolgen im Alter von 21 Jahren zusammenhängt (Bussière, Hébert & Knighton, 2010).

Nachdem die Lesekompetenz bei PISA 2009 zum zweiten Mal Schwerpunkt der Erhebung war, wurde sie 2012 als Nebendomäne erfasst. In den folgenden Abschnitten wird die Konzeption der Lesekompetenz näher beschrieben, auf welcher der Lesekompetenztest von PISA 2012 aufbaut (OECD, 2013).

Rahmenkonzeption der Domäne Lesen

Die theoretische Rahmenkonzeption, die allen in PISA untersuchten Domänen zu Grunde gelegt wird, wurde im Bereich der Lesekompetenz für PISA 2012 unverändert aus PISA 2009 übernommen. Für Lesen als Schwerpunktbereich bei PISA 2009 wurde damals die Konzeption überarbeitet und ergänzt sowie der Aufgabenpool aufgefrischt (OECD, 2009). In PISA 2012 wird entsprechend die Lesekompetenz als mehrdimen-

sionale, facettenreiche Fähigkeit betrachtet, die weit über das Dekodieren von Informationen und wörtliche Interpretationen hinausgeht. Hierbei werden unterschiedliche Leseverstehensanforderungen (*Aspekte des Lesens*), Leseanlässe (*Situationen*) und *Textformate und Texttypen* unterschieden.

Die *Aspekte des Lesens* beschreiben Anforderungen, welche durch die Texte und die hierauf bezogenen Aufgaben an den Leser gestellt werden. Dabei werden drei Anforderungen unterschieden, die verschiedene kognitive Prozesse erfordern sollen (vgl. OECD, 2013):

- Informationen suchen und extrahieren
- textbezogen kombinieren und interpretieren
- reflektieren und bewerten

Unter *Informationen suchen und extrahieren* wird verstanden, dass die Lesenden unterschiedliche Informationen in Texten lokalisieren und dem Text entnehmen können. *Textbezogenes Kombinieren und Interpretieren* als zweite Anforderungskategorie umfasst Prozesse, die zur tieferen Verarbeitung des Gelesenen benötigt werden. Kombinieren bezieht sich darauf, dass der Text als großes Ganzes verstanden wird. Die Leser stellen zum Beispiel Verbindungen zwischen im Text enthaltenen Informationen her, um dadurch die (tiefere) Bedeutung des Textes zu erfassen; dies kann etwa durch Herausarbeiten von Ähnlichkeiten und Unterschieden geschehen, durch Vergleiche oder durch das Durchdenken von Ursache- und Wirkungsketten. Beim Interpretieren wird von den Lesenden unter anderem erwartet, dass sie in der Lage sind „zwischen den Zeilen“ zu lesen, und somit die Bedeutung des Textes auch dann erfassen, wenn nicht alle Informationen explizit im Text vorhanden sind oder mögliche Botschaften und Absichten der Autorin oder des Autors nicht direkt bewusst gemacht werden. Beim *Reflektieren und Bewerten*, der dritten Anforderungskategorie, sollen die Lesenden die aus dem Text gewonnenen Informationen mit eigenem Wissen, eigenen Vorstellungen und Einstellungen in Beziehung setzen (*reflektieren*) und gegebenenfalls evaluieren (*bewerten*).

Leseanlässe oder *Situationen* beziehen sich auf den Kontext, in dem das Lesen stattfindet und damit auch auf Anlässe, Anforderungen und Zwecke des Lesens (z. B. im Unterricht). Um eine möglichst große Bandbreite an verschiedenen Situationen und Leseanlässen abzubilden, werden in Anlehnung an das *Common European Framework of Reference* (Council of Europe, 1996) vier Situationen unterschieden:

- private Situationen (z. B. persönliche Briefe)
- öffentliche Situationen (z. B. offizielle Dokumente)
- bildungsbezogene Situationen (z. B. Lehrbuchtexte)
- berufsbezogene Situationen (z. B. Stellenausschreibungen)

Diese vier Situationen können nicht als völlig unabhängig voneinander betrachtet werden, denn sie überlappen einander. So werden zum Beispiel im Rahmen des Deutsch-

unterrichts (bildungsbezogene Situation) auch allgemeine, belletristische Werke behandelt, die die Schülerinnen und Schüler durchaus in ihrer Freizeit lesen würden (private Situation). Durch die Berücksichtigung der vier Situationen bei der Erstellung von Testaufgaben sehen sich die Jugendlichen bei der Aufgabebearbeitung mit unterschiedlichen sprachlichen Mitteln, Themen, Formaten und Zwecken konfrontiert.

Textformat bezieht sich auf die Differenzierung zwischen Texten mit unterschiedlichem formalem Aufbau. Grundlegend ist in diesem Zusammenhang die Unterscheidung von kontinuierlichen Texten (z. B. in Zeitungsartikeln, Erzählungen, Anleitungen oder Romanen) und nichtkontinuierlichen Texten (z. B. bei Diagrammen, Listen, Tabellen oder Zeitplänen). Kontinuierliche Texte zeichnen sich im Gegensatz zu nichtkontinuierlichen Texten durch zusammenhängende Sätze aus, die in größere Strukturen gebettet werden können, wie Absätze oder Kapitel. Die beiden Textformate sind nicht immer vollständig voneinander getrennt, gemischte Formate sind möglich. So kann ein längerer Prosaabschnitt zum Beispiel durch zusätzliche Tabellen oder Diagramme begleitet werden.

Der *Texttyp* bezieht sich auf die Unterscheidung von verschiedenen Textsorten hinsichtlich ihrer vorherrschenden Merkmale. Hier werden Beschreibungen, Erzählungen, Darlegungen, Argumentationen, Anleitungen und Transaktionen (Austausch von Informationen) differenziert.

Der Lesetest in PISA 2012

Bei der Erstellung der Aufgaben zur Erfassung der Lesekompetenz wurden die drei oben beschriebenen grundlegenden Gesichtspunkte (Aspekte des Lesens, Situationen, Textformate und Texttypen) systematisch berücksichtigt. Dadurch soll eine möglichst breite Erfassung der Lesekompetenz als facettenreiche Fähigkeit sichergestellt werden. Da die Lesekompetenz 2012 als Nebendomäne erfasst wurde, handelt es sich bei den verwendeten Aufgaben um solche, die bereits in früheren PISA-Erhebungsrunden zum Einsatz kamen, aber weiter geheim gehalten werden. Insgesamt wurden 13 Aufgaben mit 44 Items zur Erfassung der Lesekompetenz für PISA 2012 ausgewählt, die auf acht der insgesamt 13 Testhefte verteilt wurden (vgl. Kapitel 10).

Für die Bearbeitung der Testaufgaben mussten die Schülerinnen und Schüler verschiedene Antwortformate bewältigen. Zum einen kamen Multiple-Choice-Items zum Einsatz, bei denen die Schülerinnen und Schüler eine vorgegebene Antwort auswählen sollten, die sie für richtig erachteten. Zum anderen wurden offene Antwortformate vorgelegt, bei denen die Schülerinnen und Schüler selbstständig Antworten formulieren sollten.

Eine Differenzierung in Subskalen entsprechend der drei Aspekte des Lesens und nach kontinuierlich und nicht-kontinuierlich geschriebenen Texten (Textformate), wie sie 2009 vorgenommen wurde, ist 2012 nicht möglich, da eine geringere Aufgabenmenge verwendet wurde, was lediglich die Bildung eines Gesamtwertes für die Lesekompetenz der Schülerinnen und Schüler ermöglicht. Die Gesamtskala wurde so gebildet, dass ein

Vergleich der Lesekompetenzen der Schülerinnen und Schüler bei PISA 2012 mit dem Abschneiden der Schülerinnen und Schüler in früheren Erhebungsrunden möglich ist.

Die Kompetenzstufen

Wie in den vorangegangenen PISA-Erhebungsrunden wurden auch 2012 die Antworten der Schülerinnen und Schüler nach Modellen der Item-Response-Theorie skaliert (vgl. z. B. van der Linden, Veldkamp & Carlson, 2004; Adams & Wu, 2007; vgl. Kapitel 10). Die Darstellung der Leistungen im Lesekompetenztest erfolgt auf einer Skala, die bei PISA 2012 einen Mittelwert von 500 und eine Streuung (Standardabweichung) von 100 aufweist. Ferner können die Testleistungen der Schülerinnen und Schüler anhand sogenannter Kompetenzstufen qualitativ beschrieben werden. Dabei werden bei PISA 2012, wie auch schon in der vorangegangenen PISA-Erhebungsrunde 2009, sieben Kompetenzstufen unterschieden. Diese Differenzierung ermöglicht, im Vergleich zu den ursprünglich fünf Kompetenzstufen in den Erhebungen von 2000–2006, eine detailliertere Beschreibung der Leistungen von Schülerinnen und Schülern am oberen und am unteren Ende des Fähigkeitsspektrums der Lesekompetenz (Naumann et al., 2010). Kompetenzstufen veranschaulichen Anforderungen, die auf dem jeweiligen Niveau mit einer bestimmten Wahrscheinlichkeit gemeistert werden. Schülerinnen und Schüler auf einer bestimmten Kompetenzstufe meistern Anforderungen einer darunter liegenden Kompetenzstufe mit hoher Wahrscheinlichkeit. Mit hoher Wahrscheinlichkeit scheitern sie an den Aufgaben und Anforderungen, die einer höheren Kompetenzstufe zugeordnet sind. Der Abstand zwischen den einzelnen Kompetenzstufen umfasst jeweils etwa 73 Leistungspunkte. Ein Überblick über die Anforderungen und Texte auf den einzelnen Kompetenzstufen der Lesekompetenz findet sich in Tabelle 7.1 (vgl. Naumann et al., 2010, S. 28).

Zur besseren Veranschaulichung der Kompetenzstufen ist in Abbildung 7.1 die Aufgabeneinheit „Das Schauspiel sei das Werkzeug“ mit dem Text und den dazugehörigen Aufgaben (Items) auf unterschiedlichen Kompetenzstufen dargestellt. Bei dem Text handelt es sich um den Anfang eines Theaterstücks des ungarischen Dramatikers Ferenc Molnár; diese Information war in den Testheften am Beginn des Blocks mit dieser Aufgabe enthalten. Der Auszug aus „Das Schauspiel sei das Werkzeug“ ist ein kontinuierlich geschriebener Text, der – obwohl sicher auch mit Bezügen zu bildungsbezogenen Leseanlässen – in PISA einem persönlichen Leseanlass zugeordnet wurde.

Tabelle 7.1: Überblick über die typischen Anforderungen pro Kompetenzstufe (Ib–VI)

Kompetenzstufe	Wozu die Schülerinnen und Schüler auf der jeweiligen Kompetenzstufe im Allgemeinen in der Lage sind
VI > 698 Punkte	Jugendliche auf dieser Stufe können Schlussfolgerungen, Vergleiche und Gegenüberstellungen detailgenau und präzise anstellen. Dabei entwickeln sie ein volles und detailliertes Verständnis eines oder mehrerer Texte und verbinden dabei unter Umständen gedanklich Informationen aus mehreren Texten miteinander. Hierbei kann auch die Auseinandersetzung mit ungewohnten Ideen gefordert sein, genauso wie der kompetente Umgang mit konkurrierenden Informationen und abstrakten Interpretationskategorien sowie hohe Präzision im Umgang mit zum Teil unauffälligen Textdetails.
V 626–698 Punkte	Jugendliche auf dieser Stufe können sowohl mehrere tief eingebettete Informationen finden, ordnen und herausfinden, welche davon jeweils relevant sind, als auch ausgehend von Fachwissen eine kritische Beurteilung oder Hypothese anstellen. Die Aufgaben dieser Stufe setzen in der Regel ein volles und detailliertes Verständnis von Texten voraus, deren Inhalt oder Form ungewohnt ist. Zudem muss mit Konzepten umgegangen werden können, die im Gegensatz zum Erwarteten stehen.
IV 553–626 Punkte	Aufgaben dieser Kompetenzstufe erfordern vom Leser/von der Leserin, linguistischen oder thematischen Verknüpfungen in einem Text über mehrere Abschnitte zu folgen, oftmals ohne Verfügbarkeit eindeutiger Kennzeichen im Text, um eingebettete Informationen zu finden, zu interpretieren und zu bewerten oder um psychologische oder philosophische Bedeutungen zu erschließen. Insgesamt muss ein genaues Verständnis langer oder komplexer Texte, deren Inhalt oder Form ungewohnt sein kann, unter Beweis gestellt werden.
III 480–553 Punkte	Aufgaben dieser Kompetenzstufe erfordern vom Leser/von der Leserin, vorhandenes Wissen über die Organisation und den Aufbau von Texten zu nutzen, implizite oder explizite logische Relationen (z. B. Ursache-Wirkungs-Beziehungen) über mehrere Sätze oder Textabschnitte zu erkennen, mit dem Ziel, Informationen im Text zu lokalisieren, zu interpretieren und zu bewerten. Einige Aufgaben verlangen vom Leser/von der Leserin, einen Zusammenhang zu begreifen oder die Bedeutung eines Wortes oder Satzes zu analysieren. Häufig sind die benötigten Informationen dabei nicht leicht sichtbar oder Passagen des Textes laufen eigenen Erwartungen zuwider.
II 407–480 Punkte	Jugendliche auf dieser Stufe können innerhalb eines Textabschnitts logischen und linguistischen Verknüpfungen folgen, mit dem Ziel, Informationen im Text zu lokalisieren oder zu interpretieren, im Text oder über Textabschnitte verteilte Informationen aufeinander beziehen, um die Absicht des Autors zu erschließen. Bei Aufgaben dieser Stufe müssen unter Umständen auf der Grundlage eines einzigen Textbestandteils Vergleiche und Gegenüberstellungen vorgenommen werden oder es müssen, ausgehend von eigenen Erfahrungen oder Standpunkten, Vergleiche angestellt oder Zusammenhänge zwischen dem Text und nicht im Text enthaltenen Informationen erkannt werden.
Ia 335–407 Punkte	Aufgaben dieser Kompetenzstufe erfordern vom Leser/von der Leserin, in einem Text zu einem vertrauten Thema eine oder mehrere unabhängige, explizit ausgedrückte Informationen zu lokalisieren, das Hauptthema oder die Absicht des Autors zu erkennen oder einen einfachen Zusammenhang zwischen den im Text enthaltenen Informationen und allgemeinem Alltagswissen herzustellen. Die erforderlichen Informationen sind in der Regel leicht sichtbar, und es sind nur wenige beziehungsweise keine konkurrierenden Informationen vorhanden. Der Leser wird explizit auf die entscheidenden Elemente in der Aufgabe und im Text hingewiesen.
Ib 262–335 Punkte	Jugendliche auf dieser Stufe können in einem kurzen, syntaktisch einfachen Text aus einem gewohnten Kontext, dessen Form vertraut ist (z. B. in einer einfachen Liste oder Erzählung), eine einzige, explizit ausgedrückte Information lokalisieren, die leicht sichtbar ist. Der Text enthält in der Regel Hilfestellungen für den Leser, wie Wiederholungen, Bilder oder bekannte Symbole. Es gibt kaum konkurrierende Informationen. Bei anderen Aufgaben müssen einfache Zusammenhänge zwischen benachbarten Informationsteilen hergestellt werden.

Das Schauspiel sei das Werkzeug

Spielt in einem Schloss nahe an einem Strand in Italien.

ERSTER AKT

- 5 *Ein reich verziertes Gästezimmer in einem sehr schönen Schloss in Strandnähe. Türen rechts und links. Das Wohnzimmer befindet sich in der Mitte der Bühne: Sofa, Tisch, zwei Stühle. Im Hintergrund große Fenster. Sternklare Nacht. Die Bühne liegt im Dunkeln.*
- 10 *Als der Vorhang aufgeht, hört man Männer, die sich hinter der linken Tür lautstark unterhalten. Die Tür geht auf, und drei Herren im Smoking treten auf. Einer schaltet sofort das Licht ein. Sie gehen schweigend in die*
- 15 *Mitte und stehen um den Tisch herum. Sie setzen sich alle drei gleichzeitig, Gál auf den Stuhl links, Turai in den auf der rechten Seite, Ádám auf das Sofa in der Mitte. Sehr langes, fast unangenehmes Schweigen. Ein*
- 20 *gemächliches Sichstrecken. Stille. Dann:*

GÁL

Warum bist du so in Gedanken versunken?

TURAI

- 25 Ich denke darüber nach, wie schwierig es ist, ein Theaterstück zu beginnen. Die Hauptfiguren am Anfang vorzustellen, wenn das alles beginnt.

ÁDÁM

Ich nehme an, dass das schwierig ist.

- 30 TURAI

- Es ist – teuflisch schwer. Das Theaterstück beginnt. Das Publikum wird ruhig. Die Schauspielerinnen und Schauspieler betreten die Bühne und die Qual beginnt. Es vergeht
- 35 eine Ewigkeit, manchmal eine ganze Viertelstunde, bis die Zuschauerinnen und Zuschauer herausfinden, wer wer ist und wer was im Schilde führt.

GÁL

- 40 Ein recht eigenartiges Gehirn hast du schon. Kannst du deinen Beruf nicht für eine einzige Minute vergessen?

TURAI

Das ist unmöglich.

- 45 GÁL

Es vergeht keine halbe Stunde, ohne dass du über das Theater, die Schauspielerinnen und Schauspieler oder die Theaterstücke diskutieren möchtest. Es gibt andere Dinge auf dieser Welt.

TURAI

Die gibt es nicht. Ich bin ein Dramatiker. Das ist mein Fluch.

GÁL

- 55 Du solltest nicht Sklave deiner Arbeit werden.

TURAI

- Wenn du sie nicht beherrscht, bist du ihr Sklave. Es gibt nichts dazwischen. Glaub mir, es ist kein Spaß, einen guten Anfang für ein Theaterstück zu finden. Es ist eines der schwierigsten Probleme des Theaterhandwerks. Alle Figuren umgehend vorzustellen. Schauen wir uns diese Szene
- 60 an – uns drei. Drei Herren im Smoking. Nehmen wir an, sie betreten nicht einen Raum in diesem edlen Schloss, sondern eine Bühne, genau am Anfang des Theaterstückes. Sie müssten über viele uninteressante Dinge sprechen, bis es herauskommen würde, wer wir sind. Wäre es nicht viel einfacher, damit zu beginnen, dass alle aufstehen und sich selber vorstellen? *Er steht auf.* Guten Abend.
- 70 Wir drei sind Gäste auf diesem Schloss. Wir kommen eben aus dem Speisezimmer, wo wir ein ausgezeichnetes Essen zu uns genommen und zwei Flaschen Champagner getrunken haben. Mein
- 80 Name ist Sandor Turai, ich habe während der letzten dreißig Jahre Theaterstücke geschrieben, das ist mein Beruf. Punkt. Du bist an der Reihe.

- 85 GÁL

Steht auf. Mein Name ist Gál, ich bin auch ein Stückeschreiber. Ich schreibe auch Theaterstücke, alle in Gesellschaft mit diesem Herren hier. Wir sind ein

90 berühmtes Dramatiker-Duo. Auf allen

Programmheften guter Komödien oder Operetten steht „geschrieben von Gál und Turai“. Natürlich ist das auch mein Beruf.

GÁL und TURAI

95 *Zusammen.* Und dieser junge Mann ...

ÁDÁM

- 100 *Steht auf.* Dieser junge Mann ist, wenn Sie mir erlauben, Albert Ádám, fünfundzwanzig Jahre alt, Komponist. Ich habe die Musik für die letzte Operette dieser zwei lebenswürdigen Herren geschrieben. Das ist mein erstes Werk für die Bühne. Diese zwei älteren Engel haben mich entdeckt und jetzt würde ich gerne mit ihrer Hilfe berühmt werden. Sie haben mich in dieses Schloss eingeladen. Sie haben mir einen Frack und einen Smoking schneidern lassen.
- 105 Mit anderen Worten, ich bin noch arm und unbekannt. Darüber hinaus bin ich Waise und wurde von meiner Großmutter großgezogen.
- 110 Meine Großmutter ist gestorben. Ich bin ganz allein auf dieser Welt. Ich habe keinen Namen, ich habe kein Geld.

TURAI

Aber du bist jung.

115 GÁL

Und talentiert.

ÁDÁM

Und ich bin in die Solistin verliebt.

TURAI

120 Das hättest du nicht hinzufügen sollen. Alle im Publikum würden das sowieso merken.

Sie setzen sich alle.

TURAI

125 Also, wäre das nicht die einfachste Art, ein Theaterstück zu beginnen?

GÁL

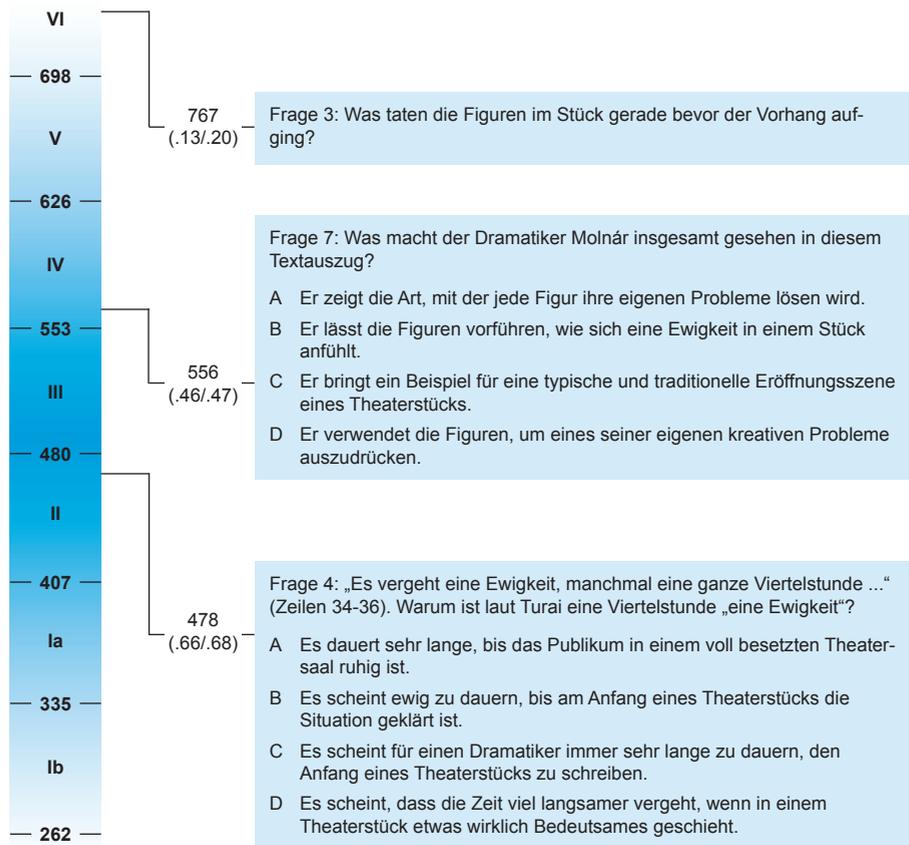
Wenn wir das dürften, wäre es einfach, Theaterstücke zu schreiben.

130 TURAI

Glaub mir, so schwer ist es nicht. Sieh das Ganze einfach einmal an als ...

GÁL

135 Schon gut, schon gut, schon gut, fang nicht schon wieder damit an, übers Theater zu sprechen. Ich habe genug davon. Morgen sprechen wir weiter, wenn du möchtest.



Anmerkung. Die Werte an den Verbindungslinien zwischen der Kompetenzsäule und den Beispielitems geben das für eine 62 %-Lösungswahrscheinlichkeit erforderliche Kompetenzniveau und die Werte in Klammern die beobachteten relativen internationalen (OECD-Durchschnitt) und deutschen Lösungshäufigkeiten bei PISA 2009 an.

Abbildung 7.1: Aufgabenbeispiel „Das Schauspiel sei das Werkzeug“

7.2 Lesekompetenz im internationalen Vergleich

Im Mittelpunkt der folgenden Ausführungen steht die Frage, wie die Schülerinnen und Schüler in Deutschland im Vergleich zu anderen Ländern – und vorrangig der Bezugsgruppe der OECD-Staaten – im Lesekompetenztest abschnitten. In diesem Zusammenhang werden zum Beispiel auch Unterschiede zwischen Mädchen und Jungen im internationalen Vergleich betrachtet.

Bei PISA 2012 liegt die durchschnittliche Lesekompetenz über alle OECD-Staaten bei 496 Punkten, die Standardabweichung bei 94 Punkten. Diese leichte Abnahme im Mittelwert der OECD-Staaten über die Zeit kann nicht unbedingt als Verschlechterung der Leseleistung in der OECD interpretiert werden, denn seit der ersten Erhebung im Jahr 2000 wurde die Gruppe der OECD-Staaten um einige Staaten erweitert. Die Veränderung im Mittelwert über die Zeit kann also auch der veränderten Zusammensetzung der OECD-Staaten geschuldet sein.

Lesekompetenz: Mittelwerte im Vergleich

Abbildung 7.2 zeigt unter anderem die Mittelwerte und Standardabweichungen der Lesekompetenz in den einzelnen OECD-Teilnehmerstaaten und differenziert drei Gruppen in Relation zum Durchschnittswert der OECD-Staaten (statistisch signifikant *über* oder *unter* dem OECD-Durchschnitt sowie statistisch nicht von diesem zu unterscheiden). Wie der Abbildung entnommen werden kann, liegen 15 Staaten signifikant über dem OECD-Durchschnitt (z. B. Frankreich, Deutschland, Schweiz, Belgien, Niederlande, Polen, Finnland, Korea, Japan). Vier Staaten weichen in den Leseleistungen nicht signifikant vom OECD-Durchschnitt ab (Vereinigtes Königreich, Vereinigte Staaten, Dänemark, Tschechische Republik) und 15 Staaten liegen signifikant unter dem OECD-Durchschnitt (z. B. Italien, Österreich, Spanien, Luxemburg, Schweden). Bei PISA 2012 gehört Deutschland mit einem Mittelwert von 508 Punkten erstmals zu der Gruppe der OECD-Staaten, in der die Lesekompetenz der Fünfzehnjährigen signifikant über dem OECD-Durchschnitt liegt. In den Erhebungsrunden von 2003, 2006 und 2009 lagen die Schülerinnen und Schüler in Deutschland lediglich im mittleren OECD-Leistungsbereich, bei PISA 2000 sogar signifikant unter dem OECD-Durchschnitt. Im internationalen Vergleich ist in Deutschland damit eine positive Entwicklung der Lesekompetenz festzustellen; in Kapitel 7.3 werden die Ergebnisse für Deutschland differenzierter betrachtet.

Die besten Leseleistungen in der Gruppe der OECD-Staaten wurden 2012 in Japan mit durchschnittlich 538 Punkten und Korea mit 536 Punkten erzielt. Finnland erreicht 524 Punkte. Die Unterschiede innerhalb der Gruppe der Staaten über dem OECD-Durchschnitt sind zum Teil substantiell. Der Abstand von circa 30 Punkten etwa zwischen Japan oder auch Korea einerseits und Deutschland andererseits steht für deutliche Unterschiede in der Ausprägung der Lesekompetenz, die fast einer halben Kompetenzstufe entsprechen. Bei den leistungsschwächsten OECD-Staaten handelt es sich bei

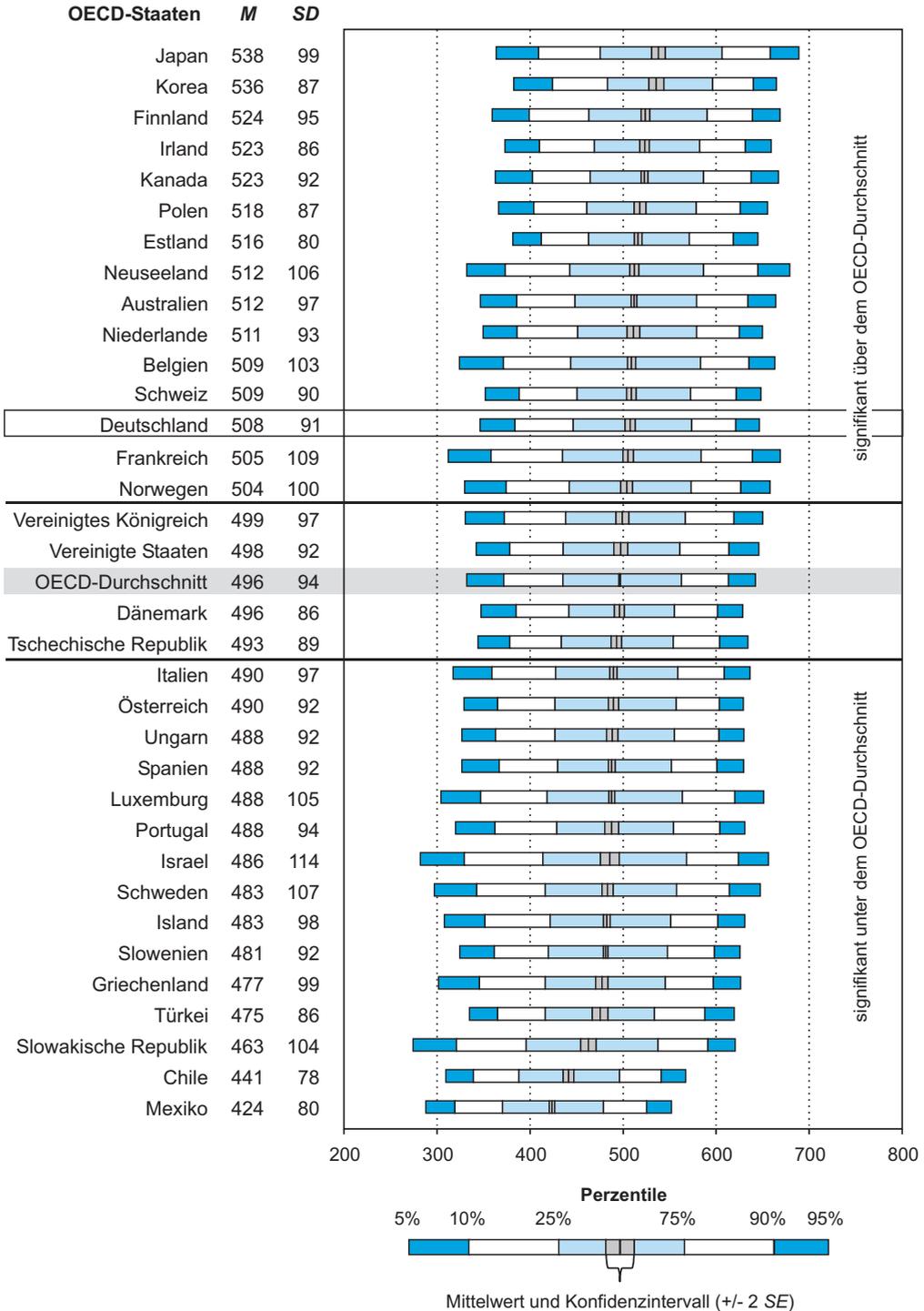


Abbildung 7.2: Perzentilbänder der Lesekompetenz in den OECD-Staaten

PISA 2012, wie auch schon bei PISA 2009, um Mexiko ($M = 424$) und Chile ($M = 441$). Dabei veränderten sich die Lesekompetenzen der Schülerinnen und Schüler in diesen beiden Staaten auch im absoluten Sinne im Verlauf der letzten drei Jahre kaum.

Ausgezeichnete und bemerkenswerte Leistungen im Lesen können in einigen OECD-Partnerstaaten festgestellt werden, in denen die Leseleistung über dem OECD-Durchschnitt liegt. Besonders hervorzuheben ist hier Shanghai, wo die Werte der Lesekompetenz der Schülerinnen und Schüler mit 570 Punkten deutlich (statistisch signifikant) besser ausfallen als bei den OECD-Spitzenreitern Japan und Korea. Bemerkenswert sind ebenfalls die Leseleistungen in Hongkong und Singapur mit 545 und 542 Punkten. Die Leistung der Schülerinnen und Schüler in Chinesisch Taipeh liegt mit 523 Punkten ebenfalls über dem OECD-Durchschnitt. Eine zusammenfassende Übersicht über die Lesekompetenzen in den teilnehmenden OECD-Staaten sowie in den OECD-Partnerstaaten findet sich im Anhang (Tabelle 7).

Die Streuung der Lesekompetenz

Nicht nur das mittlere Abschneiden beim Lesekompetenztest ist von Interesse, um Aussagen über die Lesekompetenz in einzelnen Staaten zu treffen. Die Standardabweichung als Maß der Streuung liefert zusätzliche Informationen darüber, wie sehr die Unterschiede in der Lesekompetenz innerhalb der einzelnen Teilnehmerstaaten ausgeprägt sind. Große Streuungen bedeuten in diesem Zusammenhang, dass sich die leistungsstärksten und leistungsschwächsten Leserinnen und Leser innerhalb eines Landes hinsichtlich ihrer Lesefertigkeiten besonders unterscheiden. Je kleiner die Streuung ausfällt, umso geringer sind die Unterschiede zwischen den Leistungsstärksten und Leistungsschwächsten. Die durchschnittliche Streuung in den teilnehmenden OECD-Staaten beträgt bei PISA 2012 94 Punkte. Besonders hohe Streuungen innerhalb der OECD-Staaten finden sich zum Beispiel in Frankreich ($SD = 109$), Schweden ($SD = 107$) oder auch in Luxemburg ($SD = 105$). Schon bei PISA 2009 gab es in Frankreich und Schweden große Leistungsstreuungen. In Schweden jedoch fiel sie noch vor drei Jahren signifikant kleiner aus ($SD = 99$). Die geringsten Unterschiede in der Streuung der Lesekompetenz lassen sich in Chile, Mexiko und Estland finden. Dort beträgt die Standardabweichung lediglich 78 (Chile) beziehungsweise 80 Punkte (Mexiko, Estland). Hervorzuheben ist, dass in OECD-Staaten mit Spitzenleistungen wie Japan oder Korea die Leistungsstreuungen durchaus recht unterschiedlich ausfallen. So beträgt bei PISA 2012 die Streuung in Japan 99 Punkte und liegt damit signifikant über dem OECD-Durchschnitt. In Korea hingegen fällt die Streuung deutlich geringer aus als in Japan (statistisch signifikant) und liegt mit 87 Punkten auch signifikant unter dem OECD-Durchschnitt.

Die Spitzenreiter der OECD-Partnerstaaten (Shanghai, Hongkong, Singapur, Chinesisch Taipeh) zeigen sich hinsichtlich ihrer Leistungsstreuungen recht unterschiedlich. Diese schwanken zwischen 80 und 101 Punkten. Shanghai ($SD = 80$), Hongkong ($SD = 85$) und Chinesisch Taipeh ($SD = 91$) weisen Leistungsstreuungen auf, die signifikant unter dem OECD-Durchschnitt liegen. In Singapur fällt die Streuung hinge-

gen überdurchschnittlich groß aus ($SD = 101$). Eine Übersicht über die Streuungen der Lesekompetenzen in den teilnehmenden OECD-Partnerstaaten findet sich im Anhang (Tabelle 7).

Für Deutschland zeigt sich, dass die Streuung der Lesekompetenz mit rund 91 Punkten signifikant von der durchschnittlichen OECD-Streuung abweicht. Deutschland setzt damit seinen (erfreulichen) Trend der Verringerung der Leistungsstreuung weiter fort. Im Vergleich zu PISA 2000, wo es in Deutschland eine im internationalen Vergleich hohe Leistungsstreuung gab, zeichnet sich damit in Deutschland ein Trend in Richtung homogenerer Leistungen ab.

Das oben beschriebene Abschneiden der OECD-Teilnehmerstaaten wird in Abbildung 7.2 durch die graphische Darstellung der Perzentilbänder pro Staat verdeutlicht. Die Breite der Perzentilbänder spiegelt dabei das Ausmaß an Kompetenzunterschieden zwischen den 5 Prozent der schwächsten und den 5 Prozent der stärksten Schülerinnen und Schülern eines Landes im Bereich der Lesekompetenz wider.

Verteilung auf die Stufen der Lesekompetenz

Das PISA-Design erlaubt es, die im Test erzielten Leseleistungen der Schülerinnen und Schüler mit den korrespondierenden Aufgabenanforderungen in Beziehung zu setzen. Dies erfolgt anhand der Kompetenzstufen. Hier sind jene Schülerinnen und Schüler von besonderem Interesse, die sich den obersten beziehungsweise den untersten Kompetenzstufen zuordnen lassen. Abbildung 7.3 gibt einen Überblick über den Anteil Jugendlicher in den OECD-Teilnehmerstaaten, deren Lesekompetenz sich auf den Kompetenzstufen V oder VI befindet beziehungsweise auf den Kompetenzstufen Ia, Ib oder darunter.

Im OECD-Durchschnitt liegen 18.0 Prozent der Schülerinnen und Schüler auf den untersten Kompetenzstufen (Ia, Ib und unter Ib). In Deutschland können, ähnlich wie bei PISA 2009, 14.5 Prozent der Schülerinnen und Schüler auf den untersten Kompetenzstufen verortet werden. Weniger erfreulich scheint die Lage zum Beispiel in Schweden, Österreich oder auch in Italien zu sein. Hier liegen die Lesekompetenzen von 22.7 Prozent (Schweden) beziehungsweise 19.5 Prozent (Österreich und Italien) der Schülerinnen und Schüler maximal auf Kompetenzstufe Ia. Bemerkenswert sind hingegen die Anteile von Schülerinnen und Schülern in den Staaten der OECD-Spitzengruppe. Hier können in Japan 9.8 Prozent und in Korea 7.6 Prozent der Fünfzehnjährigen auf den untersten Kompetenzstufen eingeordnet werden. Diese Staaten zeigen, dass es im Prinzip möglich ist, den Anteil der schwachen Leserinnen und Leser sehr gering zu halten.

Auf den obersten Kompetenzstufen (V und VI) liegen im OECD-Durchschnitt 8.4 Prozent der Schülerinnen und Schüler. Deutlich geringer ist der Anteil von Jugendlichen zum Beispiel in Dänemark (5.4 Prozent), Österreich (5.5 Prozent) oder auch Spanien (5.5 Prozent). Die Schülerinnen und Schüler in Deutschland verteilen sich auch hier ähnlich wie bei PISA 2009 und liegen mit 8.9 Prozent im internationalen Durchschnitt. Dieser Anteil in Deutschland deutet an, dass es in den letzten Jahren zwar

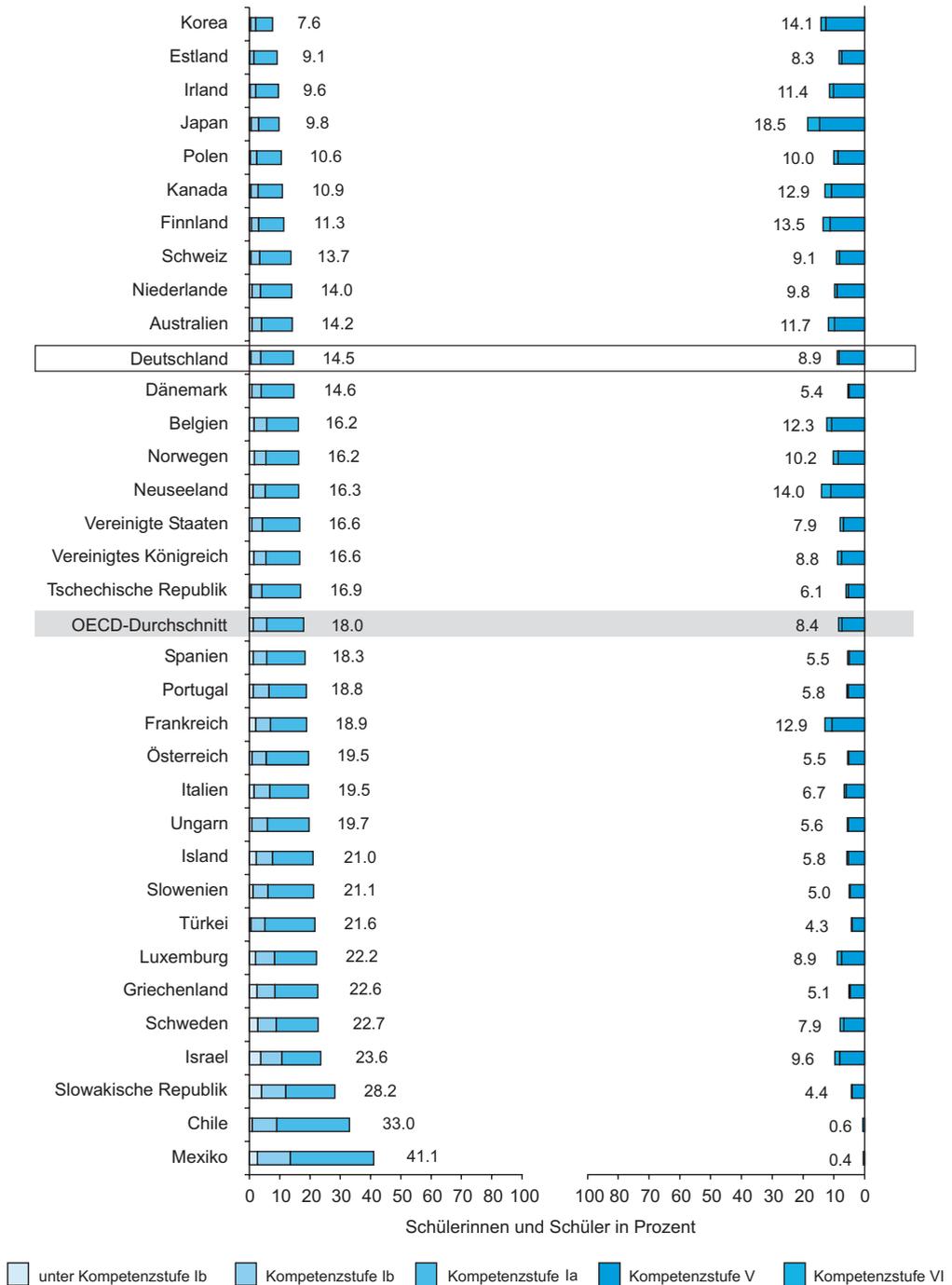


Abbildung 7.3: Prozentuale Anteile von Schülerinnen und Schülern auf Kompetenzstufe Ia, Kompetenzstufe Ib oder darunter sowie auf Kompetenzstufe V und Kompetenzstufe VI

gelingen ist, die Gruppe sehr guter Leserinnen und Leser etwas zu vergrößern, doch fehlt nach wie vor eine Spitzengruppe, die vergleichbar ist mit der in Staaten wie Korea (14.1 Prozent), Japan (18.5 Prozent) oder auch Neuseeland (14.0 Prozent). Bei den OECD-Partnerstaaten sind vor allem Shanghai (25.1 Prozent), Singapur (21.2 Prozent) oder Hongkong (16.8 Prozent) mit einem besonders hohen Anteil von Schülerinnen und Schülern auf den obersten Kompetenzstufen zu nennen.

Zu den OECD-Staaten, in denen sowohl ein großer Anteil an Schülerinnen und Schülern auf den untersten als auch ein geringer Anteil auf den obersten Kompetenzstufen zu verorten ist, gehören unter anderem Griechenland, die Türkei oder auch Island. In Griechenland können 22.6 Prozent der Fünfzehnjährigen auf den untersten und lediglich 5.1 Prozent auf den obersten Kompetenzstufen verortet werden. In der Türkei sind 21.6 Prozent der Jugendlichen auf den untersten und 4.3 Prozent auf den obersten Kompetenzstufen. In Island ist die Situation sehr ähnlich, hier befinden sich 21.0 Prozent der Fünfzehnjährigen auf den untersten und lediglich 5.8 Prozent auf den obersten Kompetenzstufen.

Unterschiede zwischen Mädchen und Jungen in der Lesekompetenz

Seit der ersten Erhebungsrunde von PISA (2000) wird immer wieder deutlich, dass sich die Leistungen von Mädchen und Jungen im Leseverständnis in den OECD-Staaten stark unterscheiden. Dabei schneiden die Jungen erheblich schlechter ab als die Mädchen. Auch bei PISA 2012 hat sich dieses Bild nicht geändert. In allen OECD-Staaten liegen die Leistungen der Mädchen signifikant über denen der Jungen (vgl. Abbildung 7.4). Die Differenz zwischen der Lesekompetenz von Mädchen und Jungen beträgt in Deutschland 44 Punkte – das entspricht einem durchschnittlichen Leistungsabstand von mehr als einer halben Kompetenzstufe. Diese Differenz hat sich seit den letzten PISA-Erhebungsrunden kaum geändert. Deutschland gehört damit zu den OECD-Staaten mit den größten Leistungsunterschieden zwischen Mädchen und Jungen im Lesen. Ähnlich große Unterschiede wie in Deutschland gibt es zum Beispiel in Norwegen (46 Punkte), Frankreich (44 Punkte) oder Polen (42 Punkte). Besonders gravierend fällt der Unterschied, wie auch schon bei PISA 2009, zwischen Mädchen und Jungen in Slowenien und in Finnland aus. Hier erreichen die Mädchen im Lesekompetenztest im Durchschnitt fast 56 Punkte (Slowenien) beziehungsweise 62 Punkte (Finnland) mehr als die Jungen. Im Mittel liegen die Leistungen der Mädchen damit fast eine ganze Kompetenzstufe über denen der Jungen. Im Vergleich zu vor drei Jahren hat sich die Leistungsdiskrepanz zwischen den Geschlechtern in Slowenien kaum geändert, in Finnland hingegen ist sie um weitere 7 Punkte angestiegen. Interessanterweise fällt die Geschlechterdifferenz sowohl in den leistungsstärksten OECD-Staaten Japan (24 Punkte) und Korea (23 Punkte) als auch in den leistungsschwächsten Ländern Mexiko (24 Punkte) und Chile (23 Punkte) am geringsten aus.

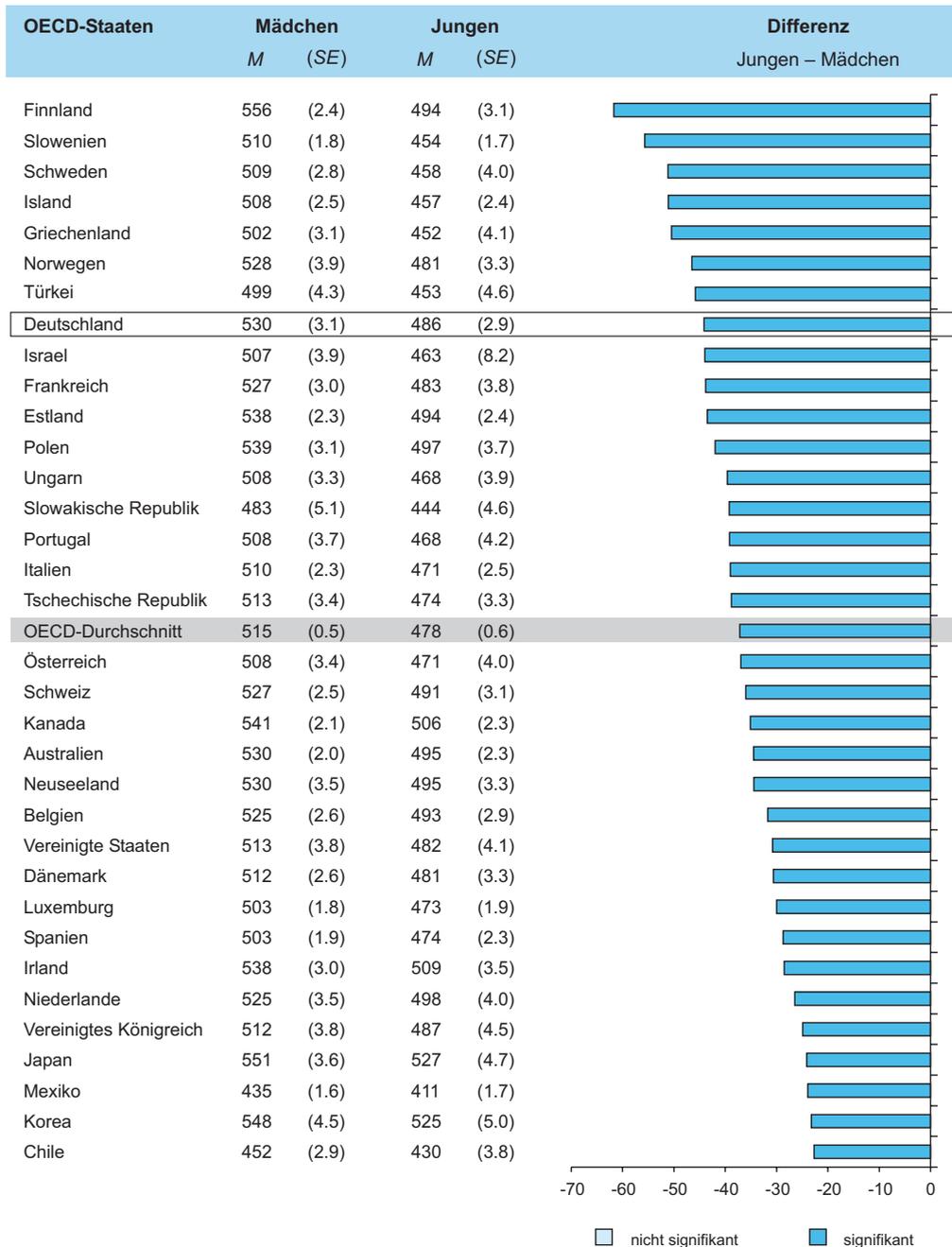


Abbildung 7.4: Mittelwerte der Lesekompetenz von Mädchen und Jungen in den OECD-Staaten

7.3 Vertiefende Analysen der Lesekompetenz in Deutschland

Im Folgenden werden die Ergebnisse der Lesekompetenz in Deutschland in Abhängigkeit von der Schulart betrachtet und es wird der Frage nachgegangen, wie sich die Lesekompetenz seit PISA 2000 in Deutschland entwickelt hat.

7.3.1 Unterschiede zwischen Schularten

Charakteristisch für das Schulsystem in Deutschland ist die Differenzierung in verschiedene Schularten. Wie im Einleitungskapitel dargelegt, kann in diesem nationalen Bericht zu PISA 2012 zwischen sieben Schularten unterschieden werden: Hauptschulen, Schulen mit mehreren Bildungsgängen, Integrierte Gesamtschulen, Realschulen, Gymnasien sowie berufliche Schulen und Sonder- beziehungsweise Förderschulen. Im Nachfolgenden wird die Lesekompetenz der Schülerinnen und Schüler in Deutschland in Abhängigkeit von den besuchten Schularten berichtet. Da die Stichproben der Jugendlichen aus beruflichen Schulen und aus Sonder- beziehungsweise Förderschulen für diese Schularten nicht repräsentativ sind und die Ergebnisse damit nicht belastbar wären, wird die Darstellung auf die verbleibenden fünf Schularten beschränkt.

Lesekompetenz: Mittelwerte und Streuungen im Vergleich

Es ist zu erwarten, dass sich die Lesekompetenz von Schülerinnen und Schülern innerhalb eines gegliederten Schulsystems, das nach Leistung differenziert, zwischen den Schularten unterscheidet. Die Ergebnisse von PISA 2012 bestätigen dies (vgl. Tabelle 7.2). Das mittlere Kompetenzniveau in den Hauptschulen liegt bei 417 Punkten und damit über eine Kompetenzstufe unter dem Mittelwert der Gesamtstichprobe in Deutschland von 508 Punkten. In Schulen mit mehreren Bildungsgängen erreichen die Schülerinnen und Schüler im Mittel 458 Punkte, was knapp eine Kompetenzstufe unter dem bundesweiten Durchschnitt liegt. Und auch in den Integrierten Gesamtschulen liegt die mittlere Leistung der Schülerinnen und Schüler mit 477 Punkten deutlich unter dem deutschen Durchschnitt. Die durchschnittliche Leseleistung der Schülerinnen und Schüler an den Realschulen liegt mit 509 Punkten nahe am Mittelwert für Deutschland. Die bei PISA 2012 gemessene Lesekompetenz fällt erwartungsgemäß an den Gymnasien am größten aus. Der Mittelwert für Gymnasien beträgt 579 Punkte.

Tabelle 7.2: Mittelwerte und Standardabweichungen der Lesekompetenz in Deutschland nach Schulart und für die Gesamtstichprobe

Schulart	<i>M</i>	(<i>SE</i>)	<i>SD</i>	(<i>SE</i>)
Hauptschule	417	(8.3)	69	(3.0)
Schule mit mehreren Bildungsgängen	458	(7.6)	77	(2.8)
Integrierte Gesamtschule	477	(4.4)	67	(2.0)
Realschule	509	(3.9)	65	(1.6)
Gymnasium	579	(2.1)	59	(1.2)
Gesamtstichprobe	508	(2.8)	91	(1.7)

Bei der Interpretation dieser Kennwerte für die einzelnen Schularten ist es notwendig, die jeweilige Bildungsbeteiligung zu berücksichtigen. Aus diesem Grund ist in Abbildung 7.5 mit angegeben, wie groß der prozentuale Anteil der Jugendlichen der untersuchten Altersgruppe ist, der die jeweilige Schulart besucht. Will man zum Beispiel den Mittelwert der Gymnasien international einordnen, dann ist zu berücksichtigen, dass hier die Leistungen einer besonders leistungsstarken Teilgruppe von 36 Prozent der Fünfzehnjährigen in Deutschland mit denen aller Schülerinnen und Schüler in den jeweils anderen Staaten verglichen werden.

Die Streuung der Leseleistungen in den einzelnen Schularten fällt in den Schulen mit mehreren Bildungsgängen am größten aus ($SD = 77$) und in den Gymnasien am geringsten ($SD = 59$). Auch dieser Befund überrascht nicht, denn das Etikett „Schule mit mehreren Bildungsgängen“ steht für eine gewisse Vielfalt von Schulmodellen in den Ländern, wohingegen das Gymnasium ein über die Länder weitgehend übereinstimmendes Schulkonzept repräsentiert – wenn auch die Bildungsbeteiligung am Gymnasium zwischen den Ländern beträchtlich variiert (vgl. Pant, Stanat, Schroeders, Roppelt, Siegle & Pöhlmann, 2013). In den verbleibenden Schularten liegen die Leistungsstreuungen zwischen circa 65 und knapp 70 Punkten, was nahezu einer Kompetenzstufe entspricht. Somit unterscheiden sich die leistungsstärksten und leistungsschwächsten Schülerinnen und Schüler innerhalb einer Schulart um fast eine Kompetenzstufe hinsichtlich ihrer Lesekompetenzen.

Die Werte der Lesekompetenz und der Leistungsstreuung für die einzelnen Schularten sind in Abbildung 7.5 anhand von Perzentilbändern noch einmal veranschaulicht und in Relation zu denen der Gesamtstichprobe gesetzt. Die Gesamtstichprobe umfasst dabei alle in Deutschland getesteten Schülerinnen und Schüler. Die Perzentilbänder können zum Beispiel genutzt werden, um in etwa den Anteil der Schülerinnen und Schüler in Vergleichsstaaten abzuschätzen, der über dem Mittelwert der Gymnasien in Deutschland liegt. Einfach ist der Vergleich mit Shanghai: Dort liest ein Großteil der Schülerinnen und Schüler auf dem Niveau der Gymnasien in Deutschland. Darüber hinaus machen die Perzentilbänder deutlich, dass sich die Verteilungen der

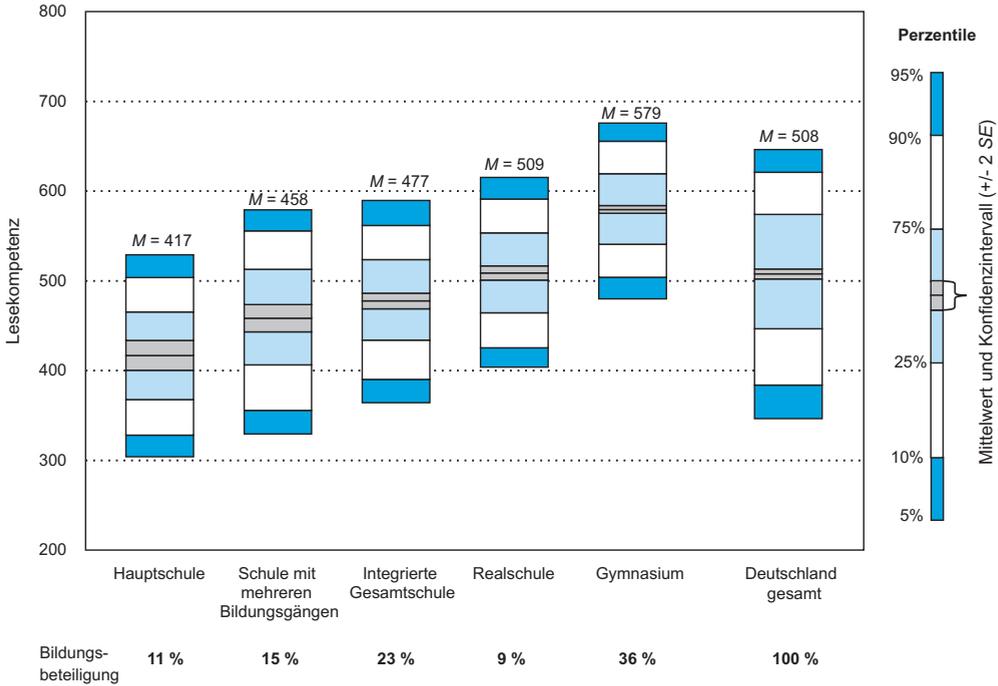


Abbildung 7.5: Perzentilbänder für die Lesekompetenz in Deutschland nach Schulart und für die Gesamtstichprobe

Leistungen der Schülerinnen und Schüler in den verschiedenen Schularten in Deutschland deutlich überlappen. Dies gilt sogar für die leistungsschwächste Schulart Hauptschule und die leistungstärkste Schulart Gymnasium. So liegt das 90. Perzentil in der Hauptschule mit 504 Punkten in der Größenordnung des 10. Perzentils im Gymnasium (ebenfalls 504 Punkte): Die besten 10 Prozent der Hauptschülerinnen und Hauptschüler verfügen also über genauso stark ausgeprägte Lesekompetenzen wie die schlechtesten 10 Prozent der Gymnasiastinnen und Gymnasiasten. Deutlich größere Überschneidungen zeigen sich in Schulen mit mehreren Bildungsgängen und Integrierten Gesamtschulen. Diese sind sich besonders in den oberen Perzentilen der Verteilung ähnlich. Das 90. Perzentil in Schulen mit mehreren Bildungsgängen (555 Punkte) und das 90. Perzentil in Integrierten Gesamtschulen (562 Punkte) sind nahezu identisch. Größere Unterschiede lassen sich beim Vergleich dieser beiden Schularten eher in den unteren Leistungsbereichen ausfindig machen. In Schulen mit mehreren Bildungsgängen liegen die schwächsten 10 Prozent der Schülerinnen und Schüler (356 Punkte) auf einem sehr ähnlichen Niveau wie die schwächsten 5 Prozent der Integrierten Gesamtschülerinnen und -schüler (364 Punkte).

Verteilung auf die Stufen der Lesekompetenz

Mit Blick auf die schulartspezifischen Anteile von Jugendlichen auf den jeweiligen Kompetenzstufen (vgl. Abbildung 7.6) ist vor allem der Anteil von leseschwachen Fünfzehnjährigen (Kompetenzstufe Ia oder darunter) von Interesse. Dieser Anteil von Schülerinnen und Schülern beträgt an Hauptschulen 43,8 Prozent, an den Schulen mit mehreren Bildungsgängen 25,6 Prozent und an den Integrierten Gesamtschulen 14,4 Prozent. Selbst an den Realschulen lesen noch knapp 6 Prozent der Schülerinnen und Schüler auf einem Niveau von Kompetenzstufe Ia oder darunter. Für jene Schülerinnen und Schüler kann ein beträchtliches Risiko für die weitere Schul- und Berufsbiographie wie auch für die gesellschaftliche Teilhabe angenommen werden. Auch wenn im Jahr 2012 ein erfreulicher Rückgang des Anteils schwacher Leserinnen und Leser seit der ersten Erhebungsrunde berichtet werden kann, zeigt die Aufschlüsselung nach Schularten nach wie vor Handlungsbedarf zur Stärkung der grundlegenden Lesefähigkeit und zur frühen Identifikation und Förderung von Schülerinnen und Schülern mit sich abzeichnenden Leseproblemen (vgl. Schneider et al., 2013).

Betrachtet man die andere Seite der Kompetenzstufenverteilung und damit die Anteile der Schülerinnen und Schüler in der Spitzengruppe (Kompetenzstufen V und VI), dann findet man außerhalb der Gymnasien nur wenig Schülerinnen und Schüler auf diesem Niveau (z. B. Realschule 3,3 Prozent). Der Anteil der entsprechenden Spit-

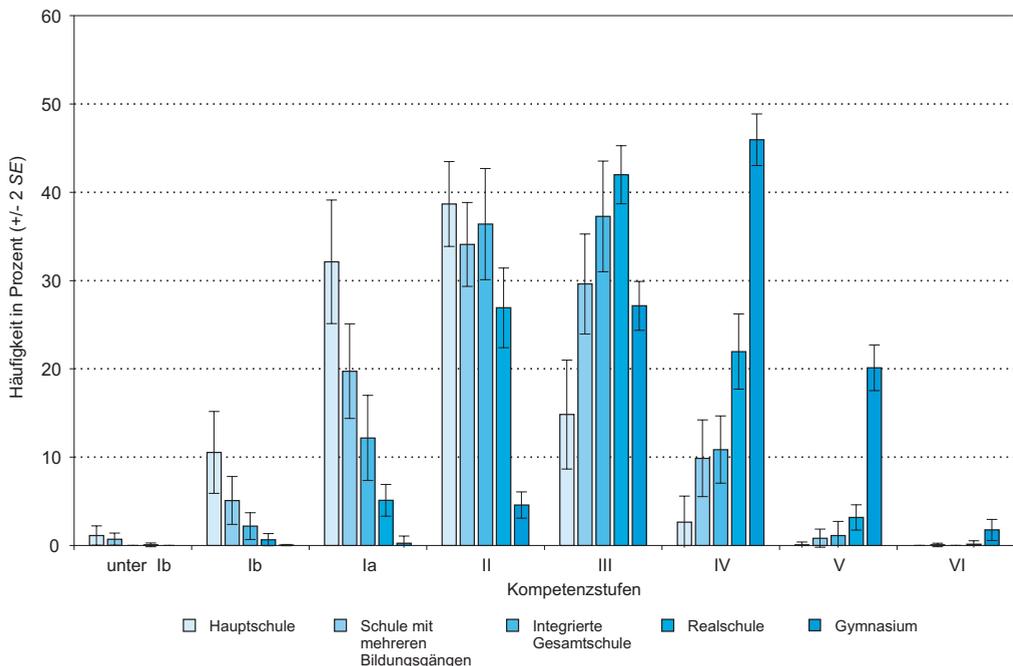


Abbildung 7.6: Prozentuale Anteile der Schülerinnen und Schüler in Deutschland auf den Stufen der Lesekompetenz nach Schulart

zengruppe im Gymnasium beträgt 21.9 Prozent. Dieser Wert erscheint auf den ersten Blick erfreulich hoch zu sein. Im internationalen Vergleich zeichnet sich allerdings ab, dass in anderen Staaten – und auch in solchen mit einem nicht differenzierten Schulsystem – deutlich mehr Schülerinnen und Schüler den obersten Kompetenzstufen (V und VI) zugeordnet werden können. Die bessere Ausgangslage der Gymnasien im gegliederten System Deutschlands, die durch die Auswahl besonders leistungsstarker Schülerinnen und Schüler am Ende der Grundschulzeit gekennzeichnet ist, scheint nicht ausreichend für die Entwicklung eines größeren Anteils von ausgezeichneten Leserinnen und Lesern genutzt zu werden.

7.3.2 Unterschiede zwischen Mädchen und Jungen in der Lesekompetenz

Im Abschnitt zum internationalen Vergleich wurde bereits auf die beträchtlichen Unterschiede zwischen Mädchen und Jungen in der Lesekompetenz hingewiesen. Um dieses Ergebnis für Deutschland differenzierter betrachten zu können, ist in Abbildung 7.7 die Verteilung von Mädchen und Jungen auf die sieben Kompetenzstufen dargestellt. Wie auch schon bei PISA 2009 liegen deutlich mehr Jungen als Mädchen auf den unteren Kompetenzstufen. 20.1 Prozent der Jungen liegen auf der Kompetenzstufe Ia oder darunter, wohingegen lediglich 5.2 Prozent der Jungen auf den beiden obersten Kompetenz-

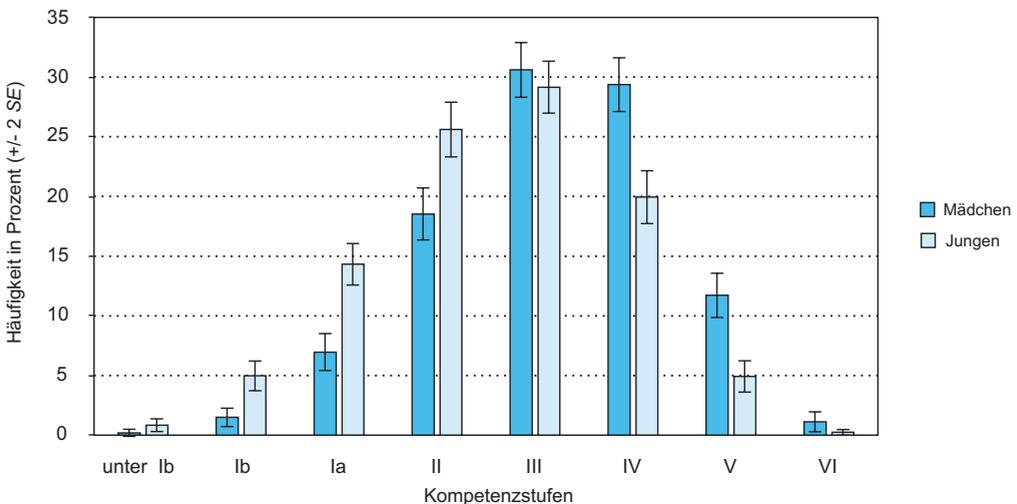


Abbildung 7.7: Prozentuale Anteile der Mädchen und Jungen in Deutschland auf den Stufen der Lesekompetenz

stufen (V und VI) zu verorten sind. Bei den Mädchen jedoch liegen knapp 13 Prozent auf den beiden obersten Kompetenzstufen, und nur 8.7 Prozent befinden sich auf den untersten Kompetenzstufen.

Beim Vergleich der Lesekompetenz von Jungen und Mädchen in den Jahren 2009 und 2012 zeichnet sich für beide Geschlechter eine Tendenz zur Verringerung des Anteils von Schülerinnen und Schülern auf den untersten Kompetenzstufen ab, bei den Jungen und den Mädchen jeweils um fast 4 Prozent. Zugleich ist der Anteil auf den obersten Kompetenzstufen leicht gestiegen, bei den Jungen um 0.8 Prozent und bei den Mädchen um 1.8 Prozent. Die Veränderung in den Verteilungen auf den Kompetenzstufen von Jungen und Mädchen zwischen PISA 2009 und 2012 ist in Abbildung 7.8 veranschaulicht.

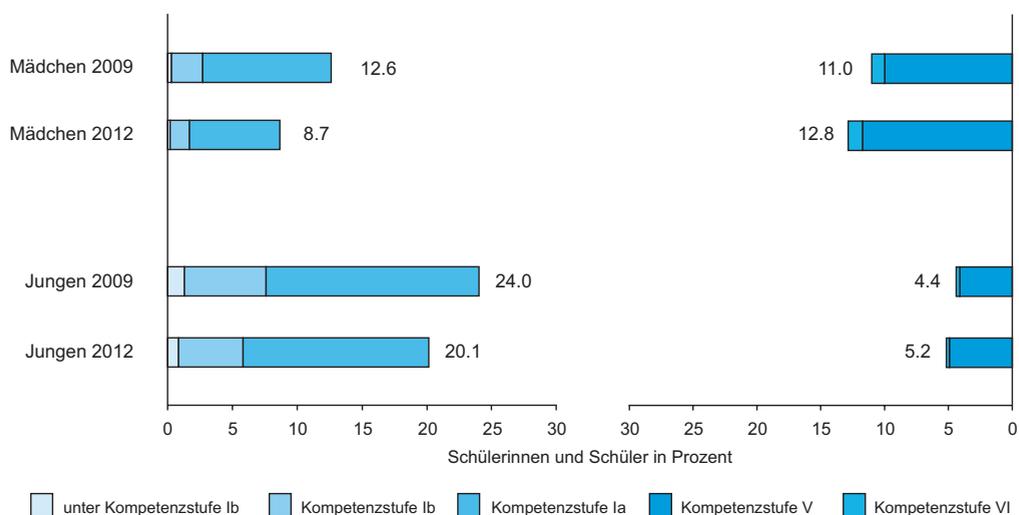


Abbildung 7.8: Prozentuale Anteile von Mädchen und Jungen in Deutschland bei PISA 2009 und 2012 auf Kompetenzstufe Ia, Ib oder darunter sowie auf Kompetenzstufe V und VI der Gesamtskala Lesen

7.3.3 Veränderung der Lesekompetenz seit PISA 2000

Wie bereits berichtet, verbesserte sich das Leistungsniveau der Schülerinnen und Schüler in Deutschland hinsichtlich der Lesekompetenz über die einzelnen PISA-Erhebungsrunden Stück für Stück. Dieser Trend setzte sich auch bei PISA 2012 fort, wobei die Fünfzehnjährigen in Deutschland 2012 erstmals signifikant oberhalb des OECD-Durchschnitts lagen. Auch der Leistungszuwachs von 2009 (497 Punkte) bis 2012 (508 Punkte) kann als statistisch signifikant abgesichert werden. Somit kann die Zunahme der Lesekompetenz der Schülerinnen und Schüler in Deutschland als substantiell erachtet wer-

den. Abbildung 7.9 veranschaulicht die Entwicklung der mittleren Lesekompetenz in Deutschland über die PISA-Erhebungsrounden.

Die positive Weiterentwicklung der Lesekompetenz in Deutschland kann auch im Vergleich mit anderen OECD-Staaten hervorgehoben werden. Zwar gibt es einige Nachbarstaaten Deutschlands, in denen die Lesekompetenz der Schülerinnen und Schüler über die PISA-Erhebungsrounden ebenfalls angestiegen ist (z. B. Schweiz, Polen), jedoch blieb sie bei einem Großteil der deutschen Nachbarländer über die Zeit nahezu unverändert (z. B. Belgien, Niederlande, Dänemark). Dass es keineswegs selbstverständlich ist, eine hohe Lesekompetenz über die Zeit aufrecht zu erhalten, zeigen zum Beispiel die Veränderungen in Schweden oder auch Finnland. In Schweden ist die Lesekompetenz der Jugendlichen seit PISA 2000 von ursprünglich 516 Punkten auf aktuell 483 Punkte gesunken. Auch in Finnland ist seit PISA 2006 ein deutlicher Rückgang der Leseleistung von 547 auf 524 Punkte festzustellen.

In Deutschland verringerte sich außerdem die Streuung der Leistungen im Lesetest. Dies bedeutet, dass die Unterschiede zwischen den Leistungsstärksten und Leistungsschwächsten über die Zeit geringer wurden. Während die Streuung bei PISA 2000 noch 111 Punkte betrug, wurde sie über die PISA-Erhebungsrounden immer kleiner und lag bei PISA 2012 mit 91 Punkten im OECD-Durchschnitt.

Da sich die Zusammensetzung der Schülerinnen und Schüler in den Schularten seit PISA 2000 erheblich verändert hat, können die Kennwerte für die einzelnen Schularten nicht sinnvoll über die Erhebungsrounden verglichen werden. Allenfalls können die Mittelwerte für die Gymnasien über die Zeit verglichen werden. Obwohl sich für diese

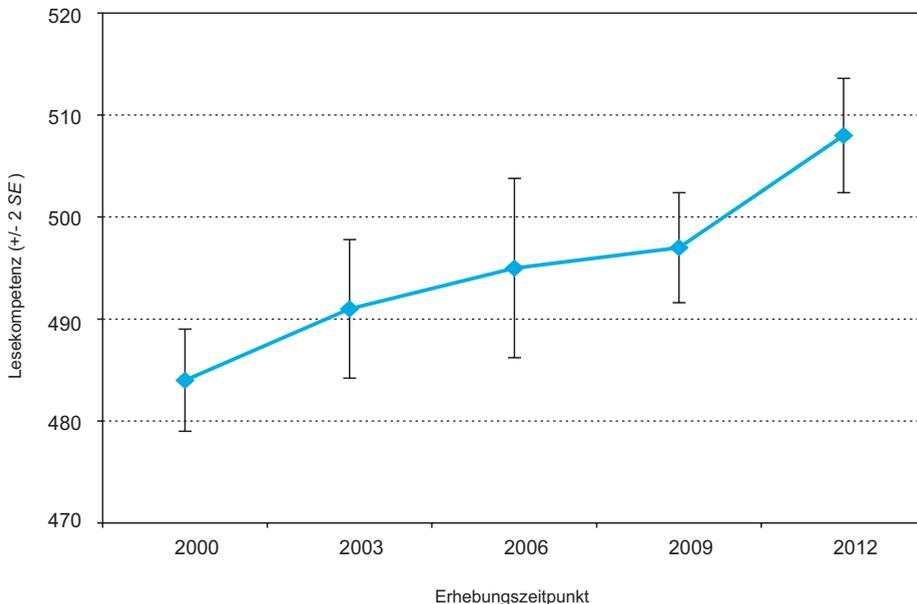


Abbildung 7.9: Veränderung der mittleren Lesekompetenz in Deutschland von 2000 bis 2012

Schulart die Bildungsbeteiligung deutlich vergrößert hat, rekrutiert sie nach wie vor den oberen Leistungsbereich. Außerdem ist an der Grundkonzeption gymnasialer Bildung in den vergangenen zwölf Jahren festgehalten worden. Bei PISA 2000 erreichten die Schülerinnen und Schüler an Gymnasien einen Mittelwert von 582 Punkten; bei PISA 2009 lag der Kennwert bei 575 Punkten, und 2012 bei 579 Punkten. Die durchschnittliche Leseleistung an den Gymnasien ist damit seit PISA 2000 ziemlich genau auf dem gleichen Niveau geblieben. Allerdings hat sich die Bildungsbeteiligung seit PISA 2000 deutlich geändert, von 28,3 Prozent auf 33,5 Prozent (PISA 2009) und schließlich auf 36,0 Prozent (PISA 2012). Die (um knapp 8 Prozent) gestiegene Bildungsbeteiligung am Gymnasium hat zu keinem Niveauverlust geführt: Im Durchschnitt lesen die Schülerinnen und Schüler an den Gymnasien genau so gut wie im Jahr 2000 – aber es sind deutlich mehr Schülerinnen und Schüler geworden, die entsprechend gut lesen.

Betrachtet man die Leistungsentwicklungen von Jungen und Mädchen getrennt über die PISA-Erhebungsrunden (vgl. Abbildung 7.10), so wird bei beiden Geschlechtern ein kontinuierlicher Aufwärtstrend sichtbar, wobei sich die Leistungen der Jungen denen der Mädchen nach wie vor nicht annähern. Obwohl die Mädchen seit PISA 2000 im durchschnittlichen beziehungsweise überdurchschnittlichen Fähigkeitsbereich liegen, konnten sie im Hinblick auf PISA 2012 ihre Leseleistungen erneut steigern.

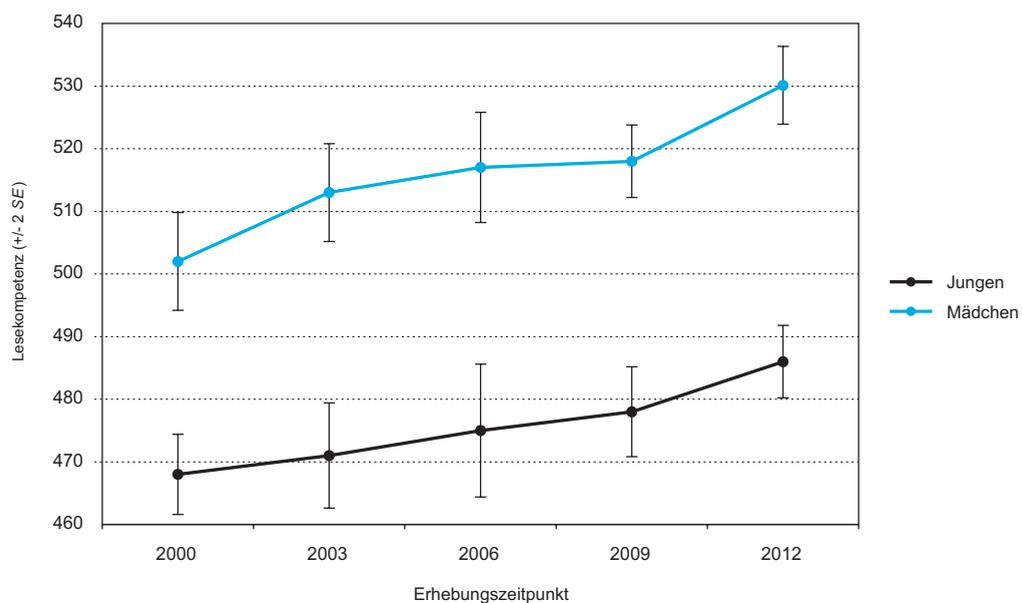


Abbildung 7.10: Veränderung der mittleren Lesekompetenz in Deutschland von 2000 bis 2012 für Mädchen und Jungen

7.4 Diskussion und Ausblick

Seit dem „PISA-Schock“ im Jahr 2000 und der damit einhergehenden Erkenntnis, dass die Schülerinnen und Schüler in Deutschland lediglich über unterdurchschnittliche Lesekompetenzen verfügen, hat sich einiges getan. Die einzelnen PISA-Erhebungsrounden ließen immer wieder Leistungszuwächse bei den Jugendlichen in Deutschland erkennen. Dieser Trend setzte sich auch bei PISA 2012 fort. Hier lagen die Lesekompetenzen in Deutschland sogar erstmals signifikant über dem OECD-Durchschnitt. Darüber hinaus geht dieser Befund mit einer weiteren, wenn auch nicht signifikanten, Verringerung der Leistungsstreuung und einer Reduktion des Anteils von leistungsschwachen Schülerinnen und Schülern einher. Von dieser Verbesserung sind Mädchen und Jungen in gleicher Weise betroffen. Trotzdem bleibt die sehr große Leistungsdiskrepanz zwischen den Mädchen und Jungen in Deutschland weiterhin bestehen. Diese gilt es in Zukunft durch eine gezielte und frühzeitig beginnende Leseförderung der Jungen zu verringern.

Seit der Veröffentlichung der Ergebnisse der ersten PISA-Erhebungsrunde sind nun 12 Jahre vergangen, genügend Zeit also für die Implementierung kurz-, mittel- und langfristiger angelegter Maßnahmen zur Förderung der Lesekompetenz. Diese scheinen mit den aktuellen PISA-Ergebnissen erstmals sichtbare Früchte zu tragen. Vergleicht man jedoch die Leistungsveränderungen in der Lesekompetenz mit denen in den Bereichen Mathematik und Naturwissenschaften, so wird deutlich, dass die Verbesserung der Lesekompetenz von fünfzehnjährigen Schülerinnen und Schülern in Deutschland eher zögerlich voranging.

Da eine ausgeprägte Lesekompetenz eine bedeutsame Grundlage für die aktive Teilhabe an unserer Gesellschaft darstellt, ist im Besonderen die Reduktion des Anteils von sehr leistungsschwachen Schülerinnen und Schülern (Kompetenzstufe Ia und darunter) erstrebenswert. Dieser Anteil liegt bei PISA 2012 in Deutschland bei 14,5 Prozent und ist somit im Vergleich zu PISA 2009 zwar leicht gesunken, aber immer noch zu hoch. Gleichzeitig ist der Anteil leistungsstarker Leserinnen und Leser (Kompetenzstufe V und VI) im Vergleich zu 2009 leicht angestiegen, von 7,6 auf 8,9 Prozent. Trotzdem stellt sich die Frage, ob ein Anteil von knapp 9 Prozent leistungsstarker Leserinnen und Leser ausreichend ist. Betrachtet man zum Beispiel die OECD-Spitzenreiter Japan oder Korea, so wird deutlich, dass es im oberen Leistungsbereich in Deutschland durchaus noch Entwicklungspotential gibt. Dieser Eindruck kann auch durch den Vergleich mit direkten Nachbarstaaten Deutschlands untermauert werden, wie Frankreich (12,9 Prozent) oder Belgien (12,3 Prozent).

Die Ergebnisse der Analysen für die verschiedenen Schularten lassen sich nur sehr eingeschränkt mit den vorherigen PISA-Erhebungsrounden vergleichen, da die Bildungsbeteiligungen über die Zeit stark variieren und sich darüber hinaus in den verschiedenen Schularten unterschiedliche Veränderungen vollziehen. Die Ergebnisse für PISA 2012 zeigen jedoch, dass es zwischen den Schularten in Deutschland nach wie vor große Leistungsdifferenzen gibt. So liegen zwischen der leistungsschwächsten Schul-

art (Hauptschule) und der leistungsstärksten Schulart (Gymnasium) mehr als 160 Leistungspunkte; dies entspricht Leistungsunterschieden von über zwei Kompetenzstufen. Trotz der in den letzten Jahren gestiegenen Bildungsbeteiligung an den Gymnasien hat sich dort das Leistungsniveau nicht verändert und liegt mit aktuell 579 Punkten weiterhin signifikant über der mittleren Lesekompetenz in Deutschland.

Zusammenfassend kann gesagt werden, dass sich die Lesekompetenz der Schülerinnen und Schüler in Deutschland seit PISA 2000 deutlich verbessert hat und inzwischen über dem Durchschnitt der OECD-Staaten einzuordnen ist. Das Bewusstsein, dass die Lesekompetenz auch nach der Grundschule – systematisch und in allen Fächern – weiter gefördert werden muss, scheint sich inzwischen verbreitet zu haben. Trotz der erfreulichen Leistungszunahme gibt es jedoch weiterhin Handlungsbedarf, um die Lesekompetenz in Deutschland insgesamt auf ein noch höheres Niveau zu entwickeln, die Anteile schwacher Leserinnen und Leser weiter zu reduzieren und den Anteil von Schülerinnen und Schülern in den Spitzengruppen auszubauen.

Literatur

- Adams, R. & Wu, M. (2007). A generalized form of the Rasch Model. In M. v. Davier & C. H. Carstensen (Hrsg.), *Multivariate and mixture distribution Rasch models. Extensions and applications* (Statistics for social and behavioral sciences, 1. Aufl., S. 57–76). New York: Springer.
- Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J. et al. (2007). *Expertise. Förderung von Lesekompetenz*. Bonn, Berlin: BMBF. Zugriff am 20.09.2013. Verfügbar unter: http://www.bmbf.de/pub/bildungsreform_band_siebzehn.pdf
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. 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. 69–137). Opladen: Leske + Budrich.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I. et al. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Bayerisches Staatsministerium für Unterricht und Kultus und Staatsinstitut für Schulqualität und Bildungsforschung (Hrsg.). (2010). *ProLesen. Auf dem Weg zur Leseschule. Leseförderung in den gesellschaftswissenschaftlichen Fächern. Aufsätze und Materialien aus dem KMK-Projekt ProLesen*. Donauwörth: Auer Verlag.
- Bos, W., Tarelli, I., Bremerich-Vos, A. & Schwippert, K. (Hrsg.). (2012). *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bussière, P., Hébert, R. & Knighton, T. (2010). Educational outcomes at age 21 associated with reading ability at age 15. *School Libraries in Canada*, 28, 98.
- Council of Europe (1996). *Modern Languages: Learning, Teaching, Assessment. A Common European Framework of Reference*. Strasbourg: Council of Europe.

- Drechsel, B. & Artelt, C. (2007). Lesekompetenz. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme et al. (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 225–247). Münster: Waxmann.
- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2002). *PISA 2000 – Zentrale Handlungsfelder. Zusammenfassende Darstellung der laufenden und geplanten Maßnahmen in den Ländern*. Beschluss der 299. Kultusministerkonferenz vom 17./18.10.2002. Zugriff am 20.09.2013. Verfügbar unter <http://www.kmk.org/fileadmin/pdf/PresseUndAktuelles/2002/massnahmen.pdf>
- KMK = Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2004). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss*. Neuwied: Luchterhand.
- 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 et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23–71). Münster: Waxmann.
- OECD. (2009). *PISA 2009 assessment framework. Key competencies in reading, mathematics and science*. Paris: OECD.
- OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Redder, A., Schwippert, K., Hasselhorn, M., Forschner, S., Fickermann, D. & Ehlich, K. et al. (2011). *Bilanz und Konzeptualisierung von strukturierter Forschung zu „Sprachdiagnostik und Sprachförderung“*. ZUSE-Berichte Band 2 [Elektronische Version]. Hamburg: Hamburger Zentrum für Unterstützung der wissenschaftlichen Begleitung und Erforschung schulischer Entwicklungsprozesse (ZUSE).
- Schaffner, W., Schiefele, U., Drechsel, B. & Artelt, C. (2004). Lesekompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 93–110). Münster: Waxmann.
- Schneider, W., Baumert, J., Becker-Mrotzek, M., Hasselhorn, M., Kammermeyer, G., Rauschenbach, T. et al. (2013). *Expertise „Bildung durch Sprache und Schrift (BISS)“*. Zugriff am 20.09.2013. Verfügbar unter http://www.bmbf.de/pubRD/BISS_Expertise.pdf
- Staatsinstitut für Schulqualität und Bildungsforschung. (2005). *Leseförderung nach PISA*. Zugriff am 20.09.2013. Verfügbar unter <http://www leseforum.bayern.de/download.asp?DownloadFileID=8f82cec95e5e709390d5af5ae4dcbc80>
- van der Linden, J. W., Veldkamp, B. P. & Carlson, J. E. (2004). Optimizing balanced incomplete block designs for educational assessments. *Applied Psychological Measurement*, 28(5), 317–331.

8

Soziale Herkunft als Bedingung der Kompetenzentwicklung

Katharina Müller und Timo Ehmke

8.1 Einleitung

Zu den großen Herausforderungen von Bildungssystemen gehört es, allen Mitgliedern der Gesellschaft gerechte Chancen zum Lernen und zur Entwicklung von Kompetenzen zu bieten. Seit der ersten Erhebungsrunde spielt die Frage nach der Bildungsgerechtigkeit (*Equity*) eine maßgebliche Rolle bei PISA (vgl. OECD, 2001). Allerdings ist nicht immer klar, was mit „gerechten“ Bildungschancen wirklich gemeint ist. Spätestens seit den Diskussionen in den siebziger Jahren ist bekannt, dass „gleiche“ Chancen im Bildungsbereich nicht unbedingt „gerecht“ sind (z. B. Jencks, 1973). Wenn Kinder unter wenig anregenden Bedingungen im Elternhaus aufwachsen, werden Entwicklungsdefizite durch eine Gleichbehandlung im Bildungssystem nicht kompensiert. „Gerechte“ Chancen auf Bildung verlangen häufig eine „ungleiche“ Behandlung. Im Kontext von Bildung stehen die Ansprüche auf Gerechtigkeit und Gleichheit in einem nicht einfach aufzulösenden Spannungsverhältnis (vgl. VBW, 2007). Wenn Eltern sich in besonderem Maße und geschickt für die Entwicklung ihres Kindes engagieren, ist das aus pädagogischer Sicht höchst erfreulich. Ein Vorwurf, sie würden damit zu ungleichen Bildungschancen beitragen, wäre höchst ungerecht. Auf der anderen Seite ist es ebenso ungerecht, wenn ein Kind aufgrund eines vernachlässigenden Elternhauses seine Talente nicht entfalten kann. Um hier für gerechte Chancen zu sorgen, muss das Bildungssystem möglicherweise Maßnahmen ergreifen und zusätzliche Leistungen erbringen.

Was kann nun eine internationale Vergleichsstudie in Anbetracht dieser komplexen Problematik an nützlichen Erkenntnissen über Bildungssysteme bereitstellen? Zunächst liefert PISA wertvolle Informationen, indem Unterschiede in den familiären Lebensverhältnissen der getesteten Jugendlichen nachgezeichnet werden. Der Blick richtet sich dabei auf strukturelle Merkmale der Elternhäuser, von denen man annimmt, dass sie sich auf die Entwicklung von Kompetenzen und die Bildungskarriere auswirken. In einem zweiten Schritt beschreibt PISA mit statistischen Mitteln, wie sehr diese Merkmale der Elternhäuser mit Unterschieden in den Kompetenzen (Mathematik, Naturwissenschaften, Lesen) gekoppelt sind. Wenn junge Menschen aus Familien mit unterschiedlichen sozialen Lagen Differenzen in den Kompetenzen aufweisen, spricht man

von Disparitäten. Der internationale Vergleich sozialer Disparitäten ist dann von besonderem Interesse, wenn es Staaten gibt, in denen hohe Kompetenzniveaus bei gleichzeitig geringen sozialen Unterschieden erreicht werden.

In PISA 2000 und PISA 2003 war für Deutschland berichtet worden, dass die erreichten Kompetenzen von Jugendlichen eng mit ihrer sozialen Herkunft zusammenhängen. Dies zeigte sich insbesondere an den Unterschieden in der Lesekompetenz zwischen Schülerinnen und Schülern aus unterschiedlichen sozialen Lagen, aber auch in Bezug auf die Mathematik und die Naturwissenschaften. Im internationalen Vergleich gab es damals nur wenige OECD-Staaten, in denen die Kopplung zwischen dem sozio-ökonomischen Status der Eltern und den gemessenen Kompetenzen ihrer Kinder so eng war wie in Deutschland (Baumert & Schümer, 2001; Ehmke, Hohensee, Heidemeier & Prenzel, 2004; OECD, 2001; 2004).

In der dritten und vierten Erhebungsrunde von PISA im Jahr 2006 und 2009 wurde für Deutschland festgestellt, dass die Kennwerte für den sozialen Gradienten der Lesekompetenz im Vergleich zu den Befunden aus PISA 2000 bedeutsam abgenommen hatten (Ehmke & Baumert, 2007; Ehmke & Jude, 2010; OECD, 2007; 2010). Die nationalen Ländervergleichsstudien zur Überprüfung der Bildungsstandards in Deutschland wiesen zudem darauf hin, dass bis heute zwischen den Bundesländern bedeutsame Unterschiede im Grad der Kopplung zwischen sozialer Herkunft und den erreichten Kompetenzen bestehen (Knigge & Leucht, 2010; Pant, Stanat, Schroeders, Roppelt, Siegle & Pöhlmann, 2013; Richter, Kuhl & Pant, 2012).

Mit PISA 2012 liegen jetzt Ergebnisse für den fünften Erhebungszeitpunkt im internationalen Vergleich vor. Veränderungen in den Kennwerten für soziale Disparitäten können inzwischen über einen Zeitraum von mehr als einem Jahrzehnt untersucht werden. Der vorliegende Bericht behandelt insbesondere folgende Fragen: (1) Wie hoch ist der Zusammenhang zwischen Merkmalen der sozialen Herkunft der Jugendlichen und ihrer Mathematikkompetenz in den OECD-Staaten? (2) Inwieweit hat sich dieser Zusammenhang zwischen den Erhebungen PISA 2003 und PISA 2012 in den einzelnen OECD-Staaten verändert? (3) Wie unterscheiden sich Jugendliche aus unterschiedlichen sozialen Lagen in Deutschland hinsichtlich bildungsrelevanter Merkmale ihrer Elternhäuser? (4) Haben sich zwischen PISA 2000 und 2012 die sozialen Disparitäten in den Kompetenzen und in der Bildungsbeteiligung für die Kohorte der Fünfzehnjährigen verringert oder erhöht?

Zu den beiden ersten Fragestellungen werden international vergleichende Analysen zwischen den OECD-Staaten berichtet. Dabei wurde geprüft, wie sich die Kopplung zwischen sozialer Herkunft und Mathematikkompetenz der Jugendlichen in den Erhebungen seit PISA 2003 darstellt. Den Fragestellungen drei und vier wird im Rahmen nationaler Analysen nachgegangen. Diese zielen auf die detailliertere Untersuchung sozialer Disparitäten in Deutschland. Um langfristige Entwicklungen abbilden zu können, wird bei diesen Analysen zusätzlich die Lesekompetenz einbezogen. Da Lesen im Jahr 2000 erstmals Hauptdomäne war, lassen sich damit Trendanalysen über einen Zeitraum von zwölf Jahren darstellen.

8.2 Erfassung der sozialen Herkunft in PISA

In der empirischen Bildungsforschung wird die soziale Herkunft von Schülerinnen und Schülern in der Regel durch die sozioökonomische Stellung der Familie bestimmt. Der sozioökonomische Status einer Person bezeichnet die relative Position in einer vertikalen sozialen Gliederung, die mit unterschiedlichen Gestaltungs- und Einflussmöglichkeiten verbunden ist. Diese Position wird von Merkmalen wie Einkommen, Macht oder Prestige bestimmt. Neben der sozioökonomischen Stellung werden in kapitaltheoretischen Ansätzen noch weitere Aspekte der sozialen Herkunft unterschieden, nämlich die kulturelle Praxis und das soziale Kapital von Familien (Bourdieu, 1982; 1983). Das kulturelle Kapital wird durch die kulturbezogenen Sachgüter und Bildungszertifikate, die in Familien vorhandenen Wahrnehmungs-, Denk- und Deutungsmuster sowie Wertorientierungen und Einstellungen bestimmt. Zum sozialen Kapital werden das soziale Netzwerk und die sozialen Beziehungen gezählt, in denen Kinder und Jugendliche aufwachsen. Während die sozioökonomische Stellung die mit den Einkommensverhältnissen und dem materiellen Wohlstand verknüpften Ressourcen in einer Familie widerspiegelt, werden mit dem sozialen und kulturellen Kapital Lebensbedingungen berücksichtigt, die mit dem Lebensstil und dem sozialen Milieu verbundenen sind. Auch diese Merkmale können als Ressourcen verstanden werden, durch die sich unterschiedlich lernförderliche Umgebungen und damit differentielle Entwicklungsmilieus abbilden lassen.

In der PISA-Berichterstattung kommen mit dem *International Socio-Economic Index of Occupational Status* (ISEI), der *Erikson-Goldthorpe-Portocarero-Klassifikation* (EGP) und dem *Index of Economic, Social and Cultural Status* (ESCS) drei verschiedene Indizes zum Einsatz, um die soziale Herkunft der Schülerinnen und Schüler genauer zu beschreiben. Die drei Indizes, die alle auf der *International Standard Classification of Occupation* (ISCO) basieren, berücksichtigen jeweils unterschiedliche Indikatoren sozialer Herkunft und beleuchten unterschiedliche Aspekte sozial bedingter Bildungsungleichheiten. Im Folgenden werden diese in PISA verwendeten Indizes (Indikatorensysteme) zur Erfassung der sozialen Herkunft genauer beschrieben.

In den Fragebögen wurden die Schülerinnen und Schüler und in einigen Staaten auch deren Eltern um Angaben über die Berufstätigkeit der Bezugspersonen (Eltern beziehungsweise Erziehungsberechtigte) gebeten. Die Angaben wurden mit Hilfe der Internationalen Standardklassifikation der Berufe kodiert (ISCO). Mit der ISCO-Kodierung lassen sich die Berufe auf der Grundlage der jeweils erforderlichen Tätigkeiten international vergleichbar klassifizieren und auf verschiedenen Ebenen aggregieren (International Labour Office, 1990). In PISA 2012 erfolgt die Klassifikation international erstmals auf der Grundlage der 2011 eingeführten revidierten Fassung ISCO-08 (International Labor Office, 2012). Bei dieser revidierten Fassung wurde das bisherige Klassifikationssystem ISCO-88 aktualisiert und mit Blick auf neue Berufsgruppen, etwa aus der IT-Branche, stärker ausdifferenziert. Auf der Grundlage dieser ISCO-Kodierung werden zwei aussagekräftige Indizes gebildet: der *International Socio-Economic Index of Occupational Status* (ISEI) und die *EGP-Klassifikation*.

Der von Ganzeboom und Kollegen (Ganzeboom, de Graaf & Treiman, 1992; Ganzeboom & Treiman, 2003) entwickelte *International Socio-Economic Index of Occupational Status* (ISEI) bestimmt die sozioökonomische Stellung auf der Grundlage von beruflichen Qualifikationsmerkmalen, dem Bildungsabschluss und dem Einkommen. Die Autoren gehen davon aus, dass Berufe und die mit ihnen verbundenen erforderlichen Qualifikationen den Bildungsabschluss in Einkommen umsetzen und dadurch einer sozialen Gliederung zugeordnet werden können. Entsprechend der Revision der ISCO-Codes wurde auch der ISEI aktualisiert (Ganzeboom, 2010; Ganzeboom & Treiman, 2012). Für die vorliegenden Analysen wurde jeweils der höchste ISEI-Wert (Highest ISEI = HISEI) im Elternhaus – also von der Mutter *oder* dem Vater – herangezogen. Da außer den sozioökonomischen Ressourcen keine weiteren Informationen in diesen Index eingehen, ist der HISEI gut geeignet, um den Zusammenhang zwischen den im Elternhaus verfügbaren sozioökonomischen Ressourcen und den Kompetenzen ihrer Kinder zu beschreiben.

Im Vergleich zur eindimensionalen Gliederung des sozioökonomischen Index (HISEI) werden mit der sogenannten *EGP-Klassifikation* (Erikson, Goldthorpe & Portocarero, 1979; Erikson & Goldthorpe, 2002) qualitative Abstufungen zwischen sozialen Klassen vorgenommen. Die Grundlage der EGP-Klassifikation ist ein Kategoriensystem, bei dem die Berufe nach der Art der Tätigkeit, der Stellung im Beruf und der Weisungsbefugnis geordnet werden. Die Verbindung von gegliederter Abstufung und typologischer Klassifikation der Berufe erlaubt es, im Vergleich zum sozioökonomischen Index Berufsgruppen theoretisch fundiert anschaulicher zu beschreiben. Anhand der EGP-Klassifikation können Unterschiede in der sozialen Entwicklungsumwelt der Schülerinnen und Schüler sichtbar gemacht werden, die mit dem Einkommen der Eltern und deren Zugang zu Bildung, Macht und gesellschaftlicher Anerkennung verknüpft sind. Deshalb wurde in den nationalen Analysen zu PISA seit Beginn auch die EGP-Klassifikation vorgenommen und berichtet. Dabei wurde und wird statt des elfstufigen Modells von Erikson und Kollegen (1979) ein Klassifikationsschema mit sechs Klassen verwendet.

Zur Quantifizierung sozialer Disparitäten wird speziell in PISA auf internationaler Ebene ein globaler Index gebildet, der soziokulturelle und sozioökonomische Merkmale der sozialen Herkunft zusammenfasst. Der (*z*-standardisierte) *Index of Economic, Social and Cultural Status* (ESCS) berücksichtigt folgende Herkunftsvariablen: den sozioökonomischen Index, den elterlichen Bildungsabschluss und Informationen über den Besitz von Kultur- und Wohlstandsgütern (OECD, 2009). Die Vorhersagekraft des ESCS ist insgesamt höher als die der einzelnen in ihm enthaltenen Indikatoren, denn der ESCS wurde bei PISA mit der Zwecksetzung gebildet, durch Herkunftsmerkmale Unterschiede in den Kompetenzen möglichst breit vorherzusagen. Der ESCS aggregiert und verliert zugleich Information: Die differenzierte Betrachtung der Einzelindikatoren gibt die Möglichkeit, die Bedeutung der verschiedenen (sozioökonomischen, sozialen und kulturellen) Komponenten des Konstrukts der Herkunft gegeneinander abzuschätzen (Ehmke & Siegle, 2005).

Bei der Interpretation von Analysen globaler Indikatorensysteme wie dem ESCS gilt es zu beachten, dass die erhöhte Prädiktivität mit Unschärfen in der theoretischen Grundlegung und der konzeptionellen Bedeutung einhergeht (Caro & Cortés, 2012). Da der ESCS über die verschiedenen PISA-Erhebungen hinweg zum Teil unterschiedlich operationalisiert und um länderspezifische Items ergänzt wurde, ist ferner zu berücksichtigen, dass eine vollständige Vergleichbarkeit zwischen den PISA-Erhebungsrounden nicht immer sichergestellt ist. Insgesamt sei darauf hingewiesen, dass bei Vergleichen über die Zeit aufgrund der Revision der ISCO-Kodierung bei allen genannten Indikatorensystemen in PISA 2012 mit Verschiebungen aufgrund veränderter Operationalisierungen zu rechnen ist.

8.3 Befunde zum Zusammenhang zwischen sozialer Herkunft und mathematischer Kompetenz im internationalen Vergleich

Im Zentrum des folgenden Abschnittes steht der internationale Vergleich der sozialen Herkunft der Schülerinnen und Schüler. In einem ersten Schritt wird anhand des HISEI der sozioökonomische Status der Fünfzehnjährigen in den OECD-Staaten betrachtet. Neben den Mittelwerten interessiert insbesondere die Streuung innerhalb der Staaten. Ferner wird beschrieben, ob die Schülerinnen und Schüler, deren Eltern eine höhere berufliche Stellung innehaben, höhere Werte in der Mathematikkompetenz aufweisen als Schülerinnen und Schüler aus Familien mit niedrigerem sozioökonomischem Status. Im zweiten Schritt wird der Zusammenhang zwischen der Mathematikkompetenz und der sozialen Herkunft anhand des ESCS-Index beleuchtet, der gleichzeitig ökonomische, kulturelle und soziale Indikatoren der sozialen Herkunft integriert. Abschließend wird auf die Veränderungen in den sozialen Disparitäten der Mathematikkompetenz zwischen PISA 2003 und 2012 eingegangen.

Kopplung zwischen Mathematikkompetenz und sozioökonomischem Status (HISEI)

Im ersten Abschnitt betrachten wir die Ergebnisse für den *Highest International Socio-Economic Index of Occupational Status* (HISEI). In Abbildung 8.1 sind der durchschnittliche Wert sowie die Streuung des sozioökonomischen Status für jeden OECD-Staat in Form von Perzentilbändern abgebildet. An der Länge des Perzentilbandes ist der Abstand zwischen den fünf Prozent der sozioökonomisch schwächsten und den fünf Prozent der sozioökonomisch stärksten Elternhäuser der getesteten Schülerinnen und Schüler zu erkennen. Die Breite des Bandes veranschaulicht also die Streuung des sozioökonomischen Status innerhalb eines Staats.

Deutschland liegt mit einem mittleren sozioökonomischen Status von 50.9 Punkten zusammen mit Slowenien, Estland und Japan im Bereich des OECD-Durchschnitts. Signifikant über dem OECD-Mittelwert liegt der sozioökonomische Status etwa in Norwegen, Finnland, Kanada oder der Schweiz. Einen bedeutsam geringeren Durchschnittswert erreichen beispielsweise die Eltern der Jugendlichen aus der Slowakischen Republik, Chile, Mexiko, Portugal oder der Türkei.

Betrachtet man die Streuungen des HISEI (festgemacht an den Standardabweichungen), dann findet man innerhalb der Staaten unterschiedliche Verteilungen der sozialen Lagen. Ein sehr heterogenes Sozialgefüge lässt sich in Griechenland ($SD = 22.9$), Luxemburg ($SD = 22.0$) und Mexiko ($SD = 21.9$) feststellen. Im Vergleich zum OECD-Durchschnitt sind die Streuungen im sozioökonomischen Status in Korea ($SD = 18.1$), der Tschechischen Republik ($SD = 19.3$) und Norwegen ($SD = 19.3$) besonders gering. Die Streuung des HISEI in Deutschland ($SD = 20.6$) liegt im OECD-Durchschnitt.

Vergleicht man die Kennwerte des HISEI über die letzten PISA-Erhebungen, so ist der OECD-Mittelwert leicht angestiegen (PISA 2009: 49.3 Punkte; PISA 2012: 50.6 Punkte), während sich die Standardabweichung über die OECD-Staaten deutlicher erhöht hat (PISA 2009: 16.0 Punkte; PISA 2012: 20.8 Punkte). Auch für Deutschland zeigt sich diese Tendenz: Der Mittelwert und insbesondere die Standardabweichung für den sozioökonomischen Status fallen in PISA 2012 höher aus als in der vorherigen Erhebung (PISA 2009: $M = 48.9$ Punkte, $SD = 15.6$ Punkte; PISA 2012: $M = 50.9$ Punkte, $SD = 20.6$ Punkte). Ähnliche Muster finden sich auch für eine Reihe anderer Staaten (etwa Norwegen oder Japan). Anhand einer Rekodierung der Berufsangaben im nationalen Datensatz für Deutschland konnte gezeigt werden, dass diese Veränderung weitgehend auf die neue und differenziertere Berufskodierung (ISCO-08) zurückgeführt werden kann, die in PISA 2012 erstmals Grundlage war, um die Indizes der sozialen Herkunft zu bilden.¹ Deshalb muss darauf hingewiesen werden, dass bei der Interpretation von Trendaussagen zum sozioökonomischen Status die Veränderung der Klassifikationssysteme berücksichtigt werden muss, denn der Anstieg in Mittelwert und Standardabweichung beeinflusst die Berechnung der Kopplung zwischen sozialer Herkunft und Kompetenz.

1 Bei der Rekodierung wurden die Berufsangaben zunächst sowohl anhand der ISCO-88 als auch der ISCO-08 Vorschriften kodiert und in HISEI-88 bzw. HISEI-08 Werte transformiert. Die so gebildeten Indizes korrelieren in hohem Maße miteinander, unterscheiden sich aber signifikant in der Standardabweichung. Bei der Analyse des Zusammenhangs zwischen Mathematikkompetenz und den beiden HISEI-Indizes zeigten sich tendenziell niedrigere Regressions- und Determinationskoeffizienten beim HISEI-88 als beim HISEI-08.

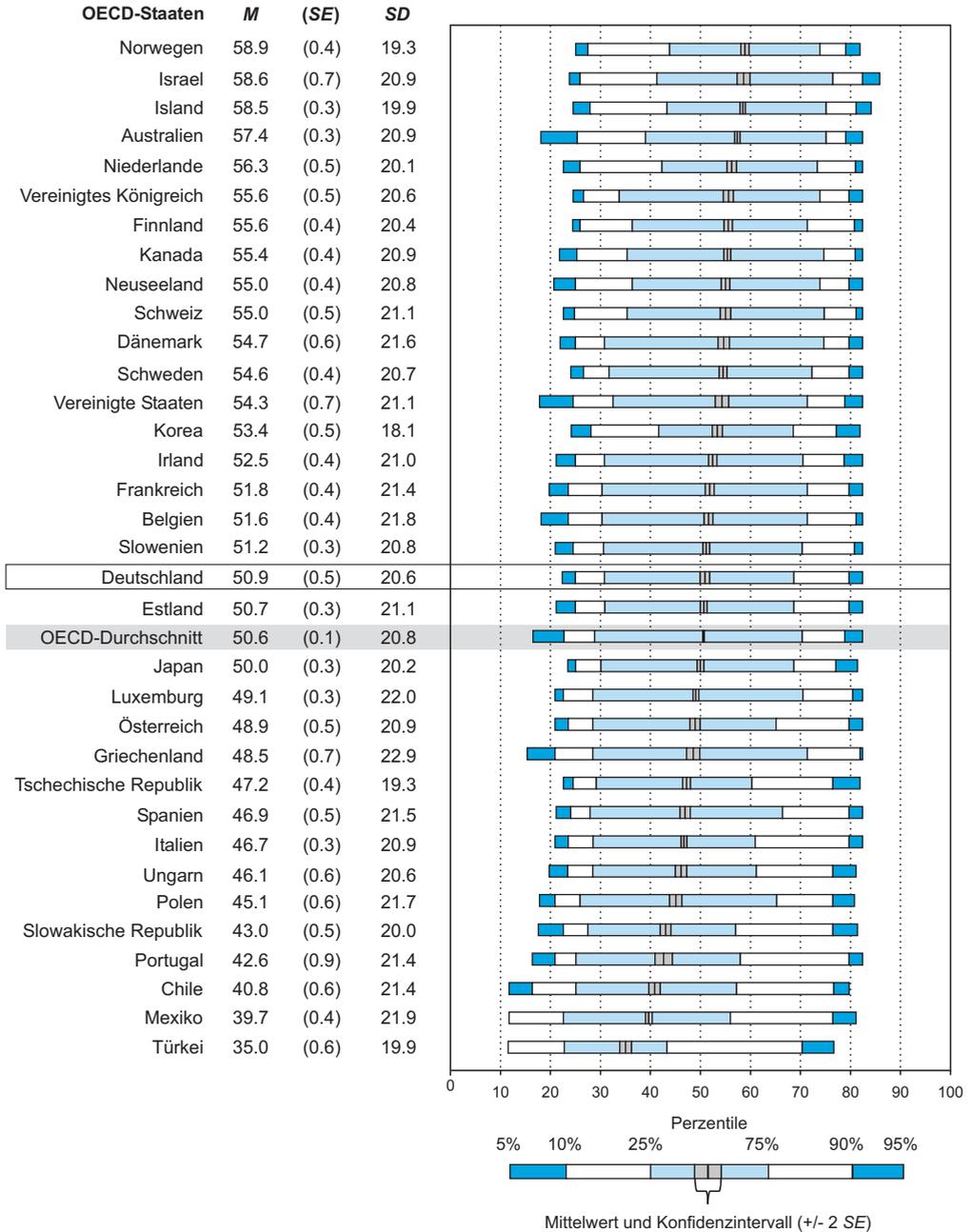


Abbildung 8.1: Verteilung des sozioökonomischen Status (HISEI) in den OECD-Staaten

Um die Kopplung zwischen sozialer Herkunft und Mathematikkompetenz zu beschreiben, werden in PISA üblicherweise zwei Kennwerte verwendet: der *soziale Gradient* und das Maß der *Varianzaufklärung*. Beide Kennwerte ergeben sich bei der linearen Regression, einem statistischen Verfahren, bei dem die Unterschiedlichkeit in einem Merkmal durch Unterschiede in einem oder mehreren anderen Merkmalen vorhergesagt werden.

Tabelle 8.1 gibt für alle OECD-Staaten die sozialen Gradienten und den Anteil der aufklärten Varianz bezogen auf die Mathematikkompetenz wieder (Spalten 4 und 6). In dieser Tabelle dient der höchste sozioökonomische Status (HISEI) als Prädiktor, das heißt, bei dieser Analyse bleiben Merkmale des *soziokulturellen* Hintergrundes, aber auch der Besitz von Wohlstandsgütern unberücksichtigt (der farblich gekennzeichnete Unterschied zum OECD-Mittelwert bezieht sich in Tabelle 8.1 auf die Stärke des Zusammenhangs R^2).

Die über alle OECD-Staaten hinweg ermittelte Steigung des *sozialen Gradienten* von 32 Punkten bedeutet, dass sich gemäß der Regressionsgeraden die durchschnittliche mathematische Kompetenz der Schülerinnen und Schüler um 32 Punkte erhöhen würde, wenn der sozioökonomische Status um eine Standardabweichung ($SD = 20.8$ Punkte, vgl. Abbildung 8.1) stiege. Der Achsenabschnitt der Mathematikkompetenz von 495 Punkten sagt hingegen jenen Kompetenzwert vorher, den eine Fünfzehnjährige oder ein Fünfzehnjähriger bei durchschnittlichem sozioökonomischem Status (OECD-Durchschnitt = 50.6 Punkte, vgl. Abbildung 8.1) erlangt.

Der an der Steigung des sozialen Gradienten gemessene Zusammenhang zwischen der Mathematikkompetenz und dem sozioökonomischen Status (HISEI) ist in allen OECD-Staaten positiv. Demnach fallen in allen Staaten die Kompetenzwerte mit zunehmendem sozioökonomischem Status tendenziell höher aus. Beim Vergleich der Steigungen der sozialen Gradienten zwischen den Staaten ist zu berücksichtigen, dass sowohl das Niveau des sozioökonomischen Index (HISEI, vgl. Abbildung 8.1) als auch die mathematische Kompetenz zwischen den Staaten (vgl. Kapitel 2) zum Teil erheblich variiert. Zu den Staaten, in denen sich die Steigung des sozialen Gradienten für den HISEI nicht signifikant vom OECD-Durchschnitt unterscheidet, gehören das Vereinigte Königreich, die Niederlande, Österreich, Polen, Belgien, die Schweiz, Spanien, Griechenland, die Vereinigten Staaten, die Türkei und Chile. Zur Gruppe der Staaten, bei denen der Zusammenhang zwischen Mathematikkompetenz und sozioökonomischem Status signifikant niedriger ist als im OECD-Mittel, zählen etwa Japan, Korea, Kanada, Italien, Irland oder auch die skandinavischen Staaten Finnland, Norwegen, Island, Dänemark und Schweden. Am höchsten ausgeprägt ist die Steigung des sozialen Gradienten für den HISEI in Portugal, Neuseeland, Frankreich, der Tschechischen Republik, Ungarn, Luxemburg, Israel, Slowenien und der Slowakischen Republik. Auch Deutschland gehört zu der Gruppe von Staaten, in denen die Steigung des sozialen Gradienten signifikant über dem Durchschnittswert der OECD liegt.

Während der soziale Gradient den Zusammenhang zwischen der Mathematikkompetenz und dem sozioökonomischen Status quantifiziert, ist die *aufgeklärte Varianz* eine statistische Maßzahl, die ausdrückt, zu welchem Anteil sich die Unterschiede in den

erreichten Kompetenzen in einem Staat durch die soziale Herkunft vorhersagen lassen. Die gesamte Varianz der Kompetenzwerte wird dabei zerlegt in einen Anteil, der auf die soziale Herkunft als Vorhersagekriterium zurückgeht, und in einen Restanteil, der sich nicht auf den sozioökonomischen Index zurückführen lässt. Im OECD-Durchschnitt liegt die Varianzaufklärung durch den HISEI bei 11.7 Prozent (vgl. Tabelle 8.1, Spalte 6).

Das Maß der aufgeklärten Varianz kann für den Vergleich der Staaten genutzt werden: Je höher der prozentuale Anteil der Gesamtvarianz der mathematischen Kompetenz in einem Staat durch den HISEI aufgeklärt wird, desto besser kann das Kompetenzniveau aus dem sozioökonomischen Status vorhergesagt werden. Aus Tabelle 8.1 wird ersichtlich, dass auch bei vergleichbaren Steigungen des sozialen Gradienten die Varianzaufklärung in einigen Staaten höher ist als in anderen. Die beiden Indikatoren erfassen also unterschiedliche Aspekte der Kopplung zwischen sozialer Herkunft und mathematischer Kompetenz: Der soziale Gradient zeigt an, um wie viele Punkte sich die Kompetenz durchschnittlich erhöhen würde, wenn der sozioökonomische Status um eine Standardabweichung anstiege. Die aufgeklärte Varianz gibt Auskunft darüber, wie präzise sich die Unterschiede in der Kompetenz durch den sozioökonomischen Status vorhersagen lassen.

In Deutschland wie auch in Neuseeland, Frankreich, Portugal oder Luxemburg ist der prozentuale Anteil der Unterschiede in der Mathematikkompetenz, die sich durch den sozioökonomischen Status erklären lassen, überdurchschnittlich hoch. In Japan, Korea, Kanada, Finnland oder Schweden hingegen ist dieser prozentuale Anteil niedriger als im OECD-Durchschnitt.

Kopplung zwischen Mathematikkompetenz und ökonomischem, kulturellem und sozialem Status (ESCS)

Bei der Berechnung des sozialen Gradienten auf der Basis des ESCS-Index, der sowohl ökonomische, kulturelle als auch soziale Herkunftsmerkmale berücksichtigt, zeigt sich wie erwartet, dass die Vorhersagekraft hinsichtlich der Kompetenzunterschiede höher ausfällt. Die Ergebnisse für den über den erweiterten Index ermittelten sozialen Gradienten sind in Tabelle 8.2 dargestellt (der farblich gekennzeichnete Unterschied zum OECD-Mittelwert bezieht sich in Tabelle 8.2 auf die Stärke des Zusammenhangs R^2).² Im OECD-Durchschnitt liegt die Steigung bei 39 Punkten, das heißt, bei einem um eine Standardabweichung höheren ESCS ($SD = 0.9$, vgl. Tabelle 8.4, Spalte 12) würde die mathematische Kompetenz der Schülerinnen und Schüler um 39 Punkte höher ausfallen. Hier zählt Deutschland zusammen mit Staaten wie Japan, Korea, Schweden, den Niederlanden, dem Vereinigten Königreich, der Schweiz, Irland, Slowenien, Österreich, Dänemark, Polen und Luxemburg zur Gruppe jener Staaten, in denen der soziale Gra-

2 Der anhand einer Rangkorrelation berechnete Vergleich der Rangpositionen zeigt, dass der Zusammenhang zwischen ESCS und HISEI beim Achsenabschnitt bei $r = 0.97$, beim sozialen Gradienten bei $r = 0.73$ und bei der Stärke des Zusammenhangs bei $r = 0.91$ liegt. Zur inhaltlichen Vertiefung der verschiedenen Komponenten des ESCS-Index siehe Analysen in Tabelle 8.5.

Tabelle 8.1: Zusammenhang zwischen Mathematikkompetenz und sozioökonomischem Status im internationalen Vergleich (Prädiktorvariable: HISEI)

OECD-Staaten	Mathematikkompetenz		Steigung des sozialen Gradienten		Stärke des Zusammenhangs	
	Achsenabschnitt	(SE)	Steigung	(SE)	R ²	(SE)
Japan	542	(3.4)	22	(2.4)	4.9	(1.0)
Korea	552	(4.1)	26	(2.7)	5.0	(1.0)
Norwegen	483	(2.7)	25	(1.9)	6.4	(0.9)
Finnland	516	(1.7)	24	(1.4)	7.0	(0.8)
Island	487	(1.8)	26	(1.9)	7.2	(1.0)
Mexiko	425	(1.2)	20	(0.9)	7.5	(0.6)
Kanada	517	(1.7)	25	(1.1)	7.8	(0.6)
Australien	500	(1.6)	28	(1.1)	8.1	(0.6)
Estland	522	(1.9)	24	(1.5)	8.2	(0.9)
Türkei	472	(5.5)	28	(2.6)	8.3	(1.4)
Italien	493	(1.9)	29	(1.1)	9.3	(0.6)
Schweden	478	(1.8)	29	(1.7)	9.6	(1.0)
Schweiz	526	(2.7)	30	(1.7)	10.0	(1.1)
Vereinigtes Königreich	493	(2.5)	31	(1.7)	10.3	(1.0)
Niederlande	519	(3.2)	31	(2.2)	10.4	(1.3)
Irland	502	(2.0)	28	(1.6)	10.5	(1.1)
Vereinigte Staaten	480	(2.9)	30	(1.6)	11.1	(1.1)
Dänemark	498	(1.6)	28	(1.6)	11.7	(1.3)
Österreich	509	(2.5)	33	(2.0)	12.0	(1.3)
Spanien	491	(1.7)	30	(1.1)	12.2	(0.8)
Belgien	518	(1.7)	35	(1.8)	12.5	(1.1)
Griechenland	458	(1.9)	30	(1.7)	13.1	(1.4)
Polen	527	(3.4)	33	(2.2)	13.5	(1.5)
Tschechische Republik	508	(2.6)	39	(2.2)	13.9	(1.3)
Neuseeland	498	(2.2)	38	(1.6)	14.3	(1.2)
Slowenien	502	(1.3)	36	(1.4)	14.4	(1.1)
Israel	460	(3.8)	42	(3.0)	16.2	(2.0)
Deutschland	520	(2.8)	40	(2.0)	15.6	(1.4)
Frankreich	498	(2.4)	38	(1.8)	15.9	(1.2)
Portugal	505	(2.7)	38	(1.9)	16.8	(1.7)
Chile	440	(2.8)	34	(2.1)	17.7	(1.9)
Slowakische Republik	505	(2.9)	45	(2.9)	17.9	(1.9)
Ungarn	490	(2.9)	41	(2.8)	18.3	(2.0)
Luxemburg	495	(1.3)	41	(1.3)	19.7	(1.1)
OECD-Durchschnitt	495	(0.4)	32	(0.3)	11.7	(0.2)

Anmerkung: Prädiktorvariable ist der höchste sozioökonomische Status (HISEI), z-standardisiert am OECD-Durchschnitt. Der farblich gekennzeichnete Unterschied zum OECD-Durchschnitt bezieht sich auf die Stärke des Zusammenhangs

signifikant über dem OECD-Durchschnitt
 nicht signifikant verschieden vom OECD-Durchschnitt
 signifikant unter dem OECD-Durchschnitt

Tabelle 8.2: Zusammenhang zwischen Mathematikkompetenz und ökonomischem, kulturellem und sozialem Status im internationalen Vergleich (Prädiktorvariable: ESCS)

OECD-Staaten	Mathematikkompetenz		Steigung des sozialen Gradienten		Stärke des Zusammenhangs	
	Achsenabschnitt	(SE)	Steigung	(SE)	R ²	(SE)
Norwegen	476	(2.8)	32	(2.4)	7.4	(1.0)
Island	470	(2.1)	31	(2.1)	7.7	(1.0)
Estland	518	(1.9)	29	(1.7)	8.6	(0.9)
Finnland	508	(1.9)	33	(1.8)	9.4	(0.9)
Kanada	508	(1.6)	31	(1.2)	9.4	(0.7)
Japan	541	(3.3)	41	(3.9)	9.8	(1.6)
Korea	553	(3.9)	42	(3.3)	10.1	(1.4)
Italien	487	(1.8)	30	(1.2)	10.1	(0.6)
Mexiko	435	(1.4)	19	(0.8)	10.4	(0.8)
Schweden	471	(1.9)	36	(1.9)	10.6	(1.1)
Australien	496	(1.6)	42	(1.3)	12.3	(0.8)
Niederlande	515	(3.2)	40	(3.1)	11.5	(1.7)
Vereinigtes Königreich	486	(2.6)	41	(2.4)	12.5	(1.2)
Schweiz	525	(2.7)	38	(1.8)	12.8	(1.2)
Türkei	494	(6.6)	32	(2.4)	14.5	(1.8)
Irland	497	(2.0)	38	(1.8)	14.6	(1.2)
Vereinigte Staaten	476	(2.7)	35	(1.7)	14.8	(1.3)
Belgien	511	(1.8)	43	(1.9)	15.0	(1.3)
Griechenland	456	(1.9)	35	(1.8)	15.5	(1.5)
Slowenien	499	(1.3)	42	(1.5)	15.6	(1.0)
Spanien	492	(1.6)	34	(1.1)	15.8	(1.0)
Österreich	503	(2.5)	43	(2.2)	15.8	(1.5)
Tschechische Republik	503	(2.5)	51	(2.7)	16.2	(1.5)
Dänemark	485	(1.7)	39	(1.7)	16.5	(1.4)
Polen	526	(3.2)	41	(2.4)	16.6	(1.7)
Deutschland	511	(2.6)	43	(2.0)	16.9	(1.4)
Israel	460	(3.8)	51	(2.6)	17.2	(1.5)
Luxemburg	488	(1.3)	37	(1.2)	18.3	(1.1)
Neuseeland	500	(2.2)	52	(1.9)	18.4	(1.3)
Portugal	506	(2.6)	35	(1.6)	19.6	(1.8)
Frankreich	500	(2.2)	57	(2.2)	22.5	(1.3)
Ungarn	490	(2.8)	47	(2.8)	23.1	(2.3)
Chile	443	(2.7)	34	(1.6)	23.1	(1.9)
Slowakische Republik	492	(2.6)	54	(2.9)	24.6	(2.1)
OECD-Durchschnitt	495	(0.5)	39	(0.4)	14.6	(0.2)

Anmerkung: Prädiktorvariable ist ESCS. Der farblich gekennzeichnete Unterschied zum OECD-Durchschnitt bezieht sich auf die Stärke des Zusammenhangs.

□ signifikant über dem OECD-Durchschnitt

□ nicht signifikant verschieden vom OECD-Durchschnitt

□ signifikant unter dem OECD-Durchschnitt

dient *nicht signifikant* vom OECD-Durchschnitt verschieden ist. In Norwegen, Island, Estland, Finnland, Kanada, Italien, Mexiko, der Türkei, den Vereinigten Staaten, Griechenland, Spanien, Portugal und Chile ist der Zusammenhang zwischen Merkmalen der sozioökonomischen bzw. soziokulturellen Herkunft und der Mathematikkompetenz geringer als im OECD-Durchschnitt.

Werden also neben Merkmalen des sozioökonomischen Status der Schülerinnen und Schüler auch kulturelle und soziale Ressourcen berücksichtigt, liegt die Kopplung zwischen der sozialen Herkunft und der Leistung in Deutschland verglichen mit den anderen Staaten im Bereich des OECD-Durchschnitts. Ähnlich verhält es sich mit der durch den ESCS-Index vorhergesagten Varianzaufklärung. Hier liegt Deutschland mit einer Varianzaufklärung von 16.9 Prozent ebenfalls im Bereich des OECD-Durchschnitts (14.6 Prozent).

Die Varianzaufklärung kann zudem gemeinsam mit der durchschnittlichen Kompetenzausprägung der Jugendlichen in den OECD-Staaten betrachtet werden, um eine Art Leistungs- und Disparitätsprofil zu erstellen (Abbildung 8.2). Neben dem durchschnittlichen Niveau der mathematischen Kompetenz sind die Staaten danach angeordnet, welcher Varianzanteil der Kompetenz durch den ökonomischen, sozialen und kulturellen Status (ESCS) aufgeklärt wird (vgl. Tabelle 8.2, Spalte 6).

Die Einteilung der Abbildung 8.2 in Quadranten gibt Auskunft darüber, welche der Staaten bezüglich der beiden Kriterien unter oder über dem Gesamtdurchschnitt der OECD (Mathematikkompetenz: 494 bzw. R^2 : 14.6 Prozent) liegen. Durch die gemeinsame Betrachtung des durchschnittlichen Kompetenzniveaus und des Maßes der Varianzaufklärung lassen sich somit Staaten identifizieren, in denen die Kompetenz hoch und die Stärke des Zusammenhangs zwischen sozialer Herkunft und Kompetenz niedrig ist. Diese Staaten sind im Quadranten oben links abgebildet. Ebenso kann durch die Zusammenschau auch die weniger wünschenswerte Konstellation abgebildet werden: Staaten, in denen die sozialen Disparitäten hoch sind und das Kompetenzniveau niedrig ausfällt (Quadrant unten rechts).

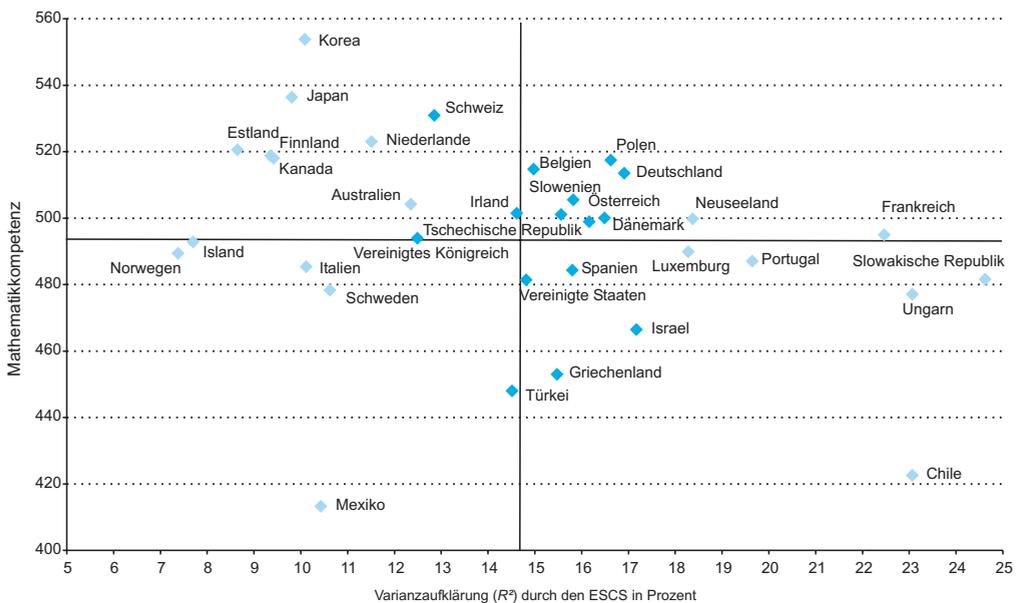
Zu den Staaten, in denen bei einem hohen Kompetenzniveau Unterschiede in den Kompetenzwerten unterdurchschnittlich stark durch die soziale Herkunft aufgeklärt werden, gehören insbesondere Japan, Korea, Finnland, Kanada, Australien und Estland. Bei einem sehr hohen Kompetenzniveau liegt für diese Staaten die Varianzaufklärung in der Mathematikkompetenz durch den ESCS signifikant unterhalb des OECD-Durchschnitts. Verglichen mit dem OECD-Durchschnitt wird das insgesamt hohe Kompetenzniveau in diesen Staaten weniger stark durch Disparitäten der ökonomischen, sozialen und kulturellen Herkunft vorhergesagt. Am Beispiel dieser Staaten wird deutlich, dass überdurchschnittliche mathematische Kompetenzen nicht zwangsläufig in hohem Maße durch die soziale Herkunft vorhergesagt werden und eine geringe Kopplung bei gleichzeitig hohen Kompetenzen möglich ist.

Der Quadrant rechts unten repräsentiert Staaten, deren durchschnittliches Kompetenzniveau geringer ausfällt als der OECD-Durchschnitt und bei denen die Varianzaufklärung durch den ESCS oberhalb des OECD-Durchschnitts liegt. Staaten in diesem

Quadranten sind zum Beispiel Luxemburg, Chile, Ungarn und die Slowakische Republik. Die insgesamt unterdurchschnittlichen mathematischen Kompetenzen werden in diesen Staaten überdurchschnittlich stark durch Unterschiede in den Merkmalen der sozialen Herkunft vorhergesagt.

Der Datenpunkt für Deutschland ist im Quadranten rechts oben zu finden. Die Kompetenzmittelwerte der Staaten in diesem Quadranten sind höher als der OECD-Durchschnitt. Gleichzeitig wird in diesen Staaten das erreichte Kompetenzniveau in relativ hohem Maß durch Unterschiede in der sozialen, kulturellen und ökonomischen Herkunft vorhergesagt. In Deutschland unterscheiden sich die durch den ESCS-Index erklärten Unterschiede in der mathematischen Kompetenz nicht signifikant vom OECD-Durchschnitt.

Darüber hinaus ist in Abbildung 8.2 zu erkennen: Es gibt keine Anhaltspunkte dafür, dass ein hohes Kompetenzniveau nur durch starke soziale Unterschiede erreicht werden kann. Vielmehr wird aus der Abbildung ersichtlich, dass es eine Reihe von Staaten gibt, in denen der ESCS einen niedrigen Vorhersagewert für die Kompetenz hat (wie etwa in Korea, Japan und Finnland) und Jugendliche dennoch ein hohes Kompetenzniveau erreichen. Dieser Befund weist darauf hin, dass auch in Deutschland eine weitere Reduzierung sozialer Disparitäten bei gleichzeitiger Erhaltung oder gar Steigerung des Kompetenzniveaus prinzipiell möglich ist.



Anmerkung: Hell gefüllte Datenpunkte unterscheiden sich bezüglich der Varianzaufklärung durch den ESCS signifikant vom OECD-Durchschnitt.

Abbildung 8.2: Mathematikkompetenz und Varianzaufklärung (R^2) durch den ökonomischen, sozialen und kulturellen Status (ESCS)

Veränderungen in den sozialen Disparitäten der Mathematikkompetenz über die Zeit

Inwieweit haben sich in den OECD-Staaten die sozialen Disparitäten in den erreichten Kompetenzen über die Zeit verändert? Um diese Frage zu klären, werden im Folgenden die Kennwerte der sozialen Gradienten aus PISA 2003 und 2012 verglichen. Die Skala für die mathematische Kompetenz bietet sich als abhängige Variable an, da Mathematik in diesen beiden Erhebungen Hauptdomäne war und die Werte direkt verglichen werden können. Als Indikator für die soziale Herkunft wird zunächst der höchste sozioökonomische Status (HISEI) verwendet. Die Kennwerte des sozioökonomischen Status der Eltern basieren ab 2012 jedoch auf einer aktualisierten und differenzierteren Berufskodierung. Wie zuvor dargestellt (Abbildung 8.1) fallen Mittelwerte und insbesondere die Standardabweichungen des HISEI über alle Staaten hinweg etwas höher aus. Mögliche Unterschiede in den Kennwerten sind daher nicht nur auf Veränderungen in den Kompetenzen zurückzuführen, sondern auch auf die veränderte Berufskodierung.

Tabelle 8.3 stellt die sozialen Gradienten der mathematischen Kompetenz beider Erhebungen in den OECD-Staaten gegenüber. In der Tabelle sind für 2003 und 2012 die Mittelwerte und die Standardabweichungen für den HISEI sowie die Steigung des sozialen Gradienten (b) und der Anteil der durch den sozioökonomischen Status aufgeklärten Varianz (R^2) an der Mathematikkompetenz angegeben. Betrachtet man die Gesamtentwicklung in den OECD-Staaten über die beiden Erhebungszyklen, so ist die Kopplung zwischen sozioökonomischem Status und der mathematischen Kompetenz im Durchschnitt über die OECD-Staaten relativ stabil geblieben. Dieser Befund gilt sowohl für die Steigung der Gradienten (b : 29 vs. 32 Punkte) als auch für die Varianzaufklärung (R^2 : 11 vs. 12 Prozent). Entsprechend bleibt in der Mehrzahl der Staaten die Höhe der sozialen Disparitäten hinsichtlich der mathematischen Kompetenz unverändert.

Eine statistisch bedeutsame Zunahme in den Steigungen der sozialen Gradienten (b) von PISA 2003 bis 2012 kann für Frankreich, Island, Luxemburg, Neuseeland, die Slowakische Republik, Spanien und die Tschechische Republik beobachtet werden. In Belgien und Mexiko hat die Steigungen der sozialen Gradienten im Vergleich zwischen PISA 2003 und 2012 statistisch bedeutsam abgenommen. Hinsichtlich der Varianzaufklärung durch den HISEI findet sich in Spanien, Island, Luxemburg, Neuseeland und der Slowakischen Republik eine signifikante Zunahme zwischen den beiden Zyklen.

Für Deutschland bleiben die Kennwerte für die sozialen Disparitäten in der mathematischen Kompetenz über die Zeit weitgehend unverändert. Der Zusammenhang zwischen dem sozioökonomischen Status des Elternhauses und der von den Jugendlichen erreichten mathematischen Kompetenz ist im internationalen Vergleich immer noch hoch ausgeprägt. Die Herausforderung, eine geringe Kopplung bei hohem Kompetenzniveau zu erreichen, bleibt damit weiterhin bestehen.

Tabelle 8.3: Veränderungen in den sozialen Disparitäten des Kompetenzerwerbs über die Zeit im internationalen Vergleich des HISEI-Index und seines Zusammenhangs mit der Mathematikkompetenz zwischen PISA 2003 und 2012

OECD-Staaten	PISA 2003								PISA 2012							
	M	(SE)	SD	(SE)	b	(SE)	R ²	(SE)	M	(SE)	SD	(SE)	b	(SE)	R ²	(SE)
Australien	52.6	0.3	15.9	0.1	29	1.3	9.6	0.7	57.4	0.3	20.9	0.1	28	1.1	8.1	0.6
Belgien	50.6	0.4	16.7	0.1	41	1.7	15.3	1.2	51.6	0.4	21.8	0.1	35	1.8	12.5	1.1
Chile	m	m	m	m	m	m	m	m	40.8	0.6	21.4	0.3	34	2.1	17.7	1.9
Dänemark	49.3	0.5	15.6	0.2	27	1.6	9.1	1.0	54.7	0.6	21.6	0.2	28	1.6	11.7	1.3
Deutschland	49.3	0.4	16.2	0.2	38	1.9	15.5	1.4	50.9	0.5	20.6	0.2	40	2.0	15.6	1.4
Estland	m	m	m	m	m	m	m	m	50.7	0.3	21.1	0.2	24	1.5	8.2	0.9
Finnland	50.2	0.4	16.9	0.1	22	1.3	7.2	0.8	55.6	0.4	20.4	0.1	24	1.4	7.0	0.8
Frankreich	48.7	0.5	16.8	0.2	32	2.0	13.0	1.4	51.8	0.4	21.4	0.2	38	1.8	15.9	1.2
Griechenland	46.9	0.7	17.0	0.3	30	2.2	10.5	1.5	48.5	0.7	22.9	0.2	30	1.7	13.1	1.4
Irland	48.3	0.5	15.9	0.2	26	1.8	10.0	1.3	52.5	0.4	21.0	0.1	28	1.6	10.5	1.1
Island	53.7	0.3	16.7	0.1	15	1.5	2.7	0.6	58.5	0.3	19.9	0.2	26	1.9	7.2	1.0
Israel	m	m	m	m	m	m	m	m	58.6	0.7	20.9	0.2	42	3.0	16.2	2.0
Italien	46.8	0.4	16.7	0.2	28	1.9	8.3	1.0	46.7	0.3	20.9	0.1	29	1.1	9.3	0.6
Japan	50.0	0.3	14.7	0.2	21	2.8	4.4	1.0	50.0	0.3	20.2	0.1	22	2.4	4.9	1.0
Kanada	52.6	0.3	15.9	0.1	24	1.1	7.5	0.7	55.4	0.4	20.9	0.1	25	1.1	7.8	0.6
Korea	46.3	0.4	13.4	0.2	22	2.7	5.5	1.3	53.4	0.5	18.1	0.2	26	2.7	5.0	1.0
Luxemburg	48.2	0.2	16.6	0.1	34	1.6	13.8	1.1	49.1	0.3	22.0	0.1	41	1.3	19.7	1.1
Mexiko	40.1	0.7	18.4	0.3	26	2.1	9.5	1.4	39.7	0.4	21.9	0.2	20	0.9	7.5	0.6
Neuseeland	51.5	0.4	16.4	0.2	29	1.6	9.1	1.0	55.0	0.4	20.8	0.2	38	1.6	14.3	1.2
Niederlande	51.3	0.4	16.0	0.2	31	2.0	12.6	1.3	56.3	0.5	20.1	0.2	31	2.2	10.4	1.3
Norwegen	54.6	0.4	15.4	0.1	27	1.5	8.9	0.9	58.9	0.4	19.3	0.2	25	1.9	6.4	0.9
Österreich	47.1	0.5	16.1	0.2	30	1.9	10.6	1.2	48.9	0.5	20.9	0.2	33	2.0	12.0	1.3
Polen	45.0	0.3	14.9	0.2	32	1.6	12.6	1.2	45.1	0.6	21.7	0.2	33	2.2	13.5	1.5
Portugal	43.1	0.5	16.1	0.3	34	1.7	14.8	1.5	42.6	0.9	21.4	0.4	38	1.9	16.8	1.7
Schweden	50.6	0.4	16.2	0.1	28	1.8	9.2	1.0	54.6	0.4	20.7	0.1	29	1.7	9.6	1.0
Schweiz	49.3	0.4	16.1	0.2	30	1.7	9.4	0.0	55.0	0.5	21.1	0.1	30	1.7	10.0	1.1
Slowakische Republik	48.8	0.4	16.3	0.2	33	1.8	13.1	1.2	43.0	0.5	20.0	0.2	45	2.9	17.9	1.9
Slowenien	m	m	m	m	m	m	m	m	51.2	0.3	20.8	0.2	36	1.4	14.4	1.1
Spanien	44.3	0.6	16.3	0.2	25	1.4	8.2	0.9	46.9	0.5	21.5	0.1	30	1.1	12.2	0.8
Tschechische Republik	50.1	0.3	14.4	0.2	33	1.7	12.6	1.2	47.2	0.4	19.3	0.1	39	2.2	13.9	1.3
Türkei	41.6	0.7	15.6	0.4	36	5.6	11.8	3.0	35.0	0.6	19.9	0.4	28	2.6	8.3	1.4
Ungarn	48.6	0.3	15.3	0.2	38	2.0	16.9	1.5	46.1	0.6	20.6	0.2	41	2.8	18.3	2.0
Vereinigte Staaten	54.5	0.4	16.3	0.2	30	1.4	10.3	0.9	54.3	0.7	21.1	0.2	30	1.6	11.1	1.1
Vereinigtes Königreich	49.6	0.4	16.6	0.2	m	m	m	m	55.6	0.5	20.6	0.2	31	1.7	10.3	1.0
OECD-Durchschnitt	49.2	0.1	16.8	0.1	29	0.4	10.5	0.2	50.6	0.1	20.8	0.0	32	0.3	11.7	0.2

Anmerkung: Indikator für die soziale Herkunft ist der höchste sozioökonomische Status (HISEI) in der Familie (M: Mittelwert; SD: Standardabweichung; b: sozialer Gradient; R²: Varianzaufklärung; m: fehlende Daten). Signifikante Unterschiede ($p < .05$) in den Kennwerten der Regression zwischen PISA 2003 und PISA 2012 sind fettgedruckt.

Zieht man für den Vergleich der sozialen Gradienten zwischen den Erhebungszeitpunkten 2003 und 2012 den ESCS-Index heran (vgl. Tabelle 8.4), zeigen sich bei der Steigung des sozialen Gradienten (b) in Belgien, Mexiko, Norwegen, der Türkei und in den Vereinigten Staaten signifikante Rückgänge. Erhöht hat sich hingegen die Steigung des sozialen Gradienten für Finnland, Frankreich, Neuseeland, Spanien und Portugal.

Anhand des ESCS lässt sich in einer Reihe von Staaten in PISA 2012 statistisch bedeutsam weniger Varianz aufklären als in PISA 2003. Dies zeigt eine Abnahme in der Kopplung zwischen ESCS und Mathematikleistung an. Zu diesen Staaten zählt neben Belgien, Mexiko, den Niederlanden, Norwegen, Schweden, der Schweiz, der Türkei und den Vereinigten Staaten auch Deutschland. Der Anteil der Varianz an der mathematischen Kompetenz, der durch soziökonomische und soziokulturelle Herkunftsmerkmale aufgeklärt wird, hat sich in diesen Staaten also reduziert. Diese Befunde zeigen insgesamt, dass im Verlauf von neun Jahren signifikante Änderungen in der Kopplung zwischen sozialer Herkunft und mathematischer Kompetenz innerhalb von Schulsystemen auftreten, die strukturell nicht verändert wurden. Offensichtlich ist die Struktur eines Schulsystems nur ein Faktor neben anderen, die für die Kopplung von Herkunft und Kompetenz relevant sind.

Was bedeuten die unterschiedlichen deskriptiven Maße und Regressionskoeffizienten für HISEI und ESCS zwischen 2003 und 2012? Zum einen ist davon auszugehen, dass die stärker differenzierende ISCO-Kodierung die Streuung des HISEI sowohl für Deutschland (Standardabweichungen 2003: 16.2 Punkte; 2012: 20.6 Punkte) als auch für den OECD-Durchschnitt (2003: 16.8 Punkte; 2012: 20.8 Punkte) deutlich vergrößert. Dieser Einfluss kommt bei der Analyse des Zusammenhangs von Mathematikkompetenz und sozioökonomischem Status stärker zum Tragen als bei den Analysen, die mit Hilfe des ESCS durchgeführt wurden. Der ESCS basiert neben dem HISEI auf weiteren Indikatoren, deren Konstruktion zwischen den PISA-Erhebungen weitgehend stabil geblieben ist (Standardabweichung ESCS im OECD-Durchschnitt: 2003: 1.0 Punkte; 2012: 0.9 Punkte). Es ist nicht auszuschließen, dass mit dem größeren Set an Indikatoren die Kopplung zwischen Herkunft und Kompetenz stabiler abgebildet wird. Generell sind, wie mehrfach angemerkt, Vergleiche zwischen den Erhebungszyklen mit Vorsicht zu interpretieren.

Vorbehaltlich dieser Einschränkung kann angenommen werden, dass in Deutschland der sozioökonomische Status in PISA 2012 einen ungefähr gleich großen Varianzanteil an der Mathematikkompetenz aufklärt wie in PISA 2003. Werden aber mit einem breiten Indikator wie dem ESCS gleichzeitig ökonomische, soziale und kulturelle Merkmale in die Analyse einbezogen, zeigt sich für Deutschland eine gewisse Entkopplung bei den sozialen Disparitäten des Kompetenzerwerbs seit PISA 2003. Die Befundlage bei PISA 2012 ist somit keineswegs mehr so alarmierend wie in den ersten Runden von PISA.

Tabelle 8.4: Veränderungen in den sozialen Disparitäten des Kompetenzerwerbs über die Zeit im internationalen Vergleich des ESCS-Index und seines Zusammenhangs mit der Mathematikkompetenz zwischen PISA 2003 und 2012

OECD-Staaten	PISA 2003								PISA 2012							
	M	(SE)	SD	(SE)	b	(SE)	R ²	(SE)	M	(SE)	SD	(SE)	b	(SE)	R ²	(SE)
Australien	0.0	0.0	0.9	0.0	40	1.8	14.0	1.1	0.2	0.0	0.8	0.0	42	1.3	12.3	0.8
Belgien	0.0	0.0	1.0	0.0	51	1.9	23.0	1.4	0.1	0.0	0.9	0.0	43	1.9	15.0	1.3
Chile	m	m	m	m	m	m	m	m	-0.6	0.0	1.1	0.0	34	1.6	23.1	1.9
Dänemark	0.1	0.0	1.0	0.0	39	1.8	17.3	1.5	0.4	0.0	0.8	0.0	39	1.7	16.5	1.4
Deutschland	0.0	0.0	1.1	0.0	44	1.6	23.8	1.4	0.2	0.0	0.9	0.0	43	2.0	16.9	1.4
Estland	m	m	m	m	m	m	m	m	0.1	0.0	0.8	0.0	29	1.7	8.6	0.9
Finnland	0.1	0.0	1.0	0.0	28	1.4	10.5	1.0	0.4	0.0	0.8	0.0	33	1.8	9.4	0.9
Frankreich	-0.3	0.0	1.0	0.0	43	2.2	20.2	1.8	0.0	0.0	0.8	0.0	57	2.2	22.5	1.3
Griechenland	-0.3	0.0	1.0	0.0	36	2.0	16.0	1.9	-0.1	0.0	1.0	0.0	34	1.8	15.5	1.5
Irland	-0.3	0.0	0.9	0.0	36	1.7	15.7	1.4	0.1	0.0	0.8	0.0	38	1.8	14.6	1.2
Island	0.6	0.0	0.9	0.0	26	1.6	7.1	0.9	0.8	0.0	0.8	0.0	31	2.1	7.7	1.0
Israel	m	m	m	m	m	m	m	m	0.2	0.0	0.9	0.0	51	2.6	17.2	1.5
Italien	-0.3	0.0	1.1	0.0	31	1.8	12.3	1.3	-0.1	0.0	1.0	0.0	30	1.2	10.1	0.6
Japan	-0.4	0.0	0.8	0.0	43	4.5	11.8	2.1	-0.1	0.0	0.7	0.0	41	3.9	9.8	1.6
Kanada	0.2	0.0	0.9	0.0	30	1.3	10.2	0.8	0.4	0.0	0.9	0.0	31	1.2	9.4	0.7
Korea	-0.4	0.0	0.9	0.0	38	2.8	14.5	2.0	0.0	0.0	0.7	0.0	42	3.3	10.1	1.4
Luxemburg	-0.1	0.0	1.1	0.0	35	1.2	16.6	1.0	0.1	0.0	1.1	0.0	37	1.2	18.3	1.1
Mexiko	-1.3	0.1	1.2	0.0	30	1.9	17.2	2.0	-1.1	0.0	1.3	0.0	19	0.8	10.4	0.8
Neuseeland	-0.1	0.0	0.9	0.0	44	1.7	16.6	1.2	0.0	0.0	0.8	0.0	52	1.9	18.4	1.3
Niederlande	-0.1	0.0	1.0	0.0	39	2.2	17.3	1.5	0.2	0.0	0.8	0.0	40	3.1	11.5	1.7
Norwegen	0.2	0.0	0.8	0.0	41	1.8	12.1	1.1	0.5	0.0	0.8	0.0	32	2.4	7.4	1.0
Österreich	-0.3	0.0	0.9	0.0	42	2.2	15.1	1.5	0.1	0.0	0.8	0.0	43	2.2	15.8	1.5
Polen	-0.4	0.0	0.9	0.0	40	1.6	16.5	1.2	-0.2	0.0	0.9	0.0	41	2.4	16.6	1.7
Portugal	-0.9	0.0	1.3	0.0	28	1.2	18.5	1.6	-0.5	0.1	1.2	0.0	35	1.6	19.6	1.8
Schweden	0.1	0.0	1.0	0.0	36	1.9	14.3	1.3	0.3	0.0	0.8	0.0	36	1.9	10.6	1.1
Schweiz	-0.2	0.0	1.0	0.0	41	1.8	18.0	1.3	0.2	0.0	0.9	0.0	38	1.8	12.8	1.2
Slowakische Republik	-0.2	0.0	0.9	0.0	48	2.5	23.6	2.0	-0.2	0.0	0.9	0.0	54	2.9	24.6	2.1
Slowenien	m	m	m	m	m	m	m	m	0.1	0.0	0.9	0.0	42	1.5	15.6	1.0
Spanien	-0.5	0.0	1.1	0.0	27	1.4	12.6	1.2	-0.2	0.0	1.0	0.0	34	1.1	15.8	1.0
Tschechische Republik	-0.1	0.0	0.9	0.0	46	2.0	18.5	1.3	-0.1	0.0	0.7	0.0	51	2.7	16.2	1.5
Türkei	-1.2	0.1	1.1	0.0	50	5.0	24.9	3.9	-1.5	0.0	1.1	0.0	32	2.4	14.5	1.8
Ungarn	-0.3	0.0	1.0	0.0	50	2.1	25.7	1.8	-0.3	0.0	1.0	0.0	47	2.8	23.1	2.3
Vereinigte Staaten	0.1	0.0	1.0	0.0	42	1.4	19.0	1.1	0.2	0.0	1.0	0.0	35	1.7	14.8	1.3
Vereinigtes Königreich	m	m	m	m	m	m	m	m	0.3	0.0	0.8	0.0	41	2.4	12.5	1.2
OECD-Durchschnitt	-0.2	0.0	1.0	0.0	39	0.4	16.8	0.3	0.0	0.0	0.9	0.0	39	0.4	14.6	0.2

Anmerkung: Indikator für die soziale Herkunft ist der ökonomische, kulturelle und soziale Index (ESCS) z-standardisiert am OECD-Durchschnitt (*M*: Mittelwert; *SD*: Standardabweichung; *b*: sozialer Gradient; *R*²: Varianzaufklärung; *m*: fehlende Daten). Signifikante Unterschiede ($p < 0.05$) in den Kennwerten für die Regression zwischen PISA 2000 und PISA 2012 sind fettgedruckt.

8.4 Die soziale Herkunft der Jugendlichen in Deutschland: vertiefende Analysen

Ziel der vorangegangenen Auswertungen war es, den Zusammenhang zwischen der sozialen Herkunft und den erreichten Kompetenzen der Jugendlichen im Vergleich zu anderen OECD-Staaten zu beschreiben. Dabei kamen mit dem HISEI und dem ESCS zwei Indikatoren zum Einsatz, die unterschiedliche Aspekte der sozialen Herkunft berücksichtigen. Der folgende Teil konzentriert sich auf die Schülerinnen und Schüler in Deutschland und untersucht diese und weitere Aspekte der sozialen Herkunft genauer. Dabei wird der Frage nachgegangen, welche bildungsrelevanten Merkmale des Elternhauses für die Entwicklung von Kompetenzen fünfzehnjähriger Schülerinnen und Schüler in Deutschland bedeutsam sind und wie sich Jugendliche aus unterschiedlichen sozialen Lagen hinsichtlich dieser bildungsrelevanten Merkmale unterscheiden. In einem Vergleich zwischen PISA 2000 und 2012 wird zudem untersucht, ob sich die sozialen Disparitäten in den Kompetenzen und in der Bildungsbeteiligung über die Zeit verringert oder erhöht haben.

Vorhersage der Mathematikkompetenz durch Merkmale der sozialen Lage der Jugendlichen

Im Folgenden wird anhand von Regressionsanalysen untersucht, wie gut sich die mathematischen Kompetenzen von Jugendlichen in Deutschland durch verschiedene bildungsrelevante Merkmale der sozialen Herkunft vorhersagen lassen. Um statistisch abschätzen zu können, wie groß der Beitrag der einzelnen Herkunftsmerkmale zur Vorhersage von Kompetenzunterschieden ist, wurden drei aufeinander aufbauende Modelle analysiert. Tabelle 8.5 zeigt die Ergebnisse der drei Modellrechnungen, in die schrittweise der sozioökonomische Status (Modell I), das Bildungsniveau und die Besitztümer der Eltern (Modell II) sowie der Kindergartenbesuch, das Einschulungsalter und Informationen zur Erwerbstätigkeit der Eltern (Modell III) einbezogen wurden. Die relative Bedeutung der einzelnen Merkmale zur Vorhersage der Mathematikkompetenz kann so geschätzt werden.

In Modell I wird die Mathematikkompetenz lediglich durch den sozioökonomischen Status (HISEI) der Schülerinnen und Schüler vorhergesagt. Dies entspricht dem Vorgehen wie im internationalen Vergleich der sozialen Gradienten in der mathematischen Kompetenz (vgl. Tabelle 8.1). Die Kennwerte für Deutschland fallen also in vergleichbarer Höhe aus: Die Steigung des Gradienten beträgt 37 Punkte in der Mathematikkompetenz und 15 Prozent der Unterschiede in der Mathematikkompetenz lassen sich allein durch den sozioökonomischen Status der Schülerinnen und Schüler erklären.

Das Modell II berücksichtigt neben dem sozioökonomischen Status auch das Bildungsniveau der Eltern und einen Index, der die kulturellen und lernrelevanten Besitztümer in einer Familie abbildet. Die Kennwerte aus Modell II zeigen: Alle drei Merk-

Tabelle 8.5: Regressionsmodelle zur Schätzung von Disparitäten in der Mathematikkompetenz: Vorhersage durch den sozioökonomischen und soziokulturellen Hintergrund sowie weitere Merkmale der sozialen Lage

	Modell I		Modell II		Modell III	
	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)
Mathematikkompetenz (Achsenabschnitt)	522	(2.7)	522	(2.6)	494	(21.8)
Sozioökonomischer Status ¹	37	(2.1)	28	(2.4)	25	(2.3)
Bildungsniveau der Eltern ¹			10	(1.9)	10	(1.9)
Häusliche Besitztümer der Eltern ¹			14	(2.4)	11	(2.3)
Kindergartenbesuch für 1 Jahr ²					-5	(7.8)
Kindergartenbesuch für länger als 1 Jahr ²					39	(7.1)
Einschulung im Alter von 5 Jahren ³					3	(20.2)
Einschulung im Alter von 6 Jahren oder älter ³					-30	(19.8)
Mind. ein Elternteil in Vollzeitbeschäftigung ⁴					25	(5.4)
<i>N</i>	3995		3995		3995	
<i>R</i> ²	0.15		0.18		0.22	

Anmerkung: Die hier angegebene Regressionskonstante ist der geschätzte Mittelwert der Mathematikkompetenz mit mittlerer Ausprägung des HISEI, des Bildungsniveaus der Eltern sowie der Besitztümer der Eltern, falls die Variablen im Modell enthalten sind.

fett: signifikante Partialregressionskoeffizienten ($p < .05$)

¹ z-standardisiert

² Referenzgruppe: Kein Kindergartenbesuch

³ Referenzgruppe: Einschulung im Alter von vor 5 Jahren

⁴ Referenzgruppe: Kein Elternteil in Vollzeitbeschäftigung

male haben jeweils einen spezifischen Vorhersagebeitrag auf die Mathematikkompetenz. Im Vergleich von Modell I und II verringert sich zwar der Koeffizient für den sozioökonomischen Status. Dem sozioökonomischen Status kommt aber auch unter Kontrolle des elterlichen Bildungsabschlusses und der häuslichen Besitztümer ein spezifischer Vorhersagebeitrag für die Mathematikkompetenz zu. Alle drei in Modell II einbezogenen Indikatoren gehen auch in den ESCS-Index ein. An der Modellrechnung II wird daher auch deutlich, dass der ESCS das Konstrukt soziale Herkunft breiter operationalisiert und entsprechend eine etwas höhere Vorhersagekraft hat.

Im dritten Modell werden zusätzlich der Erwerbstätigkeitsstatus der Eltern sowie der Kindergartenbesuch und das Einschulungsalter der Jugendlichen berücksichtigt. Die Höhe der Regressionskoeffizienten für den sozioökonomischen Status, für das elterliche Bildungsniveau und die häuslichen Besitztümer verringern sich im Vergleich zu Modell II nur geringfügig. Die spezifische Bedeutung dieser Familienmerkmale für den Kompetenzerwerb der Schülerinnen und Schüler bleibt auch dann bestehen, wenn zusätzlich Merkmale des Bildungsweges der Kinder kontrolliert werden. Zudem kommt besonders dem Kindergartenbesuch ein spezifischer Vorhersagebeitrag zu: Jugendliche, die länger als ein Jahr einen Kindergarten besucht haben, erreichen unter Kontrolle aller übrigen Merkmale eine um 39 Punkte höhere Mathematikkompetenz als Kinder, die kei-

nen Kindergarten besucht haben. Bedeutsam ist auch, ob in einer Familie mindestens ein Elternteil in Vollzeit erwerbstätig ist. Jugendliche aus diesen Familien erreichen – bei Kontrolle der anderen Herkunftsmerkmale – eine um 25 Punkte höhere Mathematikkompetenz als Kinder aus Familien, in denen kein Elternteil in Vollzeit arbeitet. Für den Einschulungszeitpunkt der Schülerinnen und Schüler lässt sich hingegen kein spezifischer Vorhersageeffekt feststellen. Insgesamt klärt das Modell III 22 Prozent der Varianz in der Mathematikkompetenz auf und zeigt damit, dass die betrachteten Merkmale der sozialen Herkunft für die Vorhersage der mathematischen Kompetenzen bedeutsam sind.

Die soziale Lage der Jugendlichen aus unterschiedlichen EGP-Klassen

Um Jugendliche unterschiedlicher sozialer Herkunft differenzierter hinsichtlich familiärer Bildungsressourcen und ihrer Bildungsbeteiligung betrachten zu können, wird in den folgenden Analysen ein Index für die soziale Herkunft verwendet, der aufgrund von zusätzlich erforderlichen Fragebogenangaben nur für Deutschland gebildet werden kann. Es handelt sich um die bereits erwähnte, von Erikson und Kollegen (1979, 2002) vorgenommene Zuordnung der Berufe der Eltern zu sozialen Klassen. Diese sogenannte EGP-Klassifikation berücksichtigt neben der Einteilung der Berufe Angaben zur Art des Beschäftigungsverhältnisses sowie zum Ausmaß der Weisungsbefugnis und führt zu einer anschaulichen und soziologisch fundierten Typologie. Die Abstufung der Berufe und die typologische Klassifikation beruhen auf der Annahme, dass die jeweiligen Lebensbedingungen in Abhängigkeit von Einkommen, Bildung, gesellschaftlicher Anerkennung und beruflichen Gestaltungsoptionen variieren (Hradil, 2005). Tabelle 8.6 beschreibt beispielhaft die (zusammengefassten) EGP-Klassen, die im Folgenden zur Differenzierung unterschiedlicher soziokultureller Strukturgefüge verwendet werden.

Die nachfolgenden Analysen basieren auf Informationen aus den Fragebögen für die Schülerinnen und Schüler sowie aus den Elternfragebögen und beziehen sich auf die Angaben zur Bezugsperson³. Fehlende Werte wurden mit der Software Norm 2.03 (Schafer, 2000; Schafer & Graham, 2002) nach dem Multiple-Imputation-Ansatz geschätzt. Um Veränderungen über den gesamten Erhebungszeitraum analysieren zu können, wird in den folgenden Analysen die Lesekompetenz der Schülerinnen und Schüler zwischen PISA 2000 und 2012 untersucht.

3 Angabe für den Vater bzw. wenn diese fehlt, für die Mutter.

Tabelle 8.6: Beispielhafte Beschreibung der EGP-Klassifikation (Quelle: Baumert & Schümer, 2001, S. 339)

EGP-Klassen
<p>Obere Dienstklasse (I)</p> <p>Dazu zählen: Freie akademische Berufe, führende Angestellte, höhere Beamte, selbstständige Unternehmer mit mehr als 10 Mitarbeitern, Hochschul- und Gymnasiallehrer.</p>
<p>Untere Dienstklasse (II)</p> <p>Dazu zählen: Angehörige von Semiprofessionen, mittleres Management, Beamte im mittleren und gehobenen Dienst, technische Angestellte mit nicht manueller Tätigkeit.</p>
<p>Routinedienstleistungen Handel und Verwaltung (III)</p> <p>Dazu zählen: Büro- und Verwaltungsberufe mit Routinetätigkeiten, Berufe mit niedrig qualifizierten, nicht manuellen Tätigkeiten, die oftmals auch keine Berufsausbildung erfordern.</p>
<p>Selbstständige (IV)</p> <p>Dazu zählen: Selbstständige aus manuellen Berufen mit wenigen Mitarbeitern und ohne Mitarbeiter, Freiberufler sofern sie keinen hoch qualifizierten Beruf ausüben.</p>
<p>Facharbeiter und Arbeiter mit Leitungsfunktion (V, VI)</p> <p>Dazu zählen: Untere technische Berufe wie Vorarbeiter, Meister, Techniker, die in manuelle Arbeitsprozesse eingebunden sind; Aufsichtskräfte im manuellen Bereich.</p>
<p>Un- und angelernte Arbeiter, Landarbeiter (VII)</p> <p>Dazu zählen: alle un- und angelernten Berufe aus dem manuellen Bereich, Dienstleistungstätigkeiten mit manuellem Charakter und geringem Anforderungsniveau, Arbeiter in der Land-, Forst- und Fischwirtschaft.</p>

Die Tabelle 8.7 weist für die Bezugsperson der Jugendlichen in Deutschland die EGP-Klassenzugehörigkeit sowie Merkmale der sozialen Lage in PISA 2012 aus. Daran lässt sich erkennen, dass sich die EGP-Klassen in ihrer jeweiligen Ausstattung mit familiären Bildungsressourcen unterscheiden. Klassenspezifische Differenzen in ökonomischen, kulturellen und sozialen Ressourcen sind relevant, da sie für die Kompetenzentwicklung und Schulkarriere der Schülerinnen und Schüler wichtige Bedingungsfaktoren darstellen.

Neben dem sozioökonomischen Status, der vor allem in den Dienstklassen überdurchschnittlich hoch ausfällt, ist auch der Erwerbstätigkeitsstatus einer Familie ein wichtiger Indikator für ökonomische Ressourcen. In den beiden Dienstklassen arbeiten überdurchschnittliche viele Eltern in Vollbeschäftigung, in der EGP-Klasse der Routinedienstleistungen (III) und der EGP-Klasse der un- und angelernten Arbeiter (VII) sind hingegen weniger Eltern in Vollzeit erwerbstätig. Schülerinnen und Schüler, deren Eltern zu den beiden Dienstklassen zählen, können damit auf überdurchschnittlich hohe sozioökonomische Ressourcen zurückgreifen im Vergleich zu Jugendlichen, deren Eltern den anderen EGP-Klassen zugeordnet werden.

Mit den Indikatoren zum höchsten Bildungsabschluss der Eltern (umgerechnet in Ausbildungsjahre) sowie zu den häuslichen Besitztümern werden Angaben zusammengefasst, die Auskunft über die kulturellen Ressourcen im Elternhaus geben können (Bourdieu, 1983). Ein Vergleich der Kennwerte belegt hier ebenfalls bedeutsame Unterschiede zwischen den EGP-Klassen. Die Eltern in den beiden Dienstklassen verfügen

Tabelle 8.7: Merkmale der sozialen Lage und der Bildungswege differenziert nach EGP-Klassen

EGP-Klassen	Bezugs- person		Sozio- ökono- mischer Status		Bildungs- dauer in Jahren		Häusliche Besitz- tümer		Familien- typ Kern- familie (zwei Eltern- teile)		Mind. ein Elternteil in Voll- zeitbe- schäfti- gung		Kinder- garten- besuch (länger als 1 Jahr)		Einschu- lung im Alter von 5 Jahren	
	%	(SE)	M	(SE)	M	(SE)	M	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)
Obere Dienstklasse (I)	24.2	(0.7)	71	(0.4)	15.9	(0.1)	0.33	(0.0)	88.6	(1.0)	93.4	(0.7)	91.6	(0.9)	7.4	(0.8)
Untere Dienstklasse (II)	18.2	(0.6)	59	(0.6)	14.8	(0.1)	0.13	(0.0)	84.9	(1.5)	90.9	(1.1)	86.9	(1.4)	4.9	(0.6)
Routinedienst- leistungen Handel und Verwaltung (III)	5.4	(0.4)	45	(1.1)	13.1	(0.2)	-0.19	(0.1)	76.7	(3.0)	80.9	(2.4)	87.7	(2.2)	2.9	(1.1)
Selbstständige (IV)	11.2	(0.5)	42	(0.8)	13.6	(0.1)	0.01	(0.0)	83.0	(1.9)	89.9	(1.4)	81.1	(2.1)	4.8	(1.0)
Facharbeiter und Arbeiter mit Leitungsfunktion (V, VI)	20.9	(0.7)	41	(0.5)	13.1	(0.1)	-0.11	(0.0)	86.8	(1.2)	86.4	(1.1)	83.5	(1.2)	3.7	(0.7)
Un- und ange- lernte Arbeiter. Landarbeiter (VII)	20.1	(0.6)	37	(0.6)	12.6	(0.1)	-0.36	(0.0)	86.6	(1.3)	80.7	(1.5)	78.8	(1.6)	3.0	(0.7)
Gesamt	100		51	(0.4)	14.0	(0.1)	0.00	(0.0)	85.8	(0.6)	88.0	(0.6)	85.2	(0.7)	4.8	(0.4)

Anmerkung: Signifikante Unterschiede ($p < .05$) in den Kennwerten für die EGP-Klassen zum Gesamtmittelwert, bzw. -prozentsatz sind fett gedruckt.

über höhere Bildungsabschlüsse und mit rund 15 bzw. 16 Jahren über bedeutsam längere Ausbildungszeiten als die übrigen Klassen (etwa durchschnittlich 13 Jahre). Auch hinsichtlich der häuslichen Besitztümer werden klassenspezifische Unterschiede sichtbar. Die Effektstärke der Mittelwertsdifferenz zwischen der oberen Dienstklasse (EGP-Klasse I) und der Klasse der Facharbeiter (EGP-Klassen V, VI) liegt für die lernrelevanten Besitztümer bei $d = 0.69$. Die EGP-Klassen gehen somit auch mit differenziellen kulturellen und häuslichen Lern- und Anregungsmilieus einher.

Unterschiede in der sozialen Herkunft wirken sich auch auf die Bildungswege der Jugendlichen aus. So besuchten 92 Prozent der Kinder, deren Eltern in der oberen Dienstklasse tätig sind, einen Kindergarten länger als ein Jahr. Jugendliche aus Familien der EGP-Klasse der un- und angelernten Arbeiter nahmen die Möglichkeiten frühkindlicher Betreuung deutlich kürzere Zeit in Anspruch (79 Prozent). Auch der Anteil der Kinder, deren Schuleintritt im Alter von fünf Jahren stattfand, ist in der oberen Dienstklasse größer (7 Prozent) als z. B. in der Klasse der ungelerten Arbeiter (3 Prozent).

Zusammengenommen belegen die Analysen, dass mit den EGP-Klassen unterschiedliche Lebenslagen verbunden sind. So weisen die beiden Dienstklassen in allen untersuchten sozioökonomischen, kulturellen und auf den Bildungsweg bezogenen Merk-

malen überdurchschnittliche Kennwerte aus, die auf ein besonders lernförderliches familiäres Umfeld schließen lassen.

Veränderungen in den sozialen Disparitäten der Lesekompetenz und der Gymnasialbeteiligung über die Zeit

Mit der Erhebung in PISA 2012 besteht die Gelegenheit zu analysieren, ob und inwieweit sich die Kompetenzunterschiede und die Gymnasialbeteiligung von Schülerinnen und Schülern aus unterschiedlichen EGP-Klassen über einen Zeitraum von zwölf Jahren verringert oder erhöht haben. Im zeitlichen Vergleich haben sich die relativen Anteile der sechs Klassen für die Bezugsperson zwischen den beiden Messzeitpunkten nur wenig verschoben. Das heißt, dass sich die Personen über die Zeit nahezu identisch auf die EGP-Klassen verteilen. Verändert haben sich in PISA 2012 im Vergleich zu PISA 2000 lediglich die Anteile der Selbstständigen (-3.3 Prozent) und die der Routine-dienstleistungen (-2.2 Prozent) zugunsten der oberen Dienstklasse (+5.1 Prozent). Diese geringen Verschiebungen in den Anteilen sprechen dafür, dass die EGP-Klassifikation auch über die Zeit ausreichend stabil ist, um Vergleiche zwischen den Messzeitpunkten durchzuführen.

Tabelle 8.8 beschreibt – differenziert nach EGP-Klassen – die Mittelwerte und Standardabweichungen der Lesekompetenz in PISA 2000 und 2012 und gibt damit Auskunft über das erreichte Kompetenzniveau und die Variabilität der Ergebnisse zwischen den EGP-Klassen. In der oberen und unteren Dienstklasse ist die Lesekompetenz über die Zeit weitgehend unverändert geblieben. Ein deutlicher Zuwachs lässt sich hingegen bei den Schülerinnen und Schülern verzeichnen, deren Eltern den EGP-Klassen der Routinedienstleistungen (III), der Facharbeiter und Arbeiter mit Leitungsfunktion (V, VI) und der un- und angelernten Arbeiter (VII) zugeordnet werden können. Verglichen mit den Ergebnissen der ersten PISA-Erhebung im Jahr 2000 konnten die Leistungen bei Schülerinnen und Schülern aus diesen Elternhäusern um rund 40 Punkte gesteigert werden. Dieser Anstieg kann als statistisch bedeutsam abgesichert werden.

Im Vergleich zwischen PISA 2000 und 2012 hat sich demnach der Abstand in der Lesekompetenz von Schülerinnen und Schülern, deren Eltern zur oberen Dienstklasse (I) zählen, und Jugendlichen aus Haushalten un- und angelernter Arbeiter (VII) verringert. Er lag in PISA 2000 bei 106 Punkten und ist in PISA 2012 auf 71 Punkte gesunken. Dies ist als positiv zu bewerten. Im Vergleich zu den sozialen Disparitäten in der mathematischen Kompetenz fallen in Deutschland die sozialen Unterschiede in der Lesekompetenz über die Zeit geringer aus.

Aus einer kriterialen Perspektive ist bedeutsam, inwieweit es in Deutschland gelingt, ein ausreichend hohes Kompetenzniveau in allen sozialen Lagen zu erzielen. Von einem unzureichenden Kompetenzniveau spricht man in PISA, wenn die Lesekompetenz auf Stufe Ia (335–407 Punkte) oder darunter liegt. Die prozentualen Anteile dieser Gruppen von Jugendlichen sind in Tabelle 8.9 differenziert nach EGP-Klassen und im Vergleich zwischen PISA 2000 und 2012 dargestellt.

Tabelle 8.8: Mittelwerte und Standardabweichungen der Lesekompetenz differenziert nach EGP-Klassen (Bezugsperson) zwischen PISA 2000 und 2012 in Deutschland

EGP-Klassen	PISA 2000			PISA 2012			ΔM	d
	<i>M</i>	(<i>SE</i>)	<i>SD</i>	<i>M</i>	(<i>SE</i>)	<i>SD</i>		
Obere Dienstklasse (I)	538	(3.4)	94	543	(4.0)	84	5	0.06
Untere Dienstklasse (II)	531	(4.0)	93	527	(3.8)	87	-4	-0.04
Routinedienstleistungen Handel und Verwaltung (III)	470	(6.4)	109	507	(6.3)	88	37	0.37
Selbstständige (IV)	480	(5.2)	94	483	(4.7)	94	3	0.03
Facharbeiter und Arbeiter mit Leitungsfunktion (V, VI)	459	(4.4)	104	497	(4.1)	89	38	0.39
Un- und angelernte Arbeiter. Landarbeiter (VII)	432	(3.9)	111	472	(3.8)	86	40	0.41

Anmerkung: Signifikante Unterschiede ($p < .05$) sind fettgedruckt.

Tabelle 8.9: Prozentuale Anteile von Schülerinnen und Schülern, deren Lesekompetenz auf Kompetenzstufe Ia oder darunter liegt, differenziert nach EGP-Klassen (Bezugsperson) zwischen PISA 2000 und 2012 in Deutschland

EGP-Klassen	PISA 2000		PISA 2012		$\Delta \%$
	%	(<i>SE</i>)	%	(<i>SE</i>)	
Obere Dienstklasse (I)	9.0	(0.6)	7.1	(1.0)	-1.9
Untere Dienstklasse (II)	9.9	(0.6)	9.9	(1.2)	0.0
Routinedienstleistungen Handel und Verwaltung (III)	26.1	(1.3)	14.3	(2.7)	-11.8
Selbstständige (IV)	20.2	(1.0)	21.5	(2.3)	1.4
Facharbeiter und Arbeiter mit Leitungsfunktion (V)	29.0	(0.9)	16.1	(1.7)	-12.9
Un- und angelernte Arbeiter. Landarbeiter (VII)	38.6	(0.9)	22.0	(1.8)	-16.6

Anmerkung: Signifikante Unterschiede ($p < .05$) sind fettgedruckt.

Der Gesamtanteil der sehr leistungsschwachen Schülerinnen und Schüler im Lesen ist seit PISA 2000 kontinuierlich zurückgegangen. Während bei der ersten PISA-Erhebung im Jahr 2000 noch 22.6 Prozent dieser Gruppe angehörten, ist der relative Anteil in PISA 2012 mit 14.5 Prozent rund 8 Prozentpunkte niedriger (2003: 20.0 Prozent; 2006: 20.0 Prozent; 2009: 18.5 Prozent). Der Blick auf die relativen Anteile (Tabelle 8.9) leistungsschwacher Schülerinnen und Schülern in den einzelnen EGP-Klassen verdeutlicht, dass der Rückgang nicht in allen sozialen Lagen gleichermaßen stattgefunden hat. Besonders große und signifikante Verschiebungen zeigen sich in der EGP-Klasse der Routinedienstleistungen (III), der Facharbeiter und Arbeiter mit Leitungsfunktion (V) sowie der un- und angelernten Arbeiter und Landarbeiter (VII). In diesen EGP-Klassen verringerte sich der prozentuale Anteil der Jugendlichen, deren Lesekompetenz auf Stufe Ia oder darunter liegt um mehr als 10 Prozentpunkte. Mit einer Abnahme von knapp 17 Prozent ist in der EGP-Klasse VII der größte Rückgang zu verzeichnen.

Damit setzt sich ein Trend fort, der sich bereits seit PISA 2006 abzuzeichnen begann: Insbesondere bei den Jugendlichen aus den unteren EGP-Klassen hat sich die mittlere Lesekompetenz gesteigert und der Anteil von leseschwachen Jugendlichen reduziert. Dieser Befund ist als eine positive und wünschenswerte Entwicklung hin zur Verringerung sozialer Disparitäten der Lesekompetenz festzuhalten.

Zusammenhang zwischen der sozialen Herkunft und der Gymnasialbeteiligung im zeitlichen Vergleich

Inwieweit sich die sozialen Disparitäten in der Bildungsbeteiligung in Deutschland zwischen PISA 2000 und 2012 verändert haben, soll im Folgenden analysiert werden. In Tabelle 8.10 ist die Gymnasialbeteiligung der Schülerinnen und Schüler differenziert nach EGP-Klassen im zeitlichen Vergleich dargestellt. Die Gegenüberstellung beschränkt sich dabei auf die Gymnasien. Die anderen Schularten werden zusammengefasst, da sich die Zusammensetzungen der Schularten über die Zeit verändert haben und somit inhaltlich nicht mehr sinnvoll verglichen werden können.

Für die Gesamtstichprobe lässt sich insgesamt ein bedeutsamer Anstieg in der Gymnasialbeteiligung von PISA 2000 (28 Prozent) nach PISA 2012 (36 Prozent) feststellen. Dabei ist die Frage, inwieweit sich durch diese Expansion auch die soziale Zusammensetzung der Schülerinnen und Schüler an Gymnasien verändert hat. Die Gymnasialbe-

Tabelle 8.10: Prozentuale Anteile der Schülerinnen und Schüler in den Schularten, differenziert nach EGP-Klassen im Vergleich zwischen PISA 2000 und 2012 in Deutschland

EGP-Klassen	PISA 2000				PISA 2012			
	Gymnasium		Andere Schularten		Gymnasium		Andere Schularten	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)
Obere Dienstklasse (I)	52	(2.0)	48	(2.0)	58	(2.7)	42	(2.7)
Untere Dienstklasse (II)	45	(2.0)	55	(2.0)	43	(2.8)	57	(2.8)
Routinedienstleistungen Handel und Verwaltung (III)	24	(2.6)	76	(2.6)	32	(3.7)	68	(3.7)
Selbstständige (IV)	26	(1.7)	74	(1.7)	26	(2.7)	74	(2.7)
Facharbeiter und Arbeiter mit Leitungsfunktion (V, VI)	16	(1.2)	84	(1.2)	27	(2.5)	73	(2.5)
Un- und angelernte Arbeiter. Landarbeiter (VII)	11	(1.0)	89	(1.0)	19	(2.0)	81	(2.0)
Gesamt	28	(1.0)	72	(1.0)	36	(2.1)	64	(2.1)

Anmerkung: Signifikante Unterschiede ($p < .05$) zwischen PISA 2000 und 2012 sind fettgedruckt.

Die Kategorie „Andere Schularten“ umfasst Hauptschulen, Realschulen, Integrierte Gesamtschulen, Schulen mit mehreren Bildungsgängen, Sonder- und Förderschulen sowie berufsbildende Schulen.

teiligung der Jugendlichen aus Familien der oberen und der unteren Dienstklasse ist über die Zeit nahezu unverändert hoch geblieben (EGP-Klasse I: 52 vs. 58 Prozent, EGP-Klasse II: 45 vs. 43 Prozent). Im Gegensatz dazu erhöhte sich die Gymnasialbeteiligung statistisch bedeutsam für die Schülerinnen und Schüler aus der EGP-Klasse der Facharbeiter (V, VI) (von 16 auf 27 Prozent) und der un- und angelernten Arbeiter (VII) (von 11 auf 19 Prozent). Dieser Befund, der ähnlich bei PISA 2009 berichtet wurde (Ehmke & Jude, 2010), ist als positiv zu werten und bestätigt die Analysen von Müller und Haun (1994), dass von einer steigenden Expansion des Gymnasiums die Jugendlichen aus der Unter- und Mittelschicht stärker profitieren als die Jugendlichen aus der oberen Schicht (vgl. dazu Maaz, Baumert, Gresch & McElvany, 2010).

Zusammenfassend belegen die Analysen zur Bildungsbeteiligung einerseits weiterhin relativ große Unterschiede in der Gymnasialbeteiligung zwischen den EGP-Klassen. Andererseits konnte auch hier ein Rückgang der sozialen Disparitäten festgestellt werden, der sich bereits in den letzten PISA-Erhebungen angedeutet hat.

8.5 Zusammenfassung und Diskussion

Ziel dieses Kapitels war es, soziale Disparitäten in der Kompetenz und der Bildungsbeteiligung für die Fünfzehnjährigen in Deutschland zu beschreiben und im Vergleich zu den OECD-Staaten sowie im Vergleich mit den früheren PISA-Erhebungen einzuordnen. Anhand verschiedener Modelle zur Differenzierung sozialer Milieus wurden die Ungleichheiten in der Bildungsbeteiligung und der Entwicklung von Kompetenzen außerdem für Deutschland genauer betrachtet.

Zunächst zeigte sich, dass der anhand des HISEI gemessene sozioökonomische Status zwischen den Staaten beträchtlich variiert. Deutschland befindet sich wie bereits in den vorangegangenen PISA-Erhebungen im Bereich des OECD-Durchschnitts. Deutlich erhöht hat sich hingegen sowohl in Deutschland als auch in den übrigen Staaten die Heterogenität in der Sozialstruktur. Es ist davon auszugehen, dass die höheren Standardabweichungen auf das neue, stärker differenzierende Berufsklassifikationssystem des ISCO-08 zurückzuführen sind. Die Veränderung in der Berufsklassifikation wirkt sich auch auf die Analysen des Zusammenhangs zwischen sozioökonomischer Herkunft und Kompetenzerwerb aus, da sowohl der sozioökonomische Index (HISEI) als auch der Index of Economic, Social and Cultural Status (ESCS) auf den ISCO-Kodierungen der elterlichen Berufe basieren.

Bei der Konstruktion des ESCS werden jedoch noch zusätzliche Indikatoren berücksichtigt, die sich zwischen den PISA-Zyklen nur wenig verändert haben. Während der HISEI lediglich den sozioökonomischen Status des Elternhauses abbildet, werden beim ESCS zudem der elterliche Bildungsabschluss und der Besitz von Kultur- und Wohlstandsgütern und damit soziale und kulturelle Komponenten der sozialen Herkunft integriert. Da bis PISA 2009 das alte Klassifikationssystem (ISCO-88) verwendet wurde, sind insbesondere die Vergleiche zwischen den verschiedenen PISA-Erhebungen unter Vor-

behalt zu interpretieren. Der ESCS-Index scheint jedoch gegenüber den Veränderungen stabiler zu sein und könnte eine etwas verlässlichere Vergleichsbasis liefern.

In PISA 2012, wie auch schon in früheren PISA-Erhebungen, kann in allen OECD-Staaten ein deutlicher Zusammenhang zwischen der sozialen Herkunft der Jugendlichen und den erreichten Kompetenzen festgestellt werden. Durch die Verwendung unterschiedlicher Indikatoren der sozialen Herkunft lässt sich dieser Zusammenhang differenziert in den Blick nehmen. Betrachtet wurde die Kopplung zwischen sozialer Herkunft und Kompetenzniveau zum einen mit Hilfe des sozioökonomischen Status (HISEI) und zum anderen anhand des ökonomischen, sozialen und kulturellen Status (ESCS).

Anhand beider Indizes zeigte sich in einer Reihe von Staaten wie Finnland, Estland oder Kanada eine ausgesprochen geringe Kopplung zwischen Mathematikkompetenz und sozialer Herkunft bei gleichzeitig hohem Durchschnittsniveau in der mathematischen Kompetenz. Für Deutschland bleibt die Herausforderung bestehen, die sozialen Disparitäten weiter abzuschwächen, ohne dabei das hohe Kompetenzniveau bei den Jugendlichen zu verringern. Am Beispiel Finnlands zeigt sich aber auch, dass disparitätsausgleichende Strukturen sowie hohe Kompetenzwerte nicht notwendigerweise über die Zeit stabil sind und es kontinuierlicher Anstrengungen bedarf, um diese Situation aufrecht zu erhalten. Zwar ist die Steigung des sozialen Gradienten in Finnland nach wie vor sehr gering bei überdurchschnittlichen Kompetenzwerten, aber der soziale Gradient erhöhte sich zwischen 2003 und 2012 signifikant, während die durchschnittlichen Kompetenzwerte deutlich zurückgingen.

Für die Fünfzehnjährigen in Deutschland liegt sowohl die Steigung des sozialen Gradienten als auch die Varianzaufklärung in PISA 2012 gemessen am sozioökonomischen Status (HISEI) oberhalb des OECD-Durchschnitts. Wird hingegen statt des sozioökonomischen Status ein breiteres Indikatorensystem verwendet und der Zusammenhang am ESCS gemessen, sind in Deutschland in beiden Kennwerten keine signifikanten Unterschiede zum OECD-Durchschnitt feststellbar. Darüber hinaus zeigt sich bei diesem Indikator zwischen PISA 2003 und 2012 eine voranschreitende Entkopplung zwischen Mathematikkompetenz und sozialer Herkunft. Werden also nicht nur sozioökonomische, sondern gleichzeitig ökonomische, soziale und kulturelle Ressourcen im Elternhaus in die Analysen einbezogen, so nimmt der Zusammenhang von sozialer Herkunft und Mathematikkompetenz über die Zeit ab. Offenbar gelingt es in Deutschland inzwischen zumindest teilweise, die neben dem sozioökonomischen Status vorhandenen kulturellen und sozialen Ressourcen der sozialen Herkunft besser zu kompensieren.

Dies bestätigt sich auch in den vertieften Analysen für Deutschland, in denen festgestellt werden konnte, dass sich die Abstände in der Lesekompetenz zwischen den sozialen Lagen über die Zeit reduziert haben. Zwar unterscheidet sich auch in PISA 2012 die mittlere Lesekompetenz von Jugendlichen, deren Eltern der oberen Dienstklasse angehören, und Schülerinnen und Schülern aus Familien von un- und angelernten Arbeitern noch deutlich, die Differenz der Mittelwerte ist jedoch von 106 Punkten in PISA 2000 auf 71 Punkte in PISA 2012 gesunken. Profitiert haben vor allem die Schülerinnen und

Schüler, die aus Arbeiterfamilien stammen oder zur EGP-Klasse der Routinedienstleistungen zählen. Analog dazu ist der Anteil der Schülerinnen und Schüler, deren Lesekompetenz auf Stufe Ia oder darunter liegt, in diesen EGP-Klassen deutlich zurückgegangen. Die Lesekompetenz der Jugendlichen von Familien aus der oberen Dienstklasse ist hingegen unverändert auf hohem Niveau geblieben. Im Hinblick auf die sozialen Disparitäten der Bildungsbeteiligung belegen die Analysen einerseits weiterhin große Unterschiede zwischen den EGP-Klassen im Gymnasialbesuch. Andererseits wurde auch hier ein tendenzieller Rückgang der sozialen Disparitäten festgestellt. Im Vergleich zwischen PISA 2000 und 2012 erhöhte sich die Gymnasialbeteiligung vor allem für die Schülerinnen und Schüler, deren Eltern den EGP-Klassen der Facharbeiter (16 vs. 27 Prozent) und un- und angelernten Arbeiter (11 vs. 19 Prozent) zuzuordnen sind. Bei den Schülerinnen und Schülern mit Eltern der Dienstklassen blieb die Gymnasialbeteiligung weiterhin auf hohem Niveau bestehen.

Zusammenfassend lässt sich festhalten: Die Verbesserungen in der mathematischen Kompetenz bei den Jugendlichen in Deutschland sind keineswegs zu Lasten erhöhter sozialer Disparitäten erfolgt. Dies ist insofern eine erfreuliche Entwicklung, als dass Befunde aus dem nationalen Ländervergleich erkennen lassen, dass Verbesserungen im Kompetenzniveau eines Landes mit einer verstärkten Kopplung zwischen sozialer Herkunft und Kompetenz einhergehen können (Kuhl, Siegle & Lenski, 2013). Dagegen zeigen die vertiefenden Analysen in PISA 2012 für Deutschland, dass sich die sozialen Disparitäten in der Lesekompetenz über die Zeit bedeutsam abgeschwächt und die Unterschiede zwischen den EGP-Klassen zugunsten der Kinder, deren Eltern als Facharbeiter oder als un- und angelernte Arbeiter tätig sind, abgenommen haben; dieser Befund bestätigt die Aussagen aus PISA 2009 (Ehmke & Jude, 2010). Hinsichtlich der Lebensbedingungen gleiche Ausgangslagen zu schaffen, kann und soll nicht Ziel von Maßnahmen im Bildungsbereich sein. Vielmehr muss im Vordergrund stehen, das vorhandene Potential und die Kompetenzen von Kindern und Jugendlichen unabhängig von ihrer sozialen Herkunft bestmöglich zu fördern. Ein gerechtes Bildungssystem zeichnet sich dadurch aus, dass es den Zusammenhang zwischen sozialer Herkunft und Leistung zu entkoppeln versucht. In Deutschland zeigt sich, dass diese Entkopplung in Teilen stattgefunden hat. Zwar ist der Zusammenhang zwischen sozioökonomischem Status und erreichter Kompetenz nach wie vor hoch, aber er hat sich über die Zeit verringert. Die Zielsetzung besteht weiterhin darin, eine ähnlich hohe Entkopplung zwischen sozialer Herkunft und Bildungsergebnissen zu erreichen, wie sie in einzelnen OECD-Staaten schon seit einem Jahrzehnt durch konsequente Weiterentwicklung des Bildungssystems erwirkt wurde (vgl. Sahlberg, 2007).

Literatur

- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–410). Opladen: Leske + Budrich.
- Bourdieu, P. (1982). *Die feinen Unterschiede: Kritik der gesellschaftlichen Urteilskraft*. Frankfurt a.M.: Suhrkamp.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (S. 183–198). Göttingen: Schwartz.
- Caro, D. H. & Cortés, D. (2012). Measuring family socioeconomic status: An illustration using data from PIRLS 2006. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 5, 9–33.
- Ehmke, T. & Baumert, J. (2007). Soziale Herkunft und Kompetenzerwerb. Vergleiche zwischen PISA 2000, 2003 und 2006. 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. 309–335). Münster: Waxmann.
- Ehmke, T., Hohensee, F., Heidemeier, H. & Prenzel, M. (2004). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 225–254). Münster: Waxmann.
- Ehmke, T. & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 231–254). Münster: Waxmann.
- Ehmke, T. & Siegle, T. (2005). ISEI, ISCED, HOMEPOS, ESCS. *Zeitschrift für Erziehungswissenschaft*, 8 (4), 521–540.
- Erikson, R. & Goldthorpe, J. H. (2002). Intergenerational inequality: A sociological perspective. *Journal of Economic Perspectives*, 16 (3), 31–44.
- Erikson, R., Goldthorpe, J. H. & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology*, 30, 341–415.
- Ganzeboom, H. B. G. (2010, 1. Mai). *A new international socio-economic Index (ISEI) of occupational Status for the international standard classification of occupation 2008 (ISCO-08) constructed with data from the ISSP 2002-2007; With an analysis of quality of occupational measurement in ISSP*. Paper presented at Annual Conference of International Social Survey Programme, Lisbon, May 1 2010. Zugriff am 13.08.2013. Verfügbar unter <http://www.harryganzeboom.nl/isco08/index.htm>
- Ganzeboom, H. B. G., de Graaf, P. M. & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21 (1), 1–56.
- Ganzeboom, H. B. G. & Treiman, D. J. (2003). Three internationally standardised measures for comparative research on occupational status. In J. H. P. Hoffmeyer-Zlotnik & C. Wolf (Hrsg.), *Advances in cross-national comparison* (S. 159–193). Berlin: Springer.
- Ganzeboom, H. B. G. & Treiman, D. (2012). *International Stratification and Mobility File: Conversion Tools*. Amsterdam: Department of Social Research Methodology. Zugriff am 02.09.2013. Verfügbar unter <http://www.harryganzeboom.nl/ismf/index.htm>

- Hradil, S. (2005). *Soziale Ungleichheit in Deutschland* (8. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- International Labour Office. (1990). *International Standard Classification of Occupations ISCO-88*. Genf: ILO.
- International Labour Office. (2012). *International Standard Classification of Occupations ISCO-08*. Genf: ILO.
- Jencks, C. (1973). *Chancengleichheit*. Reinbek: Rowohlt.
- Knigge, M. & Leucht, M. (2010). Soziale Disparitäten im Spracherwerb. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 185–201). Münster: Waxmann.
- Kuhl, P., Siegle, T. & Lenski, A. E. (2013). Soziale Disparitäten. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 275–296). Münster: Waxmann.
- Maaz, K., Baumert, J., Gresch, C. & McElvany, N. (Hrsg.) (2010). *Der Übergang von der Grundschule in die weiterführende Schule: Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten*. Bonn: Bundesministerium für Bildung und Forschung, Referat Bildungsforschung.
- Müller, W. & Haun, D. (1994). Bildungsungleichheit im sozialen Wandel. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 46 (1), 1–42.
- OECD. (2001). *Knowledge and skills for life. First results from PISA 2000*. Paris: OECD.
- OECD. (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.
- OECD. (2007). *PISA 2006: Science competencies for tomorrow's world*. Paris: OECD.
- OECD. (2009). *PISA data analysis manual. SPSS® Second edition*. Paris: OECD.
- OECD. (2010). *PISA 2009. Was Schülerinnen und Schüler wissen und können. Schülerleistungen in Lesekompetenz, Mathematik und Naturwissenschaften*. Paris: OECD.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.) (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- 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.
- Sahlberg, P. (2007). Education policies for raising student's learning: the Finnish approach. *Journal of Education Policy*, 22 (2), 147–171.
- Schafer, J. L. (2000). *Norm for Windows 95/98/NT (Version 2.03)*. University Park, PA: Penn State Department of Statistics.
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7 (2), 147–177.
- VBW – Vereinigung der Bayerischen Wirtschaft e.V. (Hrsg.) (2007). *Bildungsgerechtigkeit. Aktionsrat Bildung. Jahresgutachten 2007*. Wiesbaden: VS Verlag für Sozialwissenschaften [Aktionsrat Bildung: H.-P. Blossfeld, W. Bos, D. Lenzen, D. Müller-Böling, J. Oelkers, M. Prenzel. & L. Wößmann].

9 Mathematische Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund

Markus Gebhardt, Dominique Rauch, Julia Mang,
Christine Sälzer und Petra Stanat

Das Zuwandern in ein anderes Land ist für Familien ein Lebensereignis, das für das Aufwachsen der Kinder beträchtliche Veränderungen bedeutet und neue Anforderungen an die Familien stellt. Diese Familien werden mit einem anderen Bildungswesen und einer neuen Sprachumgebung konfrontiert. Im neuen Lebensumfeld ergeben sich sowohl für Familien, deren Kinder noch im Herkunftsland geboren sind, als auch für Familien, deren Kinder bereits selbst im Einwanderungsland geboren sind, besondere Aufgaben. Besonders wichtig ist dabei das Erlernen der Sprache des Einwanderungslandes, da diese für den Bildungserfolg zentral ist (Ehmke, Hohensee, Heidemeier & Prenzel, 2004; Segeritz, Walter & Stanat, 2012; Stanat & Christensen, 2006). Dass Schülerinnen und Schüler mit Zuwanderungshintergrund in ihrer neuen Heimat ebenso erfolgreich unterstützt werden wie Schülerinnen und Schüler ohne Zuwanderungshintergrund, ist eine wichtige Herausforderung für Bildungssysteme (OECD, 2013). In Deutschland erzielten Jugendliche mit Zuwanderungshintergrund seit der ersten PISA-Erhebung bisher in allen drei Domänen weniger gute Leistungen als ihre Mitschülerinnen und Mitschüler ohne Zuwanderungshintergrund (Stanat & Christensen, 2006; Stanat, Rauch & Segeritz, 2010; Walter & Taskinen, 2008;). Somit steht Deutschland vor ähnlichen Herausforderungen wie viele andere OECD-Staaten.

Die großen Disparitäten zwischen Schülerinnen und Schülern mit und ohne Zuwanderungshintergrund waren ein Aspekt des PISA-Schocks, der in Deutschland auf die Veröffentlichung der Resultate aus der ersten PISA-Erhebung im Jahr 2001 folgte. Jugendliche, die selbst oder deren Eltern beide in einem anderen Land geboren waren, erzielten in allen Kompetenzbereichen wesentlich schwächere Testergebnisse als Jugendliche *ohne* Zuwanderungshintergrund (Baumert & Schümer, 2001). Auch waren Jugendliche aus Zuwandererfamilien in Gymnasien gegenüber ihren Mitschülerinnen und Mitschülern ohne Zuwanderungshintergrund unterrepräsentiert. Seither findet der Bildungserfolg von Jugendlichen aus Zuwandererfamilien im Rahmen von Vergleichsstudien besondere Aufmerksamkeit.

Seit PISA 2000 wurden in Deutschland verschiedene Fördermaßnahmen auf den Weg gebracht, die sich gezielt an Kinder und Jugendliche richten, deren Familien zugewandert sind (z. B. Becker & Beck, 2009; Stanat, Becker, Baumert, Lüdtke & Eckhardt, 2012). Als der erste PISA-Zyklus nach drei Erhebungsrounden abgeschlossen war und im

Rahmen von PISA 2009 Bilanz gezogen wurde, zeigte sich, dass die Lesekompetenz der Jugendlichen mit Migrationsgeschichte zwar noch immer hinter der ihrer Mitschülerinnen und Mitschüler ohne Zuwanderungshintergrund zurückblieb, sich jedoch im Vergleich zu PISA 2000 signifikant verbessert hat (Ehmke, Klieme & Stanat, 2013; Stanat et al., 2010). Da die Lesekompetenz der Fünfzehnjährigen *ohne* Zuwanderungshintergrund im gleichen Zeitraum nur wenig zugenommen hatte, war insgesamt eine Verringerung der herkunftsbedingten Disparitäten in Deutschland zu beobachten.

In PISA 2012 steht die mathematische Kompetenz zum zweiten Mal als Hauptdomäne im Mittelpunkt. Dieses Kapitel geht deshalb der Frage nach, ob sich für diesen Kompetenzbereich seit PISA 2003 Veränderungen der herkunftsbedingten Disparitäten nachweisen lassen. In PISA 2003 erreichten die Schülerinnen und Schüler in Deutschland insgesamt 503 Punkte auf der Skala mathematischer Kompetenz und lagen damit im Bereich des OECD-Durchschnitts (Köller, Knigge & Tesch, 2010). Die Leistungen der Schülerinnen und Schüler mit Zuwanderungshintergrund in Deutschland hingegen waren im Mittel um 64 Punkte geringer. Im Durchschnitt aller OECD-Staaten betrug der Abstand zwischen Jugendlichen aus zugewanderten Familien und Jugendlichen ohne Zuwanderungshintergrund 34 Punkte zu Ungunsten der Jugendlichen aus zugewanderten Familien. Die zuwanderungsbezogenen Disparitäten waren 2003 also in Deutschland deutlich höher als in vielen anderen OECD-Staaten.

Nationale Analysen haben allerdings wiederholt gezeigt, dass die Leistungsnachteile von Schülerinnen und Schülern mit Zuwanderungshintergrund in Abhängigkeit vom Herkunftsland teilweise stark variieren. So schnitten in PISA 2003 (Ramm, Prenzel, Heidemeyer & Walter, 2004) Jugendliche der zweiten Generation (mit einem Abstand von 49 Punkten zum Mittelwert für Deutschland) weniger schwach ab als Jugendliche der ersten Generation (mit einem Abstand von 71 Punkten). Jugendliche der ersten Generation, die aus der Türkei zugewandert waren, erreichten einen Mittelwert von 411 Punkten, während die aus dem Gebiet der ehemaligen UdSSR zugewanderten 466 Punkte erreichten (Ramm et al., 2004). Diese Kompetenzunterschiede zeigten sich auch unter Kontrolle des sozioökonomischen Status beider Gruppen (Ramm et al., 2004), sodass die verbleibenden Unterschiede in der Mathematikkompetenz nicht auf eine Ungleichverteilung sozioökonomischer Ressourcen in den Herkunftsgruppen zurückgeführt werden können. In PISA 2012 erreichen die Jugendlichen in Deutschland ohne Zuwanderungshintergrund in Mathematik einen Mittelwert von 531 Punkten. Schülerinnen und Schüler mit Zuwanderungshintergrund schneiden schwächer ab, wobei sich auch innerhalb dieser Gruppe deutliche Unterschiede zeigen. Die Fünfzehnjährigen, deren Familien aus Polen stammen, erzielen im Durchschnitt 504 Punkte; die Jugendlichen mit einer Familie aus dem Gebiet der ehemaligen UdSSR erreichen 488 Punkte. Schülerinnen und Schüler mit Zuwanderungshintergrund aus der Türkei erreichen 454 Punkte in Mathematik. Jugendliche aus anderen Herkunftsländern als der Türkei, Polen und dem Gebiet der ehemaligen UdSSR erzielen im Mittel 492 Punkte und diejenigen mit unbekanntem Zuwanderungsstatus 487 Punkte.

Die beschriebenen Leistungsunterschiede in Mathematik zu Ungunsten der Jugendlichen mit Zuwanderungshintergrund in PISA 2003 bilden einen wichtigen Bezugspunkt für das vorliegende Kapitel. In den folgenden Analysen soll geprüft werden, inwieweit sich die zuwanderungsbezogenen Disparitäten in den mathematischen Kompetenzen fünfzehnjähriger Schülerinnen und Schüler in den letzten neun Jahren reduziert haben.

9.1 Methodische Vorbemerkungen

Bei Analysen der Situation von Schülerinnen und Schülern aus zugewanderten Familien müssen einige methodische Aspekte berücksichtigt werden. Dazu gehören die Definition und Differenzierung des Zuwanderungshintergrundes, der Umgang mit fehlenden Werten sowie die Einbeziehung von Schülerinnen und Schülern aus Förderschulen und beruflichen Schulen.

Definition des Zuwanderungshintergrunds

Seit der Veröffentlichung des internationalen Berichts „*Where Immigrant Students Succeed*“ der OECD (Stanat & Christensen, 2006) wird in PISA die in der Migrationsforschung gängige Unterscheidung zwischen der ersten und zweiten Zuwanderergeneration verwendet. Unter Berücksichtigung des Zeitpunktes, zu dem die Familie einer Schülerin oder eines Schülers nach Deutschland zugewandert ist, unterscheiden wir im vorliegenden Bericht zwischen Jugendlichen, die selbst gemeinsam mit ihren Eltern aus dem Ausland nach Deutschland gekommen sind (die sogenannte *erste Generation*), und solchen, deren Eltern nach Deutschland zuwanderten, die selbst jedoch bereits in Deutschland geboren wurden (die sogenannte *zweite Generation*). Dieses Vorgehen schließt an die nationalen PISA-Berichte 2006 und 2009 an (Stanat & Christensen, 2006; Stanat et al., 2010). Da die Schülerinnen und Schüler nicht nach dem Geburtsland ihrer Großeltern gefragt wurden, können keine Aussagen über Schülerinnen und Schüler der dritten Generation gemacht werden. Nach der hier eingesetzten Klassifikation zählen Schülerinnen und Schüler, die selbst ebenso wie ihre beiden Eltern in Deutschland geboren wurden, zu denen ohne Zuwanderungshintergrund, unabhängig vom Geburtsland der Großeltern.

Im Unterschied zur Vorgehensweise der OECD werden im vorliegenden Bericht zudem Schülerinnen und Schüler gesondert ausgewiesen, die selbst in Deutschland geboren sind, deren Mutter oder Vater jedoch aus dem Ausland stammen (*ein Elternteil im Ausland geboren*). Die OECD ordnet diese Jugendlichen der Gruppe der Schülerinnen und Schüler *ohne* Zuwanderungshintergrund zu, während sie in den deutschen Berichten zu PISA zur Gruppe der Schülerinnen und Schüler *mit* Zuwanderungshintergrund gezählt werden. Bereits die nationalen Berichte zu PISA 2003, PISA 2006 und PISA 2009 in Deutschland wichen vom Vorgehen der OECD ab und stellten die Ergebnisse der Jugendlichen mit nur einem im Ausland geborenen Elternteil getrennt dar.

Daran wird auch im vorliegenden Band festgehalten, um mit der Differenzierung an frühere PISA-Berichtsbände anzuknüpfen. Das Vorgehen hat sich bewährt, zumal sich in einigen Herkunftsgruppen auch für die Jugendlichen mit nur einem im Ausland geborenen Elternteil deutliche Disparitäten im Vergleich zu den Jugendlichen ohne Zuwanderungshintergrund zeigten (Stanat et al., 2010).

In den Analysen dieses Kapitels werden die folgenden Gruppen von Schülerinnen und Schülern unterschieden:

Ohne Zuwanderungshintergrund:	Kein Elternteil im Ausland geboren.
Ein Elternteil im Ausland geboren:	Ein Elternteil im Ausland, ein Elternteil in Deutschland (bzw. im jeweiligen OECD-Teilnehmerstaat) geboren.
Zweite Generation:	Beide Elternteile im Ausland geboren, Jugendliche/r in Deutschland (bzw. im jeweiligen OECD-Teilnehmerstaat) geboren.
Erste Generation:	Beide Elternteile und Jugendliche/r im Ausland geboren.

Differenzierung verschiedener Herkunftsgruppen

Wie bereits angesprochen sind bei einer Differenzierung der Jugendlichen nach dem Herkunftsland ihrer Familien teilweise große Unterschiede in den erreichten Kompetenzwerten festzustellen. Aktuell zeigt der jüngste Ländervergleich des IQB (Pant et al., 2013), dass die mittleren Kompetenzen von Jugendlichen unterschiedlicher Herkunftsländer deutlich voneinander abweichen (Pöhlmann, Haag & Stanat, 2013). Auch frühere PISA-Erhebungen fanden entsprechende Differenzen. Deshalb werden für die folgenden Analysen die Herkunftsländer, aus denen der größte Teil der Familien von Jugendlichen mit Migrationsstatus stammt (ehemalige UdSSR, Türkei und Polen), wie folgt differenziert:

Herkunftsland ehemalige UdSSR:	Mindestens ein Elternteil in Russland, Kasachstan oder einer anderen ehemaligen Sowjetrepublik geboren.
Herkunftsland Türkei:	Mindestens ein Elternteil in der Türkei geboren.
Herkunftsland Polen:	Mindestens ein Elternteil in Polen geboren.
Anderes Land:	Zusammengefasste Kategorie mehrerer Staaten, aus denen jeweils ein sehr geringer Anteil an Jugendlichen stammt, wobei mindestens ein Elternteil in einem dieser Staaten geboren wurde.

Jugendliche aus Familien, in denen der Vater und die Mutter aus unterschiedlichen Staaten zugewandert sind, wurden zugunsten einer eindeutigen Interpretation der Ergebnisse aus den Analysen ausgeschlossen. Jugendliche mit einem im Ausland und einem in Deutschland geborenen Elternteil wurden nach dem Herkunftsland des zugewanderten Elternteils einer Herkunftsgruppe zugeordnet.

Umgang mit fehlenden Werten

Nicht für alle an der PISA-Erhebung 2012 teilnehmenden Schülerinnen und Schüler liegen die zur Bestimmung des Zuwanderungshintergrundes notwendigen Angaben zum Geburtsland der Eltern und des Jugendlichen selbst vor (siehe Kapitel 1). Solche fehlenden Informationen sind in sozialwissenschaftlichen Studien ein gängiges Problem, für das es in der Regel keine vollständig befriedigende Lösung gibt. Auch die ansonsten sehr hohe Qualität der im Rahmen von PISA gewonnenen Daten ist in Bezug auf den Zuwanderungshintergrund aufgrund fehlender Angaben eingeschränkt. Wichtig sind deshalb größtmögliche Transparenz und Konsistenz in Bezug auf den Umgang mit fehlenden Angaben (Edele & Stanat, 2011).

Wie bereits in früheren PISA-Berichtsbänden für Deutschland wurde für den vorliegenden Bericht die Variable Migrationsstatus um die Kategorie „nicht zuzuordnen“ ergänzt. Dabei handelt es sich um Jugendliche, für die keine vollständigen Informationen zum eigenen Geburtsland und zum Geburtsland der Eltern vorliegen. Diese Kategorie wird in allen nationalen Analysen, die sich auf den Generationsstatus der Schülerinnen und Schüler mit Zuwanderungshintergrund beziehen, verwendet und explizit in den Tabellen ausgewiesen.

Der Anteil der Jugendlichen mit unbekanntem Zuwanderungshintergrund ist über die Erhebungsrunden hinweg bei PISA seit 2000 gestiegen. Während in PISA 2000 für lediglich 2 Prozent der Fünfzehnjährigen der Zuwanderungshintergrund nicht identifiziert werden konnte, betraf dies in PISA 2009 bereits 11 Prozent der Schülerinnen und Schüler (Stanat et al., 2010). Dieser Anteil ist in PISA 2012 noch einmal angestiegen und liegt nunmehr bei 17,6 Prozent. Tendenziell erzielen Schülerinnen und Schüler mit fehlenden Angaben zum Zuwanderungshintergrund vergleichsweise schwache Leistungen. Der wachsende prozentuale Anteil an Fünfzehnjährigen, für die keine belastbaren oder vollständigen Informationen zu ihrem Geburtsland und dem ihrer Eltern vorhanden sind, hat mehrere Gründe. Der Hauptgrund liegt darin, dass die Bearbeitung des Schülerfragebogens nur in einem kleinen Teil der Länder in Deutschland verpflichtend ist. In den anderen Ländern verlangen Datenschutzrichtlinien das Einverständnis der Erziehungsberechtigten und der Jugendlichen für die Bearbeitung der Fragebögen (und in einigen Ländern speziell für die Beantwortung der Fragen zum Zuwanderungsstatus).

Während seit 2000 die Ausschöpfung der Stichprobe durch die verpflichtende Testteilnahme von 86 Prozent (Baumert, Stanat & Demmrich, 2001, S. 37) auf mehr als 93 Prozent (PISA 2012) gestiegen ist, stieg die Bereitschaft der Schülerinnen und Schüler

(bzw. der Erziehungsberechtigten), den Fragebogen auszufüllen, nicht gleichermaßen an. Bearbeitete Schülerfragebögen liegen bei PISA 2012 von 82 Prozent der Schülerinnen und Schüler vor. Darüber hinaus haben einige Schülerinnen und Schüler, die den Fragebogen bearbeitet haben, die Fragen zum eigenen Geburtsland und dem Geburtsland der beiden Elternteile nicht beantwortet. Insgesamt resultiert daraus ein Anteil von 17.6 Prozent der getesteten Stichprobe, für den der Zuwanderungsstatus nicht sicher festgestellt werden kann.

Anzumerken ist in diesem Zusammenhang weiterhin, dass die OECD mit fehlenden Werten für die Herkunftsländer der Eltern in einem Punkt anders umgeht als die nationale Berichterstattung in Deutschland. Während die OECD im Falle einer fehlenden Angabe zum Geburtsland eines Elternteils das Geburtsland des Partners verwendet und den fehlenden Wert damit ersetzt, wird die Variable zum Zuwanderungsstatus in Deutschland so gebildet, dass zunächst die Information aus dem Schülerfragebogen herangezogen wird und, falls diese fehlt, der Elternfragebogen als Quelle dient. Fehlt die Angabe zum Herkunftsländ eines Elternteils vollständig, so wird dem jeweiligen Jugendlichen die Kategorie „nicht zuzuordnen“ zugewiesen.

Schülerinnen und Schüler in Förderschulen sowie beruflichen Schulen

Analysen der PISA-Stichprobe in Deutschland erlauben Rückschlüsse auf die Population der fünfzehnjährigen Jugendlichen, die sich in Deutschland in Schulausbildung befinden. In dieser Stichprobe sind Schülerinnen und Schüler aller Schularten der Sekundarstufe I enthalten (Hauptschulen, Schulen mit mehreren Bildungsgängen, Integrierte Gesamtschulen, Realschulen, Gymnasien, Förderschulen und berufliche Schulen). Wie in den anderen Kapiteln dieses Bandes gilt, dass bei Analysen mit der gesamten Stichprobe die Schülerinnen und Schüler aller sieben Schularten berücksichtigt sind, bei Differenzierungen nach Schularten jedoch die Förderschulen und beruflichen Schulen nicht ausgewiesen werden (vgl. Kapitel 1).

9.2 Jugendliche mit Zuwanderungshintergrund im internationalen Vergleich

Für internationale Vergleiche sind in diesem Kapitel vor allem diejenigen Staaten von Interesse, in denen mindestens 10 Prozent der Fünfzehnjährigen einen Zuwanderungshintergrund haben. Diese Jugendlichen haben zum Zeitpunkt des PISA-Tests bereits unterschiedlich lange im jeweiligen Land gelebt, mindestens jedoch ein Jahr.

Im Hinblick auf die Geschichte und Bedingungen der Zuwanderung werden die Staaten nach verschiedenen Gruppen unterteilt dargestellt (Walter & Taskinen, 2008). Diese Differenzierung hilft, Randbedingungen der Zuwanderung zu berücksichtigen, die unter anderem den Bildungsstand der Eltern oder das Verhältnis von Erstsprache

und Verkehrssprache im Zuwanderungsland betreffen. Australien, Kanada, Neuseeland und die Vereinigten Staaten sind sogenannte *klassische Einwanderungsländer*. Sie gewähren einer beträchtlichen Anzahl an Einwanderern eine unbefristete Aufenthaltserlaubnis und die Möglichkeit zur Einbürgerung. Die Einwanderung in diesen Staaten ist zumeist an bestimmte Bedingungen wie Qualifikation und Bedarfslagen gebunden und in dieser Hinsicht selektiv. *Mitteleuropäische ehemalige Kolonialstaaten* hatten größere Kolonialgebiete außerhalb Europas und deren Sprache in den Kolonien als eine der Amtssprachen eingeführt. Seit dem 20. Jahrhundert wandern aus den ehemaligen Kolonien vermehrt Menschen in die ehemaligen Kolonialstaaten ein, die teilweise bereits die Landessprache des Einwanderungslandes beherrschen. Zu den Kolonialstaaten werden Belgien, Frankreich, die Niederlande und das Vereinigte Königreich gezählt. *Zielländer für Arbeitsmigration und humanitäre Zuwanderung* sind Staaten wie Deutschland, Luxemburg, Österreich und die Schweiz, die bis Ende der 1960er Jahre im Ausland Arbeitskräfte anwarben, um den Arbeitskräftemangel auszugleichen. Mit der ersten Ölkrise beendeten die Staaten ihre aktive Anwerbepolitik und ließen Zuwanderung nahezu ausschließlich aus humanitären Gründen und zur Familienzusammenführung zu. Aufgrund des demographischen Wandels ändern diese Staaten jedoch seit einiger Zeit wieder ihre Zuwanderungspolitik, um dem sich abzeichnenden Bedarf an Fachkräften zu begegnen (vgl. OECD, 2012a).

Mit der Zuwanderungsgeschichte der genannten Staaten sind Unterschiede in der Einwanderungspolitik verbunden (Boswell, 2007), die sich unter anderem in den Anteilen von zugewanderten Personen in der Bevölkerung dieser Staaten niederschlagen. In Tabelle 9.1 sind die prozentualen Anteile der Schülerinnen und Schüler mit und ohne Zuwanderungshintergrund in den einzelnen Staaten anhand der Daten aus PISA 2012 dargestellt. Des Weiteren wurden Veränderungen dieses Anteils zwischen PISA 2003 und PISA 2012 in der mit +/- gekennzeichneten Spalte angegeben.

In der Tabelle sind OECD-Staaten mit einem substantiellen Anteil an Schülerinnen und Schülern mit Zuwanderungshintergrund aufgeführt (mindestens 10 Prozent). Dennoch unterschieden sich diese Prozentwerte in den verschiedenen Staaten beträchtlich. Einen Anteil von mehr als 30 Prozent Jugendliche mit Zuwanderungshintergrund weisen ausschließlich die *klassischen Einwanderungsländer* Australien, Kanada und Neuseeland sowie die *Zielländer für Arbeitsmigration und humanitäre Zuwanderung* Luxemburg und Schweiz auf. In den meisten der *ehemaligen Kolonialstaaten* sowie in einem Teil der *Zielländer für Arbeitsmigration und humanitäre Zuwanderung* haben zwischen 20 und 29 Prozent der Schülerinnen und Schüler einen Zuwanderungshintergrund. Lediglich in den Niederlanden und den skandinavischen Staaten Dänemark, Finnland und Norwegen ist der Anteil der Fünfzehnjährigen mit Zuwanderungshintergrund geringer. Deutschland liegt mit einem Anteil von 25,8 Prozent im Mittelfeld der *Zielländer für Arbeitsmigration und humanitäre Zuwanderung*. Der Mikrozensus des Jahres 2011 gibt für Jugendliche in Deutschland einen Anteil von 27,1 Prozent mit Zuwanderungshintergrund an (Statistisches Bundesamt, 2012, S. 33), wobei sich diese Zahl auf Jugendliche im Alter von 15 bis 20 Jahren bezieht, die nicht notwendigerweise eine allgemeinbildende Schule

Tabelle 9.1: Prozentuale Anteile von fünfzehnjährigen Schülerinnen und Schülern mit Zuwanderungshintergrund in ausgewählten OECD-Staaten

Staat	Ohne Zuwanderungshintergrund		Mit Zuwanderungshintergrund				nicht zuzuordnen				
	gültige %* (SE)	Insgesamt Zuwanderung	Ein Elternteil im Ausland geboren	Zweite Generation	Erste Generation	zuzuordnen					
	gültige %* (SE)	gültige %* (SE)	+/- gültige %* (SE)	+/- gültige %* (SE)	+/- gültige %* (SE)	+/- gültige %* (SE)	+/- zuzuordnen				
Klassische Einwanderungsländer											
Australien	60.6 (0.8)	39.4 (0.8)	-2.3 (0.8)	16.8 (0.4)	-2.2 (0.6)	12.1 (0.6)	0.5 (0.4)	-0.6 (0.4)	89.7	10.3	
Vereinigte Staaten	70.0 (2.2)	30.0 (2.2)	8.7 (2.2)	8.6 (0.7)	1.6 (1.4)	14.7 (1.4)	6.6 (0.8)	6.7 (0.8)	0.6 (0.8)	95.4	4.6
Neuseeland	57.4 (1.6)	42.6 (1.6)	6.7 (1.6)	16.1 (0.7)	-0.1 (0.8)	9.5 (0.8)	3.3 (1.0)	16.9 (1.0)	3.6 (1.0)	96.6	3.4
Kanada	60.0 (1.2)	40.0 (1.2)	9.7 (1.2)	10.5 (0.4)	0.0 (0.8)	16.4 (0.8)	7.5 (0.7)	13.1 (0.7)	2.3 (0.7)	95.1	4.9
Mitteleuropäische ehemalige Kolonialstaaten											
Belgien	71.1 (1.1)	28.9 (1.1)	5.9 (1.1)	13.8 (0.6)	2.4 (0.6)	7.9 (0.6)	1.6 (0.6)	7.2 (0.6)	1.9 (0.6)	96.6	3.4
Frankreich	73.3 (1.4)	26.7 (1.4)	0.4 (1.4)	11.9 (0.6)	-0.4 (0.8)	9.9 (0.8)	-0.7 (0.5)	4.9 (0.5)	1.5 (0.5)	96.3	3.7
Niederlande	80.9 (1.2)	19.1 (1.2)	1.0 (1.2)	8.5 (0.5)	1.2 (0.9)	8.0 (0.9)	1.1 (0.3)	2.6 (0.3)	-1.2 (0.3)	97.3	2.7
Vereinigtes Königreich	76.2 (1.3)	23.8 (1.3)	5.4 (1.3)	11.1 (0.6)	0.6 (0.5)	5.5 (0.5)	0.3 (0.8)	7.2 (0.8)	4.5 (0.8)	96.3	3.7
Zielländer für Arbeitsmigration und humanitäre Zuwanderung											
Dänemark	82.6 (0.9)	17.4 (0.9)	3.7 (0.9)	8.3 (0.4)	1.0 (0.5)	6.1 (0.5)	2.7 (0.4)	3.0 (0.2)	0.1 (0.2)	97.7	2.3
Deutschland	74.2 (1.0)	25.8 (1.0)	5.0 (1.0)	10.2 (0.5)	4.7 (0.8)	11.9 (0.8)	5.1 (0.8)	3.7 (0.4)	-4.8 (0.4)	82.4	17.6
Finnland	90.8 (0.3)	9.2 (0.3)	4.9 (0.3)	5.8 (0.2)	3.5 (0.1)	1.5 (0.1)	1.4 (0.2)	1.9 (0.2)	0.0 (0.2)	98.2	1.8
Luxemburg	37.1 (0.6)	62.9 (0.6)	13.6 (0.6)	16.9 (0.5)	0.5 (0.6)	28.6 (0.6)	12.9 (0.6)	17.3 (0.5)	0.2 (0.5)	97.0	3.0
Norwegen	81.2 (1.0)	18.8 (1.0)	5.9 (1.0)	9.3 (0.6)	2.0 (0.6)	4.7 (0.6)	2.4 (0.6)	4.8 (0.5)	1.5 (0.5)	97.1	2.9
Österreich	74.5 (1.1)	25.5 (1.1)	7.0 (1.1)	9.1 (0.6)	3.9 (0.7)	10.8 (0.7)	6.7 (0.6)	5.5 (0.6)	-3.5 (0.6)	98.2	1.8
Schweden	72.5 (1.1)	27.5 (1.1)	7.0 (1.1)	12.7 (0.6)	3.5 (0.6)	8.7 (0.6)	3.1 (0.6)	6.1 (0.5)	0.4 (0.5)	96.9	3.1
Schweiz	56.7 (1.1)	43.3 (1.1)	7.6 (1.1)	19.2 (0.6)	3.3 (0.7)	17.4 (0.7)	8.5 (0.4)	6.7 (0.4)	-4.3 (0.4)	97.4	2.6

Anmerkung: In der Tabelle werden gerundete Werte angegeben. Es können auf die Rundungen zurückzuführende vermeintliche Inkonsistenzen vorkommen.

+/-: Veränderung gegenüber PISA 2003

fett: signifikante Veränderungen gegenüber PISA 2003 ($p < .05$)

* gültige % = Prozentangaben beruhen nur auf Angaben der Schülerinnen und Schüler, die eindeutig zuzuordnen sind.

besuchen. Im Vergleich zum Mikrozensus scheint der in der PISA-2012-Erhebung gestiegene Anteil an nicht zuzuordnenden Schülerinnen und Schülern demnach nur zu einer leichten Unterschätzung des Anteils von Jugendlichen mit Zuwanderungshintergrund zu führen.

Im Vergleich zu PISA 2003 hat sich der Anteil der Schülerinnen und Schüler mit Zuwanderungshintergrund in allen Zielländern für Arbeitsmigration und humanitäre Zuwanderung erhöht, aber auch in einigen klassischen Einwanderungsländern und ehemaligen Kolonialstaaten. Besonders deutlich ist die Zunahme in Kanada und Luxemburg. In Deutschland stieg der Anteil von Schülerinnen und Schülern mit Zuwanderungshintergrund seit PISA 2003 um 5 Prozentpunkte an und wuchs damit ähnlich stark wie in den anderen europäischen Staaten, die als Zielländer für Arbeitsmigration und humanitäre Zuwanderung klassifiziert sind. In Deutschlands Nachbarländern Frankreich und Niederlande, beides ehemalige Kolonialstaaten, blieb der Anteil der Schülerinnen und Schüler mit Zuwanderungshintergrund dagegen weitestgehend konstant.

Merkmale des familiären Hintergrundes von Jugendlichen mit Zuwanderungshintergrund

Um die Lebens- und Lernumwelten der Jugendlichen mit Zuwanderungshintergrund näher zu beschreiben, wird in den nächsten Schritten die in der Familie gesprochene Sprache und der sozioökonomische Hintergrund der Familien berichtet. Dabei gibt der Vergleich zwischen erster und zweiter Generation Hinweise darauf, inwieweit eine Annäherung von Zuwanderern und Mitgliedern der Aufnahmegesellschaft im Generationsverlauf erfolgt. Dieser Prozess wird in der Migrationsforschung als *Assimilation* bezeichnet. Dabei wird *Assimilation* als deskriptives Konzept verstanden, das die Annäherung von Zuwanderergruppen und Aufnahmegesellschaft über mehrere Generationen hinweg abbildet (Alba & Nee, 2003). Man spricht von „vollständiger *Assimilation*“, wenn Unterschiede zwischen der Zuwanderergruppe und den Mitgliedern der Aufnahmegesellschaft verschwunden sind. In der Soziologie werden strukturelle, kulturelle, soziale und identifikative Dimensionen der *Assimilation* unterschieden (z. B. Esser, 2001). Während die Verwendung der Sprache des Aufnahmelandes in Familien mit Zuwanderungshintergrund als ein Indikator für die *kulturelle Assimilation* gesehen werden kann, handelt es sich beim sozioökonomischen Status um einen Indikator für *strukturelle Assimilation* (Stanat et al., 2010).

In Tabelle 9.2 sind die Anteile der Schülerinnen und Schüler mit Zuwanderungshintergrund dargestellt, die in ihrer Familie am häufigsten die Sprache des jeweiligen Einwanderungslandes sprechen (bzw. die Sprache, in der sie den PISA-Test und den Fragebogen bearbeiten). Die Fünfzehnjährigen wurden dazu im Schülerfragebogen befragt („*Welche Sprache sprichst du am häufigsten zu Hause?*“). Als Antwortkategorien standen die zehn in früheren PISA-Erhebungsrunden am häufigsten angegebenen Sprachen zur Auswahl sowie eine offene Antwortkategorie, in die die Fünfzehnjährigen bei Bedarf die Sprache selbst eintragen konnten, die sie zu Hause am häufigsten sprechen.

Tabelle 9.2: Prozentuale Anteile von fünfzehnjährigen Schülerinnen und Schülern mit Zuwanderungshintergrund, die zu Hause die Sprache des Einwanderungslandes sprechen

Staat	Mit Zuwanderungshintergrund							
	Insgesamt Zuwanderung		Ein Elternteil im Ausland geboren		Zweite Generation		Erste Generation	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)
Klassische Einwanderungsländer								
Australien	79.4	(1.0)	96.1	(0.5)	67.1	(1.8)	65.4	(1.9)
Vereinigte Staaten	54.9	(2.4)	85.9	(2.0)	48.4	(2.8)	28.7	(3.7)
Neuseeland	67.2	(1.7)	95.2	(1.0)	59.8	(2.9)	44.6	(2.4)
Kanada	62.2	(1.4)	92.5	(0.9)	61.8	(1.9)	37.8	(1.6)
Mitteuropäische ehemalige Kolonialstaaten								
Belgien	61.7	(1.8)	75.3	(2.0)	49.2	(2.7)	47.3	(4.7)
Frankreich	74.2	(1.5)	91.8	(1.5)	67.9	(2.7)	41.6	(3.5)
Niederlande	69.5	(1.8)	91.8	(1.8)	54.3	(3.1)	32.4	(5.2)
Vereinigtes Königreich	72.7	(2.0)	94.3	(1.0)	75.8	(2.7)	35.4	(3.6)
Zielländer für Arbeitsmigration und humanitäre Zuwanderung								
Dänemark	71.8	(1.3)	94.9	(0.8)	54.2	(2.3)	33.6	(3.2)
Deutschland	72.1	(1.8)	89.1	(1.7)	65.8	(2.1)	35.5	(5.3)
Finnland	63.8	(1.5)	89.9	(1.2)	27.3	(3.1)	11.2	(1.9)
Luxemburg	23.8	(0.7)	17.3	(1.2)	22.1	(1.2)	33.2	(1.6)
Norwegen	63.8	(2.3)	93.0	(1.5)	45.0	(3.7)	23.1	(2.7)
Österreich	50.4	(2.1)	88.0	(2.1)	25.7	(2.5)	29.0	(3.2)
Schweden	60.3	(1.8)	91.4	(1.7)	43.6	(2.5)	16.3	(2.7)
Schweiz	65.5	(1.2)	88.8	(1.1)	48.0	(1.8)	41.7	(2.7)

Der Anteil der Jugendlichen mit Zuwanderungshintergrund, die in ihrer Familie am häufigsten Deutsch sprechen, liegt in Deutschland insgesamt bei über 70 Prozent, wobei er in der ersten Generation deutlich geringer ist (35.5 Prozent) als in der zweiten Generation (65.8 Prozent). Dieses Muster ist mit den theoretischen Annahmen eines Assimilationsprozesses vereinbar und in den meisten Staaten zu beobachten. Ausnahmen sind dabei Österreich und Luxemburg, wo die Jugendlichen der ersten Generation häufiger zu Hause Deutsch (bzw. die Testsprache) sprechen als die Jugendlichen der zweiten Generation. In der zweiten Generation sprechen in Deutschland mehr als 60 Prozent der Jugendlichen mit Zuwanderungshintergrund zu Hause am häufigsten Deutsch. Dieser hohe Prozentsatz ist ansonsten nur in den *klassischen Einwanderungsländern* und zwei ehemaligen Kolonialstaaten, Frankreich und dem Vereinigten Königreich, zu beobachten.

Als Indikator für den sozioökonomischen Hintergrund wird in PISA unter anderem mit dem *Highest International Socio-Economic Index of Occupational Status* (HISEI, Ganzeboom, de Graaf, Treiman & de Leeuw, 1992) verwendet, der auf der Basis der sogenannten *International Standard Classification of Occupations* (ISCO) gebildet wird. Diese wurde erstmals im Jahre 1988 festgelegt und basiert auf vierstelligen Codes für jeden Beruf. Während frühere PISA-Erhebungen auf den ISCO-88 zurückgriffen (International Labour Office, 1990), wird der HISEI in PISA 2012 anhand der im Jahr 2008 neu gebildeten Berufsklassifizierung – ISCO-08 – gebildet (International Labour Office, 2007). Diese berücksichtigt die Veränderungen am Arbeitsmarkt in den letzten 25 Jahren (vgl. Kapitel 8).

Tabelle 9.3: Disparitäten im sozioökonomischen Status zwischen Familien ohne und Familien mit Zuwanderungshintergrund in ausgewählten OECD-Staaten

Staat	Ohne Zuwanderungshintergrund		Mit Zuwanderungshintergrund							
			Insgesamt Zuwanderung		Ein Elternteil im Ausland geboren		Zweite Generation		Erste Generation	
	M	(SE)	DIFF	(SE)	DIFF	(SE)	DIFF	(SE)	DIFF	(SE)
Klassische Einwanderungsländer										
Australien	57.5	(0.3)	0.4	(0.4)	1.7	(0.5)	-2.7	(0.7)	2.0	(0.7)
Vereinigte Staaten	57.4	(0.7)	-9.5	(1.4)	0.0	(1.4)	-13.3	(1.8)	-13.8	(1.9)
Neuseeland	54.6	(0.6)	1.5	(0.9)	3.6	(0.8)	-5.4	(1.6)	3.2	(1.3)
Kanada	55.8	(0.4)	-0.8	(0.8)	2.9	(0.8)	-6.2	(1.0)	2.9	(1.1)
Mitteuropäische ehemalige Kolonialstaaten										
Belgien	53.7	(0.4)	-6.8	(0.8)	-3.9	(1.0)	-10.4	(1.3)	-8.6	(1.5)
Frankreich	53.6	(0.6)	-6.2	(1.0)	0.1	(1.1)	-11.5	(1.3)	-11.1	(1.8)
Niederlande	57.5	(0.5)	-6.0	(1.0)	-0.5	(1.2)	-11.6	(1.4)	-8.7	(3.1)
Vereinigtes Königreich	55.1	(0.5)	3.0	(1.0)	4.0	(1.1)	2.0	(1.5)	2.1	(1.8)
Zielländer für Arbeitsmigration und humanitäre Zuwanderung										
Dänemark	55.7	(0.6)	-5.9	(0.9)	1.3	(1.0)	-13.8	(1.2)	-12.8	(1.7)
Deutschland	53.7	(0.5)	-10.7	(0.8)	-4.1	(1.3)	-16.6	(1.0)	-10.3	(1.9)
Finnland	55.9	(0.4)	-3.3	(0.8)	0.9	(1.0)	-9.8	(1.6)	-12.3	(1.4)
Luxemburg	56.2	(0.4)	-11.1	(0.5)	-2.2	(0.8)	-16.5	(0.7)	-10.9	(0.9)
Norwegen	59.5	(0.4)	-2.8	(0.9)	3.0	(1.2)	-7.5	(1.4)	-10.3	(1.8)
Österreich	50.3	(0.5)	-5.4	(1.1)	5.7	(1.3)	-12.3	(1.1)	-10.5	(1.9)
Schweden	55.7	(0.5)	-4.2	(1.0)	2.1	(1.1)	-9.2	(1.4)	-11.8	(2.4)
Schweiz	57.6	(0.6)	-5.9	(0.8)	4.0	(0.9)	-14.4	(0.8)	-12.2	(1.6)

Anmerkung: **fett**: signifikante Differenz zu Familien ohne Zuwanderungshintergrund ($p < .05$).

Allerdings erschwert die neue Berufsklassifikation einen Vergleich der aktuellen HISEI-Werte mit denen früherer PISA-Erhebungen. Deshalb wird auf solche Vergleiche im Folgenden verzichtet. In Tabelle 9.3 sind die Mittelwerte des HISEI für Schülerinnen und Schüler ohne Zuwanderungshintergrund eingetragen. Für die Jugendlichen mit Zuwanderungshintergrund ist die Differenz (DIFF) zu dieser Schülergruppe dargestellt.

In den klassischen Einwanderungsländern Australien, Kanada und Neuseeland haben die Familien von Schülerinnen und Schülern der ersten Generation einen signifikant höheren sozioökonomischen Status als Schülerinnen und Schüler ohne Zuwanderungshintergrund. Hierin spiegelt sich die selektive Einwanderungspolitik dieser Aufnahme-länder wider. Die zweite Generation in diesen drei Staaten weist jedoch im Vergleich zu Familien ohne Zuwanderungshintergrund einen niedrigeren sozioökonomischen Status auf. In den Vereinigten Staaten sowie in den mitteleuropäischen ehemaligen Kolonialstaaten Belgien, Frankreich und Niederlande haben die Familien sowohl von Jugendlichen der ersten als auch von Jugendlichen der zweiten Generation einen signifikant niedrigeren sozioökonomischen Status als Schülerinnen und Schüler ohne Zuwanderungshintergrund. Dasselbe gilt für alle Staaten, die den Zielländern für Arbeitsmigration und humanitäre Zuwanderung zugerechnet werden.

Die Familien von Schülerinnen und Schülern mit *inem im Ausland geborenen Elternteil* dagegen haben in den meisten Staaten einen ähnlichen oder höheren sozioökonomischen Status als die Familien von nicht zugewanderten Schülerinnen und Schülern. Lediglich in Belgien, Deutschland und Luxemburg weisen die Schülerinnen und Schüler mit einem im Ausland geborenen Elternteil einen niedrigeren durchschnittlichen Sozialstatus auf.

In Deutschland ist der durchschnittliche sozioökonomische Status (HISEI) von Jugendlichen der *ersten Generation* höher als der von Jugendlichen der *zweiten Generation*. Dieses Muster ist auch in den klassischen Einwanderungsstaaten Australien, Neuseeland und Kanada zu erkennen. Diese Unterschiede zwischen den Generationen von Zuwanderern gehen vermutlich auch auf Unterschiede in der Zusammensetzung der ersten und der zweiten Generation zurück. Im Vergleich zu anderen europäischen Nachbarstaaten unterscheidet sich in Deutschland der sozioökonomische Status der Familien mit Zuwanderungshintergrund insgesamt besonders deutlich vom Gesamtmittelwert. Der sozioökonomische Status von Zuwandererfamilien in Deutschland ist im Durchschnitt deutlich niedriger als derjenige nicht zugewanderter Familien.

Mittlere Kompetenz von Jugendlichen mit Zuwanderungshintergrund in Mathematik

In Tabelle 9.4 sind die Kompetenzunterschiede zwischen Jugendlichen mit Zuwanderungshintergrund und ihren Mitschülerinnen und Mitschülern ohne Zuwanderungshintergrund im internationalen Vergleich dargestellt; es werden jeweils der Mittelwert (M) und der Standardfehler (SE) für die einzelnen Gruppen angegeben.

Tabelle 9.4: Mittlere mathematische Kompetenz von Jugendlichen mit Zuwanderungshintergrund in ausgewählten OECD-Staaten

Staat	Ohne Zuwanderungshintergrund			Insgesamt Zuwanderung			Mit Zuwanderungshintergrund			Erste Generation					
	M	(SE)	+/-	M	(SE)	+/-	M	(SE)	+/-	M	(SE)	+/-			
Klassische Einwanderungsländer															
Australien	500	1.6	-25.2	522 ^a	2.8	-5.6	512 ^a	2.3	-19.8	540 ^a	5.4	17.6	517 ^a	3.7	-8.6
Vereinigte Staaten	487	3.9	-2.1	477	5.3	3.6	484	5.7	-11.5	479	6.7	9.7	463 ^a	9.2	10.3
Neuseeland	498	3.1	-26.9	508 ^a	3.3	-18.9	518 ^a	3.8	-22.9	490	6.8	-6.4	507	5.3	-15.5
Kanada	521	1.8	-15.3	522	3.4	-17.2	527	3.5	-14.3	513	4.5	-31.6	528	5.2	-3.7
Mitteleuropäische ehemalige Kolonialstaaten															
Belgien	532	2.1	-18.4	481 ^a	3.9	1.6	499 ^a	3.9	-13.2	470 ^a	6.1	15.5	458 ^a	7.5	20.1
Frankreich	511	3.0	-10.4	461 ^a	4.9	-27.2	488 ^a	5.5	-25.8	448 ^a	6.8	-24.5	424 ^a	10.5	-24.1
Niederlande	533	3.4	-18.9	491 ^a	6.8	-17.3	512 ^a	8.0	-30.8	475 ^a	9.3	-17.1	471 ^a	10.2	-0.1
Vereinigtes Königreich	496	3.2	-13.8	501	5.4	-7.9	516 ^a	5.4	-2.8	484	7.5	-19.9	491	11.8	11.8
Zielländer für Arbeitsmigration und humanitäre Zuwanderung															
Dänemark	509	2.3	-12.0	470 ^a	3.4	-11.8	501	4.6	-7.1	447 ^a	4.0	-1.3	428 ^a	5.4	-25.1
Deutschland	531	3.4	3.6	485 ^a	4.1	24.0	504 ^a	5.4	-2.1	476 ^a	5.3	44.9	461 ^a	9.0	5.3
Finnland	524	1.9	-22.9	485 ^a	4.3	-19.5	512 ^a	5.0	-16.9	451 ^a	4.8	25.9	426 ^a	7.8	-48.0
Luxemburg	517	2.1	7.1	477 ^a	1.7	-1.5	498 ^a	3.0	0.0	470 ^a	2.5	-7.4	469 ^a	4.0	7.4
Norwegen	496	3.0	-3.9	471 ^a	4.9	0.1	494	4.9	4.3	458 ^a	9.5	-2.3	441 ^a	6.2	2.6
Österreich	518	2.8	3.3	473 ^a	4.4	-0.6	502 ^a	6.3	-20.2	458 ^a	5.2	-1.5	454 ^a	8.6	2.6
Schweden	491	2.4	-26.7	455 ^a	3.9	-26.8	483	3.6	-31.4	445 ^a	5.3	-39.3	411 ^a	7.6	-15.6
Schweiz	552	3.3	5.6	507 ^a	3.6	12.1	534 ^a	4.1	5.2	490 ^a	3.8	5.8	472 ^a	5.8	18.4

Anmerkung: **fett** = signifikante Unterschiede zwischen 2003 und 2012 ($p < .05$).

^a signifikante Unterschiede zu Schülerinnen und Schülern ohne Zuwanderungshintergrund ($p < .05$)

+/- Mittelwertsdifferenz 2012–2003

Schülerinnen und Schüler *ohne Zuwanderungshintergrund* erreichen in Deutschland im Durchschnitt 531 Punkte auf der Skala mathematischer Kompetenz. Hingegen liegt die mathematische Kompetenz der Fünfzehnjährigen *mit Zuwanderungshintergrund* in Deutschland im Mittel bei 485 Punkten. Ein Mittelwert in vergleichbarer Höhe wird in Staaten wie den USA oder Schweden von der Gruppe der Jugendlichen *ohne Zuwanderungshintergrund* erzielt. Demnach schneiden in Deutschland die Schülerinnen und Schüler mit Zuwanderungshintergrund im Mathematiktest ähnlich ab wie die Jugendlichen in den USA oder Schweden, deren Familie nicht zugewandert ist. Dennoch handelt es sich bei der Differenz von 46 Punkten zwischen Zugewanderten und Nicht-Zugewanderten in Deutschland um einen erheblichen Kompetenzunterschied, der einem Leistungsabstand von mehr als einem Schuljahr entspricht (Autorengruppe Bildungsberichterstattung, 2012; Ehmke, Werner, Neubrand, Joran & Ulfig, 2006; Prenzel et al., 2006).

Eine Disparität zwischen Fünfzehnjährigen *mit* und solchen *ohne* Zuwanderungsgeschichte in vergleichbarer Größenordnung findet sich in fast allen Staaten bis auf die klassischen Einwanderungsländer Kanada und Vereinigte Staaten. In Australien, Neuseeland und dem Vereinigten Königreich erzielen Schülerinnen und Schüler *mit Zuwanderungshintergrund* dagegen bessere Leistungen als Schülerinnen und Schüler *ohne Zuwanderungshintergrund*. In Australien erzielen die Schülerinnen und Schüler der *zweiten Generation* die höchsten Kompetenzwerte in Mathematik, und dies obwohl diese Gruppe einen niedrigeren sozioökonomischen Status aufweist als die Schülerinnen und Schüler *ohne Zuwanderungshintergrund*.

In den meisten ehemaligen Kolonialstaaten und den Zielländern für Arbeitsmigration und humanitäre Zuwanderung erreichen die Schülerinnen und Schüler der *ersten Generation* die niedrigsten Werte auf der Mathematikskala. Die Schülerinnen und Schüler der *zweiten Generation* erzielen einige Punkte mehr, auch ihre Leistungen bleiben jedoch hinter denen der Jugendlichen ohne Zuwanderungshintergrund zurück. Auch die Schülerinnen und Schüler mit *einem im Ausland geborenen Elternteil* erreichen nicht die Leistungen der Schülerinnen und Schüler *ohne Zuwanderungshintergrund*. Eine Ausnahme bildet das Vereinigte Königreich, wo keine zuwanderungsbezogenen Disparitäten vorliegen – im Gegenteil: Die Gruppe der Schülerinnen und Schüler mit einem im Ausland geborenen Elternteil weist hier im Mittel sogar höhere Kompetenzen auf.

Tabelle 9.4 gibt darüber hinaus Auskunft über die Differenz zwischen den Mittelwerten mathematischer Kompetenz in PISA 2003 und PISA 2012. Bei positiven Differenzwerten ist der Mittelwert in PISA 2012 höher in PISA 2003, bei negativen Werten ist es umgekehrt. Die Ergebnisse zeigen, dass die Schülerinnen und Schüler *ohne Zuwanderungshintergrund* in Deutschland 2012 einen ähnlichen Wert erreichten wie 2003 (PISA 2003: 527 Punkte; PISA 2012: 531 Punkte). Die Schülerinnen und Schüler der *zweiten Generation* hingegen verbesserten sich signifikant um 45 Punkte. Diese Steigerung entspricht dem Lernzuwachs von mehr als einem Schuljahr (s. o. sowie Autorengruppe Bildungsberichterstattung, 2012). Eine Verbesserung in der mathematischen Kompetenz bei Schülerinnen und Schülern mit Zuwanderungshintergrund seit PISA 2003 zeigt sich

außer in Deutschland lediglich in der Schweiz. In zahlreichen Staaten, wie zum Beispiel Finnland, Frankreich und Schweden, hingegen haben sich die Disparitäten zwischen Schülerinnen und Schülern mit und ohne Zuwanderungshintergrund über die Zeit vergrößert.

9.3 Jugendliche mit Zuwanderungshintergrund in Deutschland

Im Folgenden wird die Situation von Schülerinnen und Schülern mit Zuwanderungshintergrund für Deutschland differenzierter beschrieben. Dabei wird zunächst nach Generationsstatus und Herkunftsland gegliedert der Anteil von Jugendlichen mit Zuwanderungshintergrund in Deutschland dargestellt, bevor wichtige Merkmale des familiären Hintergrundes dieser Schülerinnen und Schüler vorgestellt werden. Schließlich wird die Mathematikkompetenz von Jugendlichen mit Zuwanderungshintergrund zunächst im Vergleich verschiedener Generationen und Herkunftsländer beschrieben, dann in Bezug auf erreichte Kompetenzstufen dargestellt und in modellbasierten Analysen mit möglichen erklärenden Kontextvariablen in Zusammenhang gebracht.

Prozentuale Anteile von Jugendlichen mit Zuwanderungshintergrund in PISA 2003 und 2012

Differenziert man die Gruppe der Fünfzehnjährigen *mit Zuwanderungshintergrund* nach deren Generationsstatus, dann nahm der Anteil an Jugendlichen der *ersten Generation* zwischen 2003 und 2012 ab, während der Anteil der Schülerinnen und Schüler der *zweiten Generation* und der mit *einem im Ausland geborenen Elternteil* jeweils angestiegen ist. Aufgrund der teilweise sehr geringen Fallzahlen in der PISA-Stichprobe mussten in den nach Herkunftsland differenzierenden Analysen (Tabelle 9.6 und 9.8) die erste und zweite Generation zusammengefasst werden.

Insgesamt haben 25.8 Prozent der in PISA 2012 getesteten und eindeutig zuzuordnenden Schülerinnen und Schüler einen Zuwanderungshintergrund; 4.5 Prozent stammen aus Ländern der ehemaligen UdSSR, 5.1 Prozent aus der Türkei und 2.7 Prozent aus Polen. Diese Zahlen haben sich im Vergleich zu 2003 nicht signifikant verändert. Die Gruppe der Schülerinnen und Schülern aus Familien mit einem *anderen* Herkunftsland als der ehemaligen UdSSR, der Türkei und Polen hat sich dagegen vergrößert. In Tabelle 9.5 werden die Anteile der Schülerinnen und Schüler mit Zuwanderungshintergrund gegliedert nach ihrem Generationsstatus und Herkunftsland bezogen auf alle eindeutig zuzuordnenden fünfzehnjährigen Schülerinnen und Schüler dargestellt.

Tabelle 9.5: Prozentuale Anteile fünfzehnjähriger Schülerinnen und Schüler mit Zuwanderungshintergrund in Deutschland

Zuwanderungsstatus	2003		2012	
	%	(SE)	%	(SE)
Zuwanderungsstatus eindeutig zuzuordnen?				
Nicht zuzuordnen	10.8	(0.7)	17.6	(0.9)
Zuzuordnen	89.2	(0.7)	82.4	(0.9)
	gültige %*	(SE)	gültige %*	(SE)
Ohne Zuwanderungshintergrund	79.1	(1.1)	74.2	(1.0)
Mit Zuwanderungshintergrund	20.9	(1.1)	25.8	(1.0)
Generationsstatus				
Ein Elternteil im Ausland geboren	5.5	(0.4)	10.2	(0.5)
Zweite Generation	6.8	(0.8)	11.9	(0.8)
Erste Generation	8.5	(0.7)	3.7	(0.4)
Herkunftsländer				
Ehemalige UdSSR	4.8	(0.5)	4.5	(0.4)
Türkei	5.4	(0.7)	5.1	(0.5)
Polen	3.3	(0.4)	2.7	(0.3)
Anderes Land	7.4	(0.5)	13.5	(0.7)

Anmerkung: In der Tabelle werden gerundete Werte angegeben. Es können auf die Rundungen zurückzuführende vermeintliche Inkonsistenzen vorkommen.

Daten beruhen auf Schülerangaben.

fett: signifikante Unterschiede zwischen 2003 und 2012 ($p < .05$)

* gültige % = Prozentangaben beruhen nur auf Angaben der Schülerinnen und Schüler, die eindeutig zuzuordnen sind.

Die Aufschlüsselung des Migrationsstatus nach Generationenfolge in Tabelle 9.5 zeigt, dass sich die Zusammensetzung der Zuwandererfamilien seit PISA 2003 verändert hat. Beispielsweise hat sich der Anteil derjenigen, von denen ein Elternteil im Ausland geboren wurde, gegenüber PISA 2003 beinahe verdoppelt. Auch die Gruppe der Schülerinnen und Schüler, die in der zweiten Generation in Deutschland leben, hat sich von 6.8 Prozent (PISA 2003) auf 11.9 Prozent vergrößert. Kleiner geworden ist hingegen der Anteil an Jugendlichen, die selbst mit ihren Eltern nach Deutschland eingewandert sind (8.5 Prozent in PISA 2003, 3.7 Prozent in PISA 2012). Die Zuwanderung der Familien der Fünfzehnjährigen, die in PISA 2012 als „zugewandert“ gelten, fand zwischen 1997 und 2011 statt, da die Teilnahme an PISA voraussetzt, dass die Schülerinnen und Schüler mindestens ein Jahr vor dem Test in Deutschland gelebt haben. Die Abnahme des Anteils von Jugendlichen der ersten Generation bedeutet, dass im Zeitraum von 1997 bis 2011 weniger Zuwanderung von Familien mit Kindern nach Deutschland erfolgte als im vergleichbaren Zeitraum 1988 bis 2002. In Bezug auf die Zusammensetzung der Schülerschaft mit Zuwanderungshintergrund hinsichtlich der Herkunftsländer hat sich seit 2003 vor allem eine größere Heterogenität eingestellt: Fast doppelt so viele Schülerinnen und

Schüler wie 2003 stammen 2012 aus einem anderen Herkunftsland als der ehemaligen UdSSR, der Türkei oder Polen.

Tabelle 9.6 stellt genauer dar, wie sich die Gruppe der Jugendlichen aus verschiedenen Herkunftsländern hinsichtlich des Geburtslandes beider Eltern zusammensetzt. Unterschieden werden dabei wiederum Familien, bei denen *ein Elternteil im Ausland* geboren wurde, und Familien, in denen *beide Eltern im Ausland* zur Welt kamen. Diese letzte Gruppe umfasst sowohl Jugendliche, die in erster Generation in Deutschland leben, als auch Jugendliche, die in zweiter Generation in Deutschland leben.

Bei den Familien, die aus dem Gebiet der ehemaligen UdSSR nach Deutschland zugewandert sind, haben sich die Anteile dieser beiden Gruppen im Vergleich zu PISA 2003 nicht verändert. Gleiches gilt für Familien aus der Gruppe der anderen Länder. Veränderungen zeigen sich bei den Zuwandererfamilien aus der Türkei und Polen. Mittlerweile hat fast jeder dritte Jugendliche, dessen Familie aus der Türkei zugewandert ist, nur einen Elternteil, der in der Türkei geboren wurde. In PISA 2003 waren dies weniger als 10 Prozent. Entsprechend gesunken ist der Anteil derjenigen, bei denen beide Eltern in der Türkei geboren wurden (68.3 Prozent in PISA 2012 gegenüber 91.9 Prozent in PISA 2003). Ein ähnliches Bild zeigt sich für die Jugendlichen, deren Familien aus Polen stammen. Von etwas mehr als jedem Vierten war in PISA 2003 ein Elternteil in Polen geboren; in PISA 2012 ist dieser Anteil signifikant gewachsen (auf 43.2 Prozent). Von mehr als der Hälfte dieser Jugendlichen sind beide Elternteile in Polen geboren; dieser prozentuale Anteil ist damit signifikant geringer als in PISA 2003 (damals noch 72.5 Prozent).

Tabelle 9.6: Prozentuale Anteile fünfzehnjähriger Schülerinnen und Schüler mit Zuwanderungshintergrund in Deutschland nach Herkunft und Geburtsland der Eltern

Herkunftsländer	2003		2012	
	gültige %*	(SE)	gültige %*	(SE)
Ehemalige UdSSR				
Ein Elternteil im Ausland geboren	9.7	(2.4)	10.9	(2.4)
Beide Elternteile im Ausland geboren	90.3	(2.4)	89.1	(2.4)
Türkei				
Ein Elternteil im Ausland geboren	8.1	(1.5)	31.7	(2.9)
Beide Elternteile im Ausland geboren	91.9	(1.5)	68.3	(2.9)
Polen				
Ein Elternteil im Ausland geboren	27.5	(5.5)	43.2	(5.1)
Beide Elternteile im Ausland geboren	72.5	(5.5)	56.8	(5.1)
Anderes Land				
Ein Elternteil im Ausland geboren	49.8	(3.2)	50.9	(2.6)
Beide Elternteile im Ausland geboren	50.2	(3.2)	49.1	(2.6)

Anmerkung: Daten beruhen auf Schülerangaben.

fett: signifikante Unterschiede zwischen 2003 und 2012 ($p < .05$)

* gültige % = Prozentangaben beruhen nur auf Angaben der Schülerinnen und Schüler, die eindeutig zuzuordnen sind.

Merkmale des familiären Hintergrundes von Jugendlichen mit Zuwanderungshintergrund

Die Analysen zum familiären Hintergrund von Jugendlichen aus Zuwandererfamilien in Deutschland werden in den folgenden Abschnitten analog zum internationalen Vergleich in Abschnitt 9.2 dargestellt.

In Tabelle 9.7 ist der prozentuale Anteil an Jugendlichen mit Zuwanderungshintergrund dargestellt, die zu Hause mit ihrer Familie Deutsch sprechen. Fasst man alle Jugendlichen mit Zuwanderungshintergrund zusammen, so ist der Anteil derer, die zu Hause Deutsch sprechen, mit rund 70 Prozent signifikant höher als in PISA 2003. Am höchsten ist der Anteil derer, die zu Hause Deutsch sprechen, in der Gruppe der Jugendlichen mit nur *einem im Ausland geborenen Elternteil*; dieser beträgt ähnlich wie in PISA 2003 um 90 Prozent, ist jedoch signifikant zurückgegangen. Jugendliche der *zweiten Generation* sprechen 2012 signifikant häufiger zu Hause Deutsch als 2003. Allerdings sank der Anteil an Jugendlichen der *ersten Generation*, die zu Hause am häufigsten Deutsch sprechen, deutlich. Auch im Vergleich der Familien aus verschiedenen Herkunftsländern treten Unterschiede im häuslichen Sprachgebrauch hervor. So zeigt sich, dass lediglich ungefähr 54 Prozent der Jugendlichen mit einem türkischen Zuwanderungshintergrund zu Hause Deutsch sprechen, während dies bei Jugendlichen aus den übrigen Herkunftsländern jeweils mindestens etwa 70 Prozent angeben. Innerhalb der Herkunftslandgruppen ist der Anteil derer, die zu Hause Deutsch sprechen, zwischen 2003 und 2012 zum größten Teil unverändert. Lediglich für die Gruppe der Jugendlichen aus einem anderen Herkunftsland als der Türkei, dem Gebiet der ehemaligen UdSSR und Polen zeigt sich eine Steigerung auf 79,4 Prozent.

Die Anteile der Fünfzehnjährigen mit Zuwanderungshintergrund, die angeben, mit ihrer Familie am häufigsten Deutsch zu sprechen, sind in Tabelle 9.8 nach Herkunftsländern und der Frage, ob *ein* oder *beide* Elternteile im Ausland geboren wurden, aufgeschlüsselt. So sprechen von den Jugendlichen, deren Eltern beide im Gebiet der ehemaligen UdSSR geboren wurden, 69 Prozent zu Hause am häufigsten Deutsch. In PISA 2003 war dieser Anteil noch deutlich geringer (56,4 Prozent). In der Gruppe derjenigen, von denen ein Elternteil im Gebiet der ehemaligen UdSSR geboren wurde, hat sich dieser Anteil gegenüber 2003 nicht verändert. Ein ähnliches Bild zeigt sich für die Jugendlichen aus Zuwandererfamilien, die aus einem anderen Land (außer der ehemaligen UdSSR, der Türkei oder Polen) nach Deutschland gekommen sind. Gegenüber 2003 sprechen 2012 weniger Jugendliche, die einen in der Türkei oder in Polen geborenen Elternteil haben, zu Hause Deutsch. In PISA 2003 waren dies in beiden Gruppen 100 Prozent, während es in PISA 2012 66,6 Prozent (Türkei) beziehungsweise 89,0 Prozent (Polen) sind.

Tabelle 9.7: Prozentuale Anteile fünfzehnjähriger Schülerinnen und Schüler, die zu Hause Deutsch sprechen

Zuwanderungsstatus	2003		2012	
	%	(SE)	%	(SE)
Ohne Zuwanderung	100.0	(0.0)	99.4	(0.4)
Mit Zuwanderung	64.4	(1.9)	72.1	(1.8)
Generationsstatus				
Ein Elternteil im Ausland geboren	94.1	(1.9)	89.1	(1.7)
Zweite Generation	54.2	(4.5)	65.8	(2.1)
Erste Generation	52.1	(3.8)	35.5	(5.3)
Nicht zuzuordnen	73.8	(2.9)	84.5	(6.0)
Herkunftsländer				
Ehem. UdSSR	58.7	(4.2)	69.8	(4.3)
Türkei	51.4	(5.5)	53.7	(4.6)
Polen	78.6	(4.7)	72.8	(4.6)
Anderes Herkunftsland	70.2	(2.9)	79.4	(2.0)

Anmerkung: Daten beruhen auf Schülerangaben.

fett: signifikante Unterschiede zwischen 2003 und 2012 ($p < .05$)

Tabelle 9.8: Prozentuale Anteile fünfzehnjähriger Schülerinnen und Schüler mit Zuwanderungshintergrund, die zu Hause Deutsch sprechen, nach Herkunft und Geburtsland der Eltern

Herkunftsländer	2003		2012	
	%	(SE)	%	(SE)
Ehemalige UdSSR				
Ein Elternteil im Ausland geboren	76.8	(8.8)	82.6	(8.5)
Beide Elternteile im Ausland geboren	56.4	(4.4)	69.0	(4.4)
Türkei				
Ein Elternteil im Ausland geboren	100.0	(0.0)	66.6	(6.5)
Beide Elternteile im Ausland geboren	46.7	(5.7)	46.6	(5.0)
Polen				
Ein Elternteil im Ausland geboren	100.0	(0.0)	89.0	(5.2)
Beide Elternteile im Ausland geboren	70.6	(6.8)	59.6	(6.7)
Anderes Land				
Ein Elternteil im Ausland geboren	94.4	(2.4)	94.8	(1.3)
Beide Elternteile im Ausland geboren	43.7	(4.4)	59.9	(3.2)

Anmerkung: Daten beruhen auf Schülerangaben.

fett: signifikante Unterschiede zwischen 2003 und 2012 ($p < .05$)

Um Disparitäten in den Lebens- und Lernumwelten zwischen Schülerinnen und Schülern ohne und mit Zuwanderungshintergrund in Deutschland zu beschreiben, werden in Anknüpfung an die bisherige Berichterstattung drei Indikatoren der sozialen Herkunft angegeben (Tabelle 9.9). Diese drei Indikatoren sind der sozioökonomische Status (*Highest International Socio-Economic Index of Occupational Status* [HISEI]), die Verfügbarkeit von Kulturgütern und das Bildungsniveau der Eltern in Ausbildungsjahren (OECD, 2012b).

Tabelle 9.9: Unterschiede in der sozialen Herkunft zwischen Jugendlichen mit und Jugendlichen ohne Zuwanderungshintergrund

Zuwanderungsstatus	Sozioökonomischer Status		Kulturelle Ressourcen ¹				Bildungsniveau der Eltern			
	2012		2003		2012		2003		2012	
	M	(SE)	M	(SE)	M	(SE)	M	(SE)	M	(SE)
Ohne Zuwanderung	54.3	(0.5)	0.1	(0.0)	0.1	(0.0)	14.8	(0.1)	14.4	(0.1)
Mit Zuwanderung	43.8 ^a	(0.7)	-0.1 ^a	(0.0)	-0.1 ^a	(0.0)	11.5 ^a	(0.2)	13.5^a	(0.1)
Generationsstatus										
Ein Elternteil im Ausland geboren	49.8 ^a	(1.2)	0.1	(0.1)	0.0	(0.1)	14.7	(0.2)	13.9^a	(0.2)
Zweite Generation	38.9 ^a	(0.8)	-0.2 ^a	(0.1)	-0.3 ^a	(0.0)	9.8 ^a	(0.3)	12.9^a	(0.2)
Erste Generation	43.2 ^a	(1.9)	-0.1	(0.1)	0.0 ^a	(0.1)	10.8	(0.4)	14.0^a	(0.4)
Nicht zuzuordnen	48.0 ^a	(1.1)	-0.2 ^a	(0.1)	-0.1 ^a	(0.1)	12.3 ^a	(0.4)	13.4 ^a	(0.5)
Herkunftsländer										
Ehem. UdSSR	40.9 ^a	(1.6)	0.0	(0.1)	0.0	(0.1)	11.5	(0.6)	13.9^a	(0.3)
Türkei	36.8 ^a	(1.2)	0.2 ^a	(0.1)	-0.3 ^a	(0.1)	8.9 ^a	(0.4)	11.7^a	(0.3)
Polen	48.1 ^a	(1.8)	-0.2	(0.1)	-0.1	(0.1)	13.9	(0.4)	14.2 ^a	(0.3)
Anderes Herkunftsland	46.5 ^a	(0.9)	-0.1 ^a	(0.1)	-0.1	(0.0)	12.1 ^a	(0.4)	13.8^a	(0.1)

Anmerkung: Die Angaben beruhen auf Schülerangaben.

fett: signifikante Unterschiede zwischen 2003 und 2012 ($p < .05$)

^a signifikante Unterschiede zu Schülerinnen und Schülern ohne Zuwanderungshintergrund ($p < .05$)

¹ z-standardisiert

Da die Vergleichbarkeit des sozioökonomischen Status zwischen PISA 2003 und PISA 2012 durch die Überarbeitung der Berufsklassifizierung nicht eindeutig ist, wurde auf eine Analyse der Veränderung im HISEI seit PISA 2003 verzichtet. In PISA 2012 zeigen sich in Bezug auf den sozioökonomischen Status der Familien deutliche Disparitäten zwischen den Familien mit und ohne Zuwanderungshintergrund. Jugendliche mit einem im Ausland geborenen Elternteil, Jugendliche der ersten Generation und Jugendliche der zweiten Generation, sie alle wachsen im Mittel unter signifikant schlechteren sozioökonomischen Rahmenbedingungen auf als Jugendliche ohne Zuwanderungshintergrund. Dies betrifft in besonderem Maße Schülerinnen und Schüler der zweiten Generation.

Betrachtet man die Herkunftsländer der Familien, so zeigen sich ebenfalls für alle Herkunftsländer erhebliche Disparitäten im sozioökonomischen Status. Insbesondere Familien aus der Türkei verfügen über erheblich geringere sozioökonomische Ressourcen als Familien ohne Zuwanderungsgeschichte.

Der Indikator für die kulturellen Ressourcen beruht auf Schülerangaben zum Besitz von klassischer Literatur, Gedichtbänden, Kunstwerken und Ähnlichem. Auf Basis der PISA 2012 Erhebung ist festzustellen, dass sich Familien von Fünfzehnjährigen *mit* und Fünfzehnjährigen *ohne* Zuwanderungshintergrund hinsichtlich ihrer kulturellen Ressourcen voneinander unterscheiden: Schülerinnen und Schüler der ersten und zweiten Generation leben in Familien, die über erheblich niedrigere kulturelle Ressourcen verfügen als Schülerinnen und Schüler ohne Zuwanderungshintergrund. Die kulturellen Ressourcen von Schülerinnen und Schülern ohne und mit Zuwanderungshintergrund sind gegenüber PISA 2003 unverändert. Betrachtet man die Schülerschaft getrennt nach Herkunftsland, so sind es in PISA 2003 und 2012 die Schülerinnen und Schüler mit aus der Türkei stammenden Familien, deren kulturelle Ressourcen zu Hause vergleichsweise gering sind.

Das *Bildungsniveau der Eltern* wurde anhand der Angaben von Vätern und Müttern zum höchsten jeweils erreichten Schulabschluss ermittelt. Bei unterschiedlichen Schulabschlüssen von Vater und Mutter wurde für den Indikator des Bildungsniveaus der jeweils höhere Wert von beiden herangezogen. Das Bildungsniveau der Eltern von Fünfzehnjährigen mit Zuwanderungshintergrund ist zum Zeitpunkt der PISA-2012-Erhebung signifikant niedriger als bei ihren Mitschülerinnen und Mitschülern ohne Zuwanderungshintergrund. Allerdings war das Bildungsniveau der Eltern von Jugendlichen *mit Zuwanderungshintergrund* seit der PISA-Erhebung 2003 insgesamt signifikant gestiegen. Dagegen hat sich das Bildungsniveau der Eltern von Jugendlichen *ohne Zuwanderungshintergrund* kaum verändert. Somit verkleinern sich die Disparitäten im Bildungsniveau der Eltern. Betrachtet man die Schülerinnen und Schüler mit Zuwanderungshintergrund getrennt nach Generationsstatus, so fällt auf, dass das Bildungsniveau der Zuwandererfamilien sowohl bei Jugendlichen der ersten als auch der zweiten Generation in der PISA-Erhebung 2012 höher ist als in der PISA-Erhebung 2003. In den Familien mit einem im Ausland geborenen Elternteil ist das Bildungsniveau zwischen den PISA Erhebungen 2003 und 2012 dagegen signifikant abgesunken. Nimmt man das Herkunftsland der Familien von Schülerinnen und Schülern mit Zuwanderungshintergrund in den Blick, so ist gegenüber PISA 2003 das Bildungsniveau der Eltern von Jugendlichen aus dem Gebiet der ehemaligen UdSSR, der Türkei sowie aus der Gruppe der anderen Herkunftsländer signifikant höher.

Dieser Vergleich der sozialen Herkunft von Jugendlichen *mit* und *ohne* Zuwanderungshintergrund kann durch eine getrennte Betrachtung der Schülerinnen und Schüler nach Herkunftsland weiter differenziert werden (Tabelle 9.10).

Tabelle 9.10: Unterschiede in der sozialen Herkunft zwischen Jugendlichen ohne und Jugendlichen mit Zuwanderungshintergrund nach Herkunftsland und Geburtsland der Eltern

Herkunftsländer	Sozioökonomischer Status		kulturelle Ressourcen ¹				Bildungsniveau der Eltern			
	2012		2003		2012		2003		2012	
	M	(SE)	M	(SE)	M	(SE)	M	(SE)	M	(SE)
Ohne Zuwanderungshintergrund	54.3	(0.5)	0.1	(0.0)	0.1	(0.0)	14.8	(0.1)	14.4	(0.1)
Ehemalige UdSSR										
Ein Elternteil im Ausland geboren	46.5	(5.5)	0.1	(0.2)	0.0	(0.2)	15.1	(0.9)	14.3	(0.7)
Beide Elternteile im Ausland geboren	40.4 ^a	(1.7)	0.0	(0.1)	-0.1	(0.1)	11.2	(0.6)	13.9^a	(0.3)
Türkei										
Ein Elternteil im Ausland geboren	43.4 ^a	(2.5)	0.0	(0.3)	-0.2	(0.1)	12.9 ^a	(0.8)	12.6 ^a	(0.5)
Beide Elternteile im Ausland geboren	33.9 ^a	(1.4)	-0.2 ^a	(0.1)	-0.4 ^a	(0.1)	8.7 ^a	(0.4)	11.4^a	(0.4)
Polen										
Ein Elternteil im Ausland geboren	50.7	(2.8)	0.0	(0.2)	-0.1	(0.2)	14.6	(0.4)	14.1	(0.6)
Beide Elternteile im Ausland geboren	46.0 ^a	(2.5)	-0.2	(0.1)	-0.1	(0.1)	13.7	(0.5)	14.3 ^a	(0.3)
Anderes Land										
Ein Elternteil im Ausland geboren	51.5 ^a	(1.4)	0.1	(0.1)	0.0	(0.1)	14.8	(0.3)	14.2	(0.2)
Beide Elternteile im Ausland geboren	41.4 ^a	(1.2)	-0.2 ^a	(0.1)	-0.2 ^a	(0.1)	9.4	(0.6)	13.4^a	(0.3)

Anmerkung: Daten beruhen auf Schülerangaben.

fett: signifikante Unterschiede zwischen 2003 und 2012 ($p < .05$)

^a signifikante Unterschiede zu Schülerinnen und Schülern ohne Zuwanderungshintergrund ($p < .05$)

¹ z-standardisiert

In der nach Geburtsland der Eltern und Herkunftsland differenzierenden Betrachtung lässt sich feststellen, dass die Familien aus dem Gebiet der ehemaligen UdSSR und Polen nur dann einen signifikant niedrigeren sozioökonomischen Status haben als Familien ohne Zuwanderungsgeschichte, wenn beide Elternteile im Ausland geboren wurden. Ein Vergleich der Jugendlichen mit Familien aus der Türkei und einem anderen Herkunftsland als Polen oder der ehemaligen UdSSR mit Jugendlichen ohne Zuwanderungsgeschichte ergibt ein etwas anderes Bild. Hier zeigt sich, dass sowohl Familien mit zwei als auch mit einem in der Türkei oder in einem anderen Herkunftsland geborenen Elternteil einen signifikant niedrigeren HISEI-Wert haben.

Die kulturellen Ressourcen von Familien, in denen beide Eltern in der Türkei oder beide Eltern in einem anderen Herkunftsland geboren wurden, sind 2012 signifikant geringer als die von Familien ohne Zuwanderungshintergrund.

In Bezug auf die Bildungsjahre der Eltern gilt für alle Herkunftsgruppen: Sind beide Elternteile im Ausland geboren, so verfügen die Eltern der Fünfzehnjährigen mit Zuwanderungshintergrund über signifikant niedrigere Bildungsabschlüsse als die Eltern ihrer Mitschülerinnen und Mitschüler ohne Zuwanderungshintergrund. Nur für die Jugendlichen mit Zuwanderungshintergrund aus der Türkei gilt zusätzlich, dass auch wenn nur ein Elternteil in der Türkei geboren wurde, das Bildungsniveau der Eltern deutlich geringer ist als bei Jugendlichen ohne Zuwanderungshintergrund.

Aus Tabelle 9.10 lässt sich ferner ablesen, dass das Bildungsniveau von Eltern, die beide im Gebiet der ehemaligen UdSSR geboren wurden, gegenüber PISA 2003 signifikant gestiegen ist. Dasselbe gilt für das Herkunftsland Türkei und die Gruppe der anderen Herkunftsländer.

Bildungsbeteiligung von Jugendlichen mit Zuwanderungshintergrund

Im deutschen Sekundarschulsystem ist die Bildungsbeteiligung von Schülerinnen und Schülern an den verschiedenen weiterführenden Schularten für die im späteren Leben erreichbaren beruflichen, sozialen und ökonomischen Positionen von erheblicher Bedeutung (Becker, 2009). Tabelle 9.11 stellt die Bildungsbeteiligung der fünfzehnjährigen Schülerinnen und Schüler *mit* und *ohne* Zuwanderungshintergrund getrennt nach Schulart dar. Ein Vergleich zur entsprechenden Verteilung in PISA 2003 ist aufgrund der sich derzeit stark wandelnden Sekundarschullandschaft nicht sinnvoll (vgl. Kapitel 1). Vergleichbar im Sinne einer Betrachtung von Veränderungen ist allenfalls die Bildungsbeteiligung am Gymnasium, also der prozentuale Anteil der Fünfzehnjährigen, die in verschiedenen PISA-Erhebungsrounds ein Gymnasium besuchen.

Insgesamt ist die Gymnasialbeteiligung in der Gruppe der Schülerinnen und Schüler *ohne* Zuwanderungshintergrund deutlich höher als in der Gruppe der Schülerinnen und Schüler *mit* Zuwanderungshintergrund. In PISA 2003 besuchten 34.6 Prozent der Fünfzehnjährigen ohne Zuwanderungshintergrund in Deutschland ein Gymnasium. Der Anteil der Gymnasiasten aus zugewanderten Familien lag mit 16.4 Prozent damals deutlich niedriger. In PISA 2012 ist die Bildungsbeteiligung am Gymnasium unter den Schülerinnen und Schülern *ohne* Zuwanderungshintergrund auf 40.3 Prozent gestiegen. Auch von den Fünfzehnjährigen *mit Zuwanderungshintergrund* besuchen mittlerweile größere Anteile ein Gymnasium: 33.0 Prozent mit *einem im Ausland geborenen Elternteil*, 23.1 Prozent der *ersten Generation* und 28.4 Prozent der *zweiten Generation*.

Tabelle 9.11: Prozentuale Anteile fünfzehnjähriger Schülerinnen und Schüler mit Zuwanderungshintergrund aufgeteilt auf die Schularten

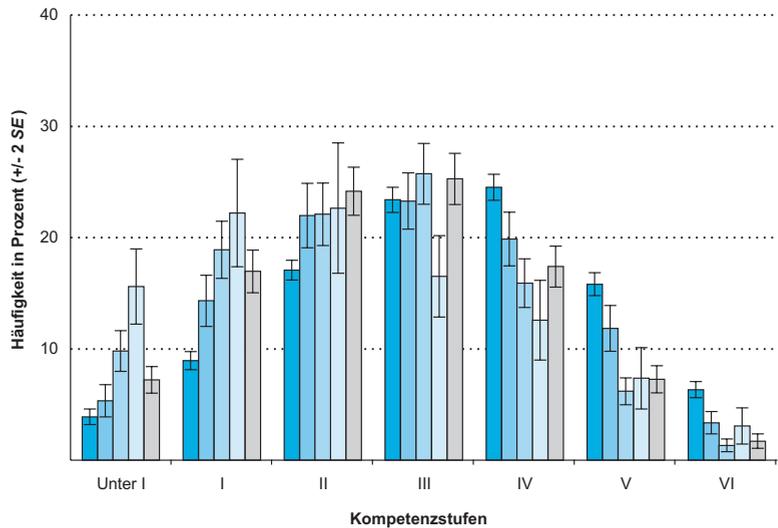
	Gymnasium		Integrierte Gesamtschule		Realschule		Schule mit mehreren Bildungsgängen		Hauptschule	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)
Ohne Zuwanderung	40.3	(2.4)	7.8	(1.1)	24.0	(1.7)	14.4	(1.9)	8.7	(1.1)
Mit Zuwanderung	29.4	(2.3)	11.1	(1.6)	27.0	(2.1)	13.6	(2.2)	14.1	(1.5)
Generationsstatus										
Ein Elternteil im Ausland geboren	33.0	(3.2)	12.9	(2.4)	25.1	(2.8)	14.2	(2.5)	11.2	(2.4)
Zweite Generation	28.4	(2.8)	9.4	(1.6)	30.9	(2.7)	12.0	(2.6)	14.5	(1.7)
Erste Generation	23.1	(3.7)	11.9	(3.1)	19.4	(4.1)	17.2	(4.8)	20.8	(3.5)
Nicht zuzuordnen	29.2	(3.4)	12.4	(2.0)	16.9	(2.0)	19.3	(3.3)	17.1	(2.0)

Verteilungen auf die Stufen mathematischer Kompetenz nach Zuwanderungshintergrund

Die in PISA 2012 erfasste mathematische Kompetenz lässt sich anschaulich anhand von Verteilungen der Jugendlichen auf Kompetenzstufen darstellen. Die inhaltliche Bedeutung dieser Stufen ist in Kapitel 2 näher beschrieben. In Abbildung 9.1 ist die Verteilung der Schülerinnen und Schüler mit und ohne Zuwanderungshintergrund auf die Stufen mathematischer Kompetenz abgebildet.

Betrachtet man die Verteilungen auf die Kompetenzstufen V und VI, so sieht man, dass mehr Schülerinnen und Schüler *ohne* Zuwanderungshintergrund diese Stufen erreichen als Schülerinnen und Schüler *mit* Zuwanderungshintergrund. Für die Schülerinnen und Schüler, die lediglich Kompetenzstufe I erreichen oder sogar die hierzu notwendige Testleistung nicht erbringen, ist die Prognose für ihre weitere Schul- und Ausbildungskarriere kritisch (vgl. Kapitel 2). Davon betroffen sind 38 Prozent der Schülerinnen und Schüler der *ersten Generation*, 29 Prozent der *zweiten Generation*, 20 Prozent der Schülerinnen und Schülern mit *einem im Ausland geborenen Elternteil* und 13 Prozent der Schülerinnen und Schüler *ohne Zuwanderungshintergrund*. Es zeigt sich also, dass Schülerinnen und Schüler mit Zuwanderungsgeschichte immer noch auf der kritischen Kompetenzstufe I und darunter deutlich überrepräsentiert sind. Dieser Befund gilt insbesondere für Jugendliche der ersten Generation. Auf der anderen Seite finden sich in der Spitzengruppe (Kompetenzstufen V und VI) vermehrt Jugendliche mit einer Zuwanderungsgeschichte: Zu dieser Gruppe zählen mehr als zehn Prozent der Jugendlichen aus der ersten Generation, mehr als sieben Prozent der zweiten Generation und mehr als fünfzehn Prozent aus der Gruppe der Jugendlichen mit einem im Ausland geborenen Elternteil.

Kennwerte zu den Kompetenzen von Schülerinnen und Schüler mit Zuwanderungshintergrund in den Domänen Lesen und Naturwissenschaften sind im Anhang (Abbildung 1 und 2) zu finden.



	% (SE)	% (SE)					
Ohne Zuwanderungshintergrund	3.9 (0.7)	8.9 (0.8)	17.1 (0.9)	23.4 (1.1)	24.5 (1.2)	15.8 (1.0)	6.3 (0.7)
Generationsstatus							
Ein Elternteil im Ausland geboren	5.3 (1.4)	14.3 (2.3)	22.0 (2.9)	23.3 (2.5)	19.9 (2.4)	11.8 (2.1)	3.4 (1.0)
Zweite Generation	9.8 (1.8)	18.9 (2.6)	22.1 (2.8)	25.7 (2.7)	15.9 (2.2)	6.2 (1.2)	1.3 (0.6)
Erste Generation	15.6 (3.4)	22.2 (4.8)	22.7 (5.9)	16.5 (3.7)	12.6 (3.6)	7.4 (2.8)	3.1 (1.6)
Nicht zuzuordnen	7.2 (1.2)	17.0 (1.9)	24.2 (2.2)	25.3 (2.3)	17.4 (1.8)	7.3 (1.2)	1.7 (0.6)

Abbildung 9.1: Kompetenzstufen von fünfzehnjährigen Schülerinnen und Schülern mit Zuwanderungshintergrund für Mathematik

Disparitäten in der mathematischen Kompetenz

Mit Hilfe von multivariaten Regressionsanalysen wurde schließlich geprüft, ob die Kompetenzunterschiede in Mathematik zwischen Jugendlichen mit Zuwanderungshintergrund und ihren Mitschülerinnen und Mitschülern ohne Zuwanderungshintergrund auch unter Kontrolle weiterer Hintergrundmerkmale bestehen bleiben. Um die Veränderungen seit der PISA-Erhebung 2003 darstellen zu können, wurden parallele Modelle für PISA 2003 und PISA 2012 spezifiziert. Die Ergebnisse für Schülerinnen und Schüler aus zugewanderten Familien sind differenziert nach Generationsstatus in Tabelle 9.12 dargestellt.

Tabelle 9.12: Regressionsmodelle zur Schätzung von Disparitäten in der Mathematikkompetenz zwischen Jugendlichen ohne Zuwanderungshintergrund und Jugendlichen mit Zuwanderungshintergrund

	Modell I		Modell II		Modell III	
	b	(SE)	b	(SE)	b	(SE)
2012						
Ohne Zuwanderung	535*	(3.4)	531*	(3.1)	531*	(3.1)
Ein Elternteil im Ausland geboren	-28 ^a	(5.9)	-19 ^a	(5.6)	-18 ^a	(5.7)
Zweite Generation	-47^a	(6.5)	-23 ^a	(6.4)	-19 ^a	(6.9)
Erste Generation	-53 ^a	(13.5)	-40 ^a	(12.3)	-33 ^a	(13.5)
Nicht zuzuordnen	-60 ^a	(12.5)	-37 ^a	(12.1)	-35 ^a	(11.4)
HISEI ¹			26 ^a	(2.5)	26 ^a	(2.5)
Kulturgüter ¹			9 ^a	(1.7)	9 ^a	(1.7)
Bildungsniveau der Eltern ¹			10 ^a	(2.1)	10 ^a	(2.1)
Sprachgebrauch ²					-13	(8.5)
N	3761		3761		3761	
R ²	0.03		0.18		0.18	
	Modell I		Modell II		Modell III	
2003	b	(SE)	b	(SE)	b	(SE)
Ohne Zuwanderung	531*	(3.7)	526*	(3.1)	526*	(3.1)
Ein Elternteil im Ausland geboren	-10	(7.2)	-11	(6.7)	-10	(6.7)
Zweite Generation	-82 ^a	(10.7)	-41 ^a	(9.9)	-28 ^a	(9.9)
Erste Generation	-69 ^a	(8.3)	-35 ^a	(7.8)	-21 ^a	(8.6)
Nicht zuzuordnen	-56 ^a	(9.6)	-42 ^a	(8.4)	-33 ^a	(8.6)
HISEI ¹			24 ^a	(1.8)	24 ^a	(1.7)
Kulturgüter ¹			12 ^a	(1.7)	12 ^a	(1.7)
Bildungsniveau der Eltern ¹			14 ^a	(2.0)	13 ^a	(2.0)
Sprachgebrauch ²					-31 ^a	(8.8)
N	3877		3877		3877	
R ²	0.08		0.23		0.24	

Anmerkung: Daten beruhen auf Schülerangaben.

* Die hier angegebene Regressionskonstante ist der geschätzte Mittelwert in der Gruppe der Schülerinnen und Schüler ohne Zuwanderungshintergrund. Für die Schülerinnen und Schüler mit Zuwanderungshintergrund gibt der Regressionskoeffizient jeweils die Abweichung von der Regressionskonstanten an.

fett: Unterschied zum Regressionskoeffizienten für 2003 statistisch signifikant ($p < .05$)

^a signifikante Partialregressionskoeffizienten ($p < .05$)

¹ z-standardisiert

² Referenzgruppe: Deutsch als Familiensprache

In die Regressionsmodelle wurden schrittweise zunächst die Variablen zur Definition des Zuwanderungshintergrundes aufgenommen (Modell I), im zweiten Schritt die drei in Tabelle 9.9 dargestellten Indikatoren des sozialen Hintergrunds (Modell II) und schließlich die zu Hause mit der Familie gesprochene Sprache (Modell III).

Bei einem Vergleich der Regressionen in PISA 2003 und 2012 wird ersichtlich, dass das Gesamtbild sehr ähnlich ist. Es resultieren jedoch einige Unterschiede, die hervorzuheben sich lohnt. In Modell I etwa zeigen sich sowohl in PISA 2003 als auch in PISA 2012 signifikante Unterschiede in der mathematischen Kompetenz zwischen Jugendlichen *mit* Zuwanderungshintergrund und solchen *ohne* Zuwanderungshintergrund. Dies gilt in PISA 2012 für alle Gruppen und in PISA 2003 für die Jugendlichen ohne Zuwanderungshintergrund oder mit einem im Ausland geborenen Elternteil. Allerdings hat sich der Abstand auf der mathematischen Kompetenzskala zwischen den Fünfzehnjährigen ohne Zuwanderungshintergrund¹ und den Schülerinnen und Schülern, die in der zweiten Generation in Deutschland leben, gegenüber PISA 2003 statistisch signifikant verringert (s. o.).

Im zweiten Modell zeigt sich sowohl für PISA 2012 als auch für PISA 2003, dass die drei Indikatoren der sozialen Herkunft signifikante Prädiktoren der mathematischen Kompetenz sind. Je höher also der sozioökonomische Status, der Besitz an Kulturgütern und das Bildungsniveau der Eltern ist, desto höher ist auch die mathematische Kompetenz. Nimmt man diese Indikatoren der sozialen Herkunft in das Regressionsmodell auf, dann sinken die Regressionskoeffizienten der Variablen für die Generationsstatusgruppen deutlich ab. Eine solche Verringerung von Regressionskoeffizienten bei Aufnahme weiterer signifikanter Prädiktoren in das Regressionsmodell kann für das Vorliegen einer (teilweisen) Mediation sprechen. Demnach könnte ein Teil des Zusammenhangs zwischen Zuwanderungsstatus und Mathematikkompetenz durch die soziale Herkunft vermittelt werden.

Ein weiterer bemerkenswerter Befund betrifft die zu Hause gesprochene Sprache im dritten Modell: Während 2003 die zu Hause gesprochene Sprache auch nach Kontrolle des sozioökonomischen Hintergrunds (HISEI, Kulturgüter, Bildungsniveau der Eltern) signifikant zur Vorhersage der Mathematikleistungen beitrug, lässt sich dieser Zusammenhang in PISA 2012 statistisch nicht mehr gegen den Zufall absichern.

1 In Tabelle 9.4 ist der Mittelwert mathematischer Kompetenz für Jugendliche ohne Zuwanderungshintergrund mit 531 Punkten angegeben. In Tabelle 9.12 handelt es sich um geschätzte Regressionsgewichte, wobei der Wert von 535 Punkten das sogenannte Intercept (Abschnitt der y-Achse) ist, an dem sich alle weiteren Regressionsgewichte relativieren. Der Unterschied zwischen beiden Werten ergibt sich aus der für die Regressionen herangezogenen, leicht eingeschränkten Stichprobe; in den Variablen *zu Hause gesprochene Sprache* und *Bildungsniveau der Eltern* gab es fehlende Werte. Die Daten von Schülerinnen und Schülern, zu denen hierzu keine Angaben vorlagen, wurden aus allen Regressionsmodellen ausgeschlossen. Missings im HISEI und in den kulturellen Ressourcen konnten hingegen durch Imputation der fehlenden Werte ausgeglichen werden.

Tabelle 9.13: Regressionsmodelle zur Schätzung von Disparitäten in der Mathematikkompetenz zwischen Jugendlichen ohne Zuwanderungshintergrund und Jugendlichen mit Zuwanderungshintergrund unterschiedlicher Herkunftsgruppen

2012	Modell I		Modell II		Modell III	
	b	(SE)	b	(SE)	b	(SE)
Ohne Zuwanderung	535*	(3.4)	531*	(3.1)	531*	(3.1)
Ehemalige UdSSR						
Ein Elternteil im Ausland geboren	3	(22.3)	10	(19.6)	11	(19.5)
Beide Elternteile im Ausland geboren	-36 ^a	(9.5)	-18 ^a	(8.7)	-14	(9.4)
Türkei						
Ein Elternteil im Ausland geboren	-65 ^a	(14.7)	-42 ^a	(13.1)	-38 ^a	(13.4)
Beide Elternteile im Ausland geboren	-67^a	(9.6)	-29^a	(9.6)	-22 ^a	(11.3)
Polen						
Ein Elternteil im Ausland geboren	-21	(14.2)	-14	(15.4)	-13	(15.4)
Beide Elternteile im Ausland geboren	-17	(14.8)	-8	(14.9)	-3	(14.7)
anderes Herkunftsland						
Ein Elternteil im Ausland geboren	-23 ^a	(6.8)	-17 ^a	(6.5)	-16 ^a	(6.5)
Beide Elternteile im Ausland geboren	-54 ^a	(8.6)	-35 ^a	(8.3)	-31 ^a	(8.9)
nicht zuzuordnen	-60 ^a	(12.5)	-37 ^a	(12.0)	-35 ^a	(11.4)
HISEI ¹			26 ^a	(2.5)	26 ^a	(2.5)
Kulturgüter ¹			9 ^a	(1.7)	9 ^a	(1.7)
Bildungsniveau der Eltern ¹			9 ^a	(2.1)	10 ^a	(2.1)
Sprachgebrauch ²					-12	(8.8)
N	3761		3761		3761	
R ²	0.04		0.18		0.18	
2003	Modell I		Modell II		Modell III	
	b	(SE)	b	(SE)	b	(SE)
Ohne Zuwanderung	531*	(3.7)	526*	(3.1)	526*	(3.1)
Ehemalige UdSSR						
Ein Elternteil im Ausland geboren	-25	(21.6)	-27	(21.8)	-22	(20.6)
Beide Elternteile im Ausland geboren	-60 ^a	(10.1)	-28 ^a	(10.3)	-17	(11.0)
Türkei						
Ein Elternteil im Ausland geboren	-33	(22.5)	-11	(22.0)	-11	(22.0)
Beide Elternteile im Ausland geboren	-112 ^a	(12.2)	-64 ^a	(11.9)	-50 ^a	(12.1)
Polen						
Ein Elternteil im Ausland geboren	-5	(18.1)	-8	(17.7)	-8	(17.7)
Beide Elternteile im Ausland geboren	-30 ^a	(11.5)	-11	(9.3)	-4	(10.2)
anderes Herkunftsland						
Ein Elternteil im Ausland geboren	-7	(8.8)	-10	(8.0)	-9	(8.0)
Beide Elternteile im Ausland geboren	-80 ^a	(10.3)	-43 ^a	(10.4)	-28 ^a	(10.5)
nicht zuzuordnen	-56 ^a	(9.6)	-43 ^a	(8.5)	-35 ^a	(8.8)
HISEI ¹			24 ^a	(1.7)	24 ^a	(1.7)
Kulturgüter ¹			12 ^a	(1.7)	12 ^a	(1.7)
Bildungsniveau der Eltern ¹			13 ^a	(2.0)	12 ^a	(2.0)
Sprachgebrauch ²					-27 ^a	(8.4)
N	3877		3877		3877	
R ²	0.09		0.24		0.24	

Anmerkung: Daten beruhen auf Schülerangaben.

* Die hier angegebene Regressionskonstante ist der geschätzte Mittelwert in der Gruppe der Schülerinnen und Schüler ohne Zuwanderungshintergrund. Für die Schülerinnen und Schüler mit Zuwanderungshintergrund gibt der Regressionskoeffizient jeweils die Abweichung von der Regressionskonstanten an.

fett: Unterschied zum Regressionskoeffizienten für 2003 statistisch signifikant ($p < .05$)

^a signifikante Partialregressionskoeffizienten ($p < .05$)

¹ z-standardisiert

² Referenzgruppe: Deutsch als Familiensprache

Zur Schätzung der Disparitäten zwischen verschiedenen Zuwanderungsgruppen wurden weitere Regressionsanalysen gerechnet. Diese werden in Tabelle 9.13 (analog zu Tabelle 9.12) für jeweils drei gleich aufgebaute Regressionsmodelle und für PISA 2003 und 2012 dargestellt. Um die Leistungen der einzelnen Schülergruppen mit Zuwanderungshintergrund genauer beschreiben zu können, mussten als Prädiktoren die Herkunftsländer mit einem vereinfachten Generationsstatus differenziert werden (analog Tabelle 9.6 und 9.8).

Bezieht man außer den Indikatoren für den Zuwanderungshintergrund keine weiteren Prädiktoren ins Modell ein (Modell I), so ergeben sich unterschiedlich ausgeprägte Disparitäten zwischen den verschiedenen Herkunftsgruppen und den Jugendlichen ohne Zuwanderungshintergrund. Während Jugendliche, deren Mütter *oder* Väter in der ehemaligen UdSSR geboren wurden, über eine vergleichbar hohe Mathematikkompetenz verfügen wie Jugendliche ohne Zuwanderungshintergrund, schneiden Jugendliche, deren Eltern beide in der ehemaligen UdSSR geboren sind, rund 36 Punkte schlechter ab.

Für die Jugendlichen, deren Familien aus der Türkei zugewandert sind, ist eine solche Differenzierung der Disparitäten nach Generationsstatus nicht zu erkennen: Sowohl Schülerinnen und Schüler, deren Eltern beide in der Türkei geboren sind, als auch Jugendliche, bei denen nur ein Elternteil in der Türkei geboren wurde, erzielen deutlich schwächere Leistungen als ihre Mitschülerinnen und Mitschüler ohne Zuwanderungshintergrund.

Jugendliche mit polnischer Herkunft schneiden ähnlich wie Jugendliche ohne Zuwanderungshintergrund ab, unabhängig vom Generationsstatus. Schülerinnen und Schüler, deren Mutter und/oder deren Vater aus einem anderen Land zugewandert sind, weisen ebenfalls eine deutlich niedrigere mathematische Kompetenz auf als Schülerinnen und Schüler ohne Zuwanderungshintergrund. Das explizite Ausweisen der Jugendlichen, die aufgrund fehlender Daten hinsichtlich ihres Migrationsstatus nicht zu bestimmen sind (Kategorie „nicht zuzuordnen“), ermöglicht es, auch deren Mathematikkompetenz mit den anderen Gruppen zu vergleichen. Sie verfügen über ähnlich niedrige Mathematikkompetenzen wie die Schülerinnen und Schüler mit Zuwanderungshintergrund aus der Türkei.

Im Vergleich zu PISA 2003 zeigt sich im Modell I, dass es zwischen den PISA-Erhebungen 2003 und 2012 zu einer signifikanten Verbesserung der Mathematikkompetenzen von Schülerinnen und Schülern kam, deren Eltern beide in der Türkei geboren wurden. Dieser Befund bleibt auch unter Kontrolle des sozialen Hintergrunds bestehen. Erst in Modell III, unter Einbezug der zu Hause gesprochenen Sprache, ist der Unterschied in den Regressionskoeffizienten zwischen 2003 und 2012 nicht mehr signifikant. Eine mögliche Erklärung hierfür könnte sein, dass in Zuwandererfamilien in PISA 2012 häufiger Deutsch zu Hause gesprochen wird.

In Modell II wird der soziale Hintergrund der Fünfzehnjährigen kontrolliert. Hierdurch verringern sich die Disparitäten zu den Schülerinnen und Schülern ohne Zuwanderungshintergrund, bleiben aber bei den Schülerinnen und Schülern türkischer Herkunft, aus dem Gebiet der ehemaligen UdSSR (von denen jeweils beide Elternteile im Gebiet der ehemaligen UdSSR geboren wurden) und aus einem anderen Land signifi-

kant. Die größte Reduzierung der Disparitäten ergibt sich für die Jugendlichen, deren Eltern beide in der Türkei geboren sind. Hier beträgt die Abnahme der Disparitäten mehr als 30 Punkte, in Modell III 28 Punkte. Auch bei den Schülerinnen und Schülern aus einem anderen Land, deren Eltern beide im Ausland geboren wurden, und den nicht zuzuordnenden Schülerinnen und Schülern verringern sich die Disparitäten um mehr als 20 Punkte. Demnach könnte in diesen Gruppen ein Teil des Zusammenhangs zwischen Zuwanderungsstatus und Mathematikkompetenz durch die soziale Herkunft mediiert (vermittelt) werden.

Die Berücksichtigung der in der Familie gesprochenen Sprache als Prädiktor der Mathematikkompetenz in Modell III verändert, wie schon in Tabelle 9.12, das Muster der Regressionskoeffizienten gegenüber Modell II kaum.

9.4 Zusammenfassung und Diskussion

In diesem Kapitel wurden die Lebens- und Lernumwelten sowie die Mathematikkompetenz der Schülerinnen und Schüler *mit* und der Jugendlichen *ohne* Zuwanderungshintergrund beschrieben. In der Tradition der bisherigen nationalen PISA-Berichterstattung wurden dabei sowohl eine internationale wie auch eine auf Deutschland zugeschnittene, vertiefende Perspektive eingenommen.

Im Vergleich zur PISA-Erhebung 2003 ist an allgemeinbildenden Schulen in Deutschland zudem der Anteil der Schülerinnen und Schüler mit Zuwanderungshintergrund signifikant gestiegen; es lässt sich zwar eine Abnahme in den Neuzuwanderungen (ersten Generation) beobachten, gleichzeitig jedoch eine Zunahme der zweiten Generation.

Im Durchschnitt sprechen 72 Prozent der Schülerinnen und Schüler *mit* Zuwanderungshintergrund zu Hause Deutsch. Betrachtet man den Generationsstatus, so gibt es Anzeichen für eine fortschreitende kulturelle Assimilation: Während nur knapp ein Viertel der Jugendlichen der ersten Generation zu Hause Deutsch spricht, ist es in der zweiten Generation bereits die Hälfte und in Familien mit nur einem im Ausland geborenen Elternteil wird sogar überwiegend (89 Prozent) Deutsch gesprochen. Im internationalen Vergleich nähert sich die Häufigkeit der Nutzung der Landessprache in zugewanderten Familien von Jugendlichen der zweiten Generation damit der Häufigkeit in einem „klassischen“ Einwanderungsland an. Es wird also mittlerweile fast so häufig Deutsch zu Hause gesprochen, wie dies in den klassischen Einwanderungsländern (dort: Englisch) der Fall ist. Differenziert man innerhalb Deutschlands weiter nach den größten Herkunftsgruppen, so zeigt sich, dass insbesondere in der Gruppe der Jugendlichen türkischer Herkunft die Nutzung des Deutschen zu Hause weniger verbreitet ist: Lediglich ungefähr 54 Prozent der Schülerinnen und Schüler mit einem türkischen Zuwanderungshintergrund sprechen zu Hause Deutsch, während dies unter Jugendlichen aus den übrigen Herkunftsländern bei jeweils mindestens 60 Prozent der Fall ist. Im Vergleich zu

der PISA-Erhebung 2003 hat die Nutzung des Deutschen zu Hause in Familien mit Herkunftsland Türkei nicht signifikant zugenommen.

Die sozioökonomische Situation von Einwandererfamilien in Deutschland ist nach wie vor ungünstiger als die von Familien ohne Zuwanderungsgeschichte. Im Vergleich zu den meisten anderen Zielländern für Arbeitsmigration und humanitäre Zuwanderung ist der Unterschied zwischen Zuwandererfamilien und Familien ohne Zuwanderungshintergrund hierzulande größer. Betrachtet man darüber hinaus die Generationenfolge, so liegt der sozioökonomische Status von Familien von Jugendlichen der *ersten Generation* höher als der von Familien von Jugendlichen der *zweiten Generation*. Der unterschiedlich hohe sozioökonomische Status ist vor dem Hintergrund einer unterschiedlichen Zusammensetzung der Gruppen von Jugendlichen der ersten und zweiten Generation vorsichtig zu interpretieren.

Neben den sozioökonomischen sind auch die bildungsbezogenen Ressourcen von Familien wichtig für die Entwicklung schulisch relevanter Kompetenzen. Ein Faktor ist das Bildungsniveau der Eltern selbst. Hier zeichnet sich eine positive Entwicklung in Familien mit Zuwanderungshintergrund ab: Das Bildungsniveau der Eltern von Jugendlichen mit Zuwanderungshintergrund war zum Zeitpunkt der PISA-Erhebung 2012 deutlich höher als noch zum Zeitpunkt der PISA-Erhebung 2003. Dies trifft für die erste, aber auch die zweite Generation von Jugendlichen mit Zuwanderungshintergrund zu, was darauf hindeutet, dass neben einer vermehrten Zuwanderung von Höherqualifizierten mittlerweile auch eine bessere Qualifikation von Eltern der in Deutschland geborenen Schülerinnen und Schüler mit Zuwanderungshintergrund vorliegt. Nach Herkunftsland differenziert betrachtet sind es insbesondere die Familien aus dem Gebiet der ehemaligen UdSSR, die mit vergleichsweise hohen Qualifikationen nach Deutschland kommen.

Bei differenzierter Betrachtung lassen sich weitere Hinweise auf eine fortschreitende Integration von Familien mit Zuwanderungshintergrund in Deutschland finden. So hat sich der Anteil der Fünfzehnjährigen, die ein im Ausland und ein in Deutschland geborenes Elternteil haben, seit der PISA-Erhebung im Jahr 2003 beinahe verdoppelt. Diese Zunahme von Ehen oder Partnerschaften zwischen einem Elternteil, der bereits in Deutschland geboren wurde, und einem Ausland geborenen Elternteil ist vor allem auf die Jugendlichen türkischer Herkunft und auf die große Gruppe der nicht näher bestimmten Herkunftsländer zurückzuführen. Eindeutig zu interpretieren ist dieser Befund jedoch nicht, da auch der in Deutschland geborene Elternteil eine familiäre Migrationsgeschichte haben könnte.

Mathematische Kompetenz von Jugendlichen mit Zuwanderungshintergrund

Betrachtet man die Entwicklung der Mathematikkompetenz bei Fünfzehnjährigen in Deutschland seit 2003, so ist eine deutliche Annäherung zwischen Schülerinnen und Schülern mit und ohne Zuwanderungshintergrund zu verzeichnen. Während die Jugendlichen mit Zuwanderungshintergrund in der PISA-Erhebung 2012 deutlich bes-

ser abschneiden als 2003, bleibt die durchschnittliche mathematische Kompetenz der Jugendlichen *ohne* Zuwanderungshintergrund etwa auf dem gleichen Niveau. Somit ist für die Jugendlichen mit Zuwanderungshintergrund ein positiver Trend in der Entwicklung der Mathematikkompetenz festzustellen. Damit bestätigt sich die erfreuliche Entwicklung, die in PISA 2009 für die Lesekompetenz gefunden wurde, auch für Mathematik (Stanat et al., 2010). Ein Großteil dieses Trends ist auf die Jugendlichen der zweiten Generation zurückzuführen, die den deutlich größten Kompetenzzuwachs von 44 Punkten zu verzeichnen hat. Unter Kontrolle der Herkunftsmerkmale in den Regressionsanalysen kann man feststellen, dass dieser Zuwachs vor allem durch die höhere Kompetenz von Jugendlichen mit türkischer Herkunft zu erklären ist.

Im internationalen Vergleich schneiden Schülerinnen und Schüler mit Zuwanderungshintergrund in Deutschland mit 485 Punkten jedoch nicht mehr so schlecht ab: Der OECD-Durchschnitt für alle Schülerinnen und Schüler (also auch der „*native students*“) in Mathematik liegt bei 494 Punkten und damit nur um 9 Punkte höher. Zudem sind die zuwanderungsbezogenen Disparitäten in der Mathematikkompetenz in Deutschland ähnlich ausgeprägt wie in den anderen europäischen Zielländern für Arbeitsmigration und humanitäre Zuwanderung und auch wie in den meisten ehemaligen Kolonialstaaten mit Ausnahme des Vereinigten Königreichs. Dennoch bleibt die Herausforderung bestehen, weiter darauf hinzuwirken, dass Schülerinnen und Schüler mit Zuwanderungshintergrund letztlich eine ähnlich hohe Mathematikkompetenz erreichen wie ihre Mitschülerinnen und Mitschüler ohne Zuwanderungshintergrund.

Auch bei PISA 2012 ist weiterhin festzustellen, dass der Anteil der Schülerinnen und Schüler mit Zuwanderungshintergrund, deren Prognose für einen erfolgreichen Übergang in Ausbildung und Beruf im Hinblick auf die Ausprägung ihrer mathematischen Kompetenz kritisch ist (vgl. Kapitel 2), immer noch viel zu hoch ist. Besonders besorgniserregend ist, dass die mathematische Kompetenz von 29 Prozent der in Deutschland geborenen Jugendlichen, deren Eltern im Ausland geboren wurden (zweite Generation), auf oder sogar unter Kompetenzstufe I liegen. Da eine grundlegende mathematische Kompetenz heute in sehr vielen Berufen notwendig ist, wird der Zugang zu einer erfolgreichen Ausbildung und Berufskarriere erschwert, aber auch die gesellschaftliche Teilhabe beeinträchtigt, wenn ein grundlegendes mathematisches Können und Verständnis fehlen. Die Identifikation und Förderung junger Talente ist in dieser Hinsicht besonders erstrebenswert, denn Deutschland braucht angesichts der aktuellen demographischen Entwicklung in Zukunft vermehrt ausbildungsfähige und selbstständige Schulabgänger.

Literatur

- Alba, R. D. & Nee, V. (2003). *Remaking the mainstream. Assimilation and contemporary immigration*. Cambridge, MA: Harvard University Press.
- Autorengruppe Bildungsberichterstattung. (2012). *Bildung in Deutschland 2012. Ein indikatorengestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf*. Bielefeld: Bertelsmann.
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 402–407). Opladen: Leske + Budrich.
- Baumert, J., Stanat, P., Demmrich, A. (2001). PISA 2000. Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider et al. (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Opladen: Leske + Budrich.
- Becker, R. (2009). Integration von Migranten durch Bildung und Ausbildung. Theoretische Erklärungen und empirische Befunde. In R. Becker (Hrsg.), *Lehrbuch der Bildungssoziologie* (1. Aufl., S. 11–36). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Becker, R. & Beck, M. (2009). Migration, Sprachförderung und soziale Integration. Eine Evaluation der Sprachförderung von Berliner Schulkindern mit Migrationshintergrund. In R. Becker (Hrsg.), *Lehrbuch der Bildungssoziologie* (1. Aufl., S. 121–138). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Boswell, C. (2007). Theorizing migration policy: Is there a third way? *International Migration Review*, 41 (1), 75–100.
- Edele, A. & Stanat, P. (2011). PISA's potential for analyses of immigrant students' educational success: The German case. In M. A. Pereyra, H.-G. Kotthoff & R. Cowen (Hrsg.), *PISA under examination: Changing knowledge, changing tests, and changing schools* (S. 185–206). Rotterdam: SensePublishers.
- Ehmke, T., Hohensee, F., Heidemeier, H. & Prenzel, M. (2004). Soziale Herkunft. In M. Prenzel, J. Baumert, B. Werner, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs* (S. 225–282). Münster: Waxmann.
- Ehmke, T., Klieme, E. & Stanat, P. (2013). Veränderungen der Lesekompetenz von PISA 2000 nach PISA 2009. Die Rolle von Unterschieden in den Bildungswegen und in der Zusammensetzung der Schülerschaft. In N. Jude & E. Klieme (Hrsg.), *PISA 2009 – Impulse für die Schul- und Unterrichtsforschung*. 59. Beiheft der Zeitschrift für Pädagogik (S. 132–150). Weinheim: Beltz.
- Ehmke, T., Werner, B., Neubrand, M., Joran, A. & Ulfig, F. (2006). Wie verändert sich die mathematische Kompetenz von der neunten zur zehnten Klassenstufe? In M. Prenzel, J. Baumert, B. Werner, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 63–86). Münster: Waxmann.
- Esser, H. (2001). *Integration und ethnische Schichtung* (Arbeitspapiere des Mannheimer Zentrums für Europäische Sozialforschung Nr. 40). Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.

- Ganzeboom, H. B. G., de Graaf, P. M., Treiman, D. J. & de Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1–56.
- International Labour Office. (1990). *ISCO-88: International Standard Classification of Occupations*. Genf: ILO.
- International Labour Office. (2007). *Resolution concerning updating the International Standard Classification of Occupations*, ILO. Zugriff am 16.09.2013. Verfügbar unter <http://www.ilo.org/public/english/bureau/stat/isco/docs/resol08.pdf>.
- Köller, O., Knigge, M. & Tesch, B. (Hrsg.). (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.
- OECD. (2012a). *International migration outlook 2012*. Paris: OECD.
- OECD. (2012b). *PISA 2009 technical report (PISA)*. Paris: OECD.
- OECD. (2013). *International migration outlook 2013*. Paris: OECD.
- Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T. & Pöhlmann, C. (Hrsg.). (2013). *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.
- Pöhlmann, C., Haag, N. & Stanat, P. (2013). Zuwanderungsbezogene Disparitäten. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 297–330). Münster: Waxmann.
- Prenzel, M., Baumert, J., Werner, B., Lehmann, R., Leutner, D., Neubrand, M. et al. (Hrsg.). (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.
- Ramm, G., Prenzel, M., Heidemeier, H. & Walter, O. (2004). Soziokulturelle Herkunft. Migration. In M. Prenzel, J. Baumert, B. Werner, R. Lehmann, D. Leutner, M. Neubrand et al. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs* (S. 273–282). Münster: Waxmann.
- Segeritz, M., Walter, O. & Stanat, P. (2012). Muster des Schulischen Erfolgs von jugendlichen Migranten in Deutschland. Evidenz für segmentierte Assimilation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62, 113–138.
- Stanat, P., Becker, M., Baumert, J., Lüdtke, O. & Eckhardt, A. G. (2012). Improving second language skills of immigrant students: A field trial study evaluating the effects of a summer learning program. *Learning and Instruction*, 22, 159–170.
- Stanat, P. & Christensen, G. S. (2006). *Where immigrant students succeed. A comparative review of performances and engagement in PISA 2003*. Paris: OECD.
- Stanat, P., Rauch, D. & Segeritz, M. (2010). Schülerinnen und Schüler mit Migrationshintergrund. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 200–228). Münster: Waxmann.
- Statistisches Bundesamt. (2012). *Fachserie 1, Reihe 2.2. Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2011*. Wiesbaden: Statistisches Bundesamt.
- Walter, O. & Taskinen, P. (2008). 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 et al. (Hrsg.), *PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich* (S. 337–366). Münster: Waxmann.

10 Technische Grundlagen des fünften internationalen Vergleichs

Jörg-Henrik Heine, Christine Sälzer, Lars Borchert,
Heiko Sibberns und Julia Mang

In diesem Kapitel werden die wichtigsten methodischen und statistischen Verfahren beschrieben, die PISA 2012 zugrunde liegen.¹ Dabei wird zunächst auf die Stichprobenziehung der Schulen sowie der Schülerinnen und Schüler in Deutschland eingegangen, die sich streng an den Vorgaben des internationalen PISA-Konsortiums orientiert (Abschnitt 10.1). Im Anschluss daran folgen eine Beschreibung der verwendeten statistischen Verfahren zur Modellierung der Kompetenzen (Abschnitt 10.2), der Messgenauigkeit dieser Verfahren (Abschnitt 10.3) sowie eine Erläuterung der Bestimmung von verschiedenen Kompetenzstufen (Abschnitt 10.4). Zuletzt werden in Abschnitt 10.5 die methodischen Grundlagen erklärt, die verwendet werden, um die Ergebnisse zwischen den verschiedenen PISA-Erhebungsrunden miteinander zu vergleichen.

10.1 Die Stichprobe der PISA-Teilnehmer in Deutschland

Wie bei den meisten anderen wissenschaftlichen Schulleistungsstudien ist es auch bei PISA – aus organisatorischen, zeitlichen und finanziellen Gründen – nicht möglich, *alle* Personen zu befragen, die im Fokus des Forschungsinteresses stehen (hier: alle fünfzehnjährigen Schülerinnen und Schüler in Deutschland). Aus diesem Grund ist eine systematische Teilerhebung notwendig. Diese Teilerhebung erfolgt durch die Ziehung einer Zufallsstichprobe, so dass anhand dieser Stichprobe Verallgemeinerungen über die Grundgesamtheit, das heißt über alle fünfzehnjährigen Schülerinnen und Schüler in Deutschland, möglich sind (vgl. Bortz & Schuster, 2010; Häder, 2010; Kish, 1995; Levy & Lemeshaw, 2008; Thompson, 2012).

Schulleistungsstudien, insbesondere *Large-Scale-Assessments*, stellen allerdings besondere Anforderungen an ihre Stichproben. So kommen bei ihnen im Allgemeinen keine einstufigen Zufallsauswahlen zum Einsatz (z. B. Ziehung einzelner Schülerinnen und Schüler in einer Schule), sondern es werden zwei- oder mehrstufige Auswahlverfahren verwendet. Wie in den übrigen an PISA 2012 teilnehmenden Ländern wurde daher auch

1 Eine detaillierte Beschreibung der in den bisherigen PISA-Erhebungsrunden verwendeten Methodologie kann den sogenannten *Technical Reports* entnommen werden. Diese können von der Website der OECD – unter: <http://www.oecd.org/pisa/pisaproducts/> – kostenlos heruntergeladen werden.

in Deutschland ein mehrstufiges Stichprobenverfahren eingesetzt. Bei diesem wurden im ersten Schritt *Schulen* als grundlegende Stichprobeneinheiten gezogen. In einem zweiten Schritt erfolgte dann eine Zufallsauswahl von fünfzehnjährigen *Schülerinnen und Schülern* innerhalb der im ersten Schritt gezogenen Schulen. Wie hierbei im Detail vorgegangen wurde, wird in den folgenden Abschnitten beschrieben. Darüber hinaus folgt eine Beschreibung der tatsächlich realisierten Stichprobe (Abschnitt 10.1.2) sowie eine Erläuterung, wie mit dem Verfahren der Gewichtung die Repräsentativität der erhobenen Daten sichergestellt wird (Abschnitt 10.1.3).

10.1.1 Stichprobenplan und Ziehung der Stichprobe

Im Rahmen von PISA 2012 wurde in allen Teilnehmerstaaten ein zweistufiges Stichprobenziehungsverfahren eingesetzt. Da in Deutschland jedoch ein erweitertes Forschungsdesign zum Einsatz kam, ist es notwendig, dass der Ablauf der Stichprobenziehung für Deutschland an dieser Stelle noch einmal präzisiert wird (vgl. auch Kapitel 1). Hierbei kann das erweiterte Forschungsdesign zunächst folgendermaßen zusammengefasst werden: Ebenso wie in den bisherigen PISA-Erhebungsrunden besteht auch bei PISA 2012 die international vorgegebene Zielpopulation aus allen Schülerinnen und Schülern einer Alterskohorte. Bei dieser handelt es sich um alle Fünfzehnjährigen, die sich in der Klassenstufe 7 oder einer höheren Klassenstufe befinden. Die genaue Altersdefinition wurde hierbei mit dem internationalen PISA-Konsortium abgestimmt. Für Deutschland wurden alle Schülerinnen und Schüler der Alterskohorte der Fünfzehnjährigen zugerechnet, wenn sie zwischen dem 1. Januar und dem 31. Dezember 1996 geboren wurden. Die aus dieser Alterskohorte stammenden Schülerinnen und Schüler sind Teil der PISA-Grundstichprobe, die in jedem an PISA teilnehmenden Staat gezogen wurde.

Darüber hinaus wurde das grundlegende international vorgesehene Forschungsdesign für Deutschland dahingehend erweitert, dass nicht nur die Alterskohorte der Fünfzehnjährigen berücksichtigt wurde, sondern zusätzlich an allgemeinbildenden Schulen zwei vollständige 9. Klassen per Zufall ausgewählt wurden. An ausgewählten Förderschulen konnte sogar eine Vollerhebung aller Neuntklässler durchgeführt werden. Der vorliegende Bericht enthält allerdings ausschließlich Analysen zur PISA-Grundstichprobe der Fünfzehnjährigen, da die Daten zu den zusätzlich gezogenen 9. Klassen zum Zeitpunkt der Berichtlegung noch nicht verfügbar waren.

Außerdem nahm Deutschland, im Hinblick auf weitere Forschungsprojekte des Zentrums für internationale Vergleichsstudien (ZIB), zusätzlich an der computerbasierten Testung (*Computer-Based Assessment: CBA*) teil – eine Option, die den an PISA teilnehmenden Ländern vom internationalen PISA-Konsortium freigestellt wurde.² Zur Auswahl standen die Kompetenzbereiche Lesen, Mathematik und Problemlö-

2 An den beruflichen Schulen wurden lediglich die Fünfzehnjährigen getestet, so dass an diesen Schulen nur eine Erweiterung um die computergestützte Testung (CBA) stattfand.

sen, wobei Deutschland sich für das Problemlösen entschied und auf die computerbasierte Erfassung von Lesen und Mathematik verzichtete. Die Teilnahme Deutschlands an der computerbasierten Testung des Kompetenzbereichs *Problemlösen* erforderte, dass aus der Stichprobe der Fünfzehnjährigen (pro Schule 25 Fünfzehnjährige) noch eine weitere Teilstichprobe von 14 Schülerinnen und Schülern aus den 25 für diese spezielle Testung gezogen wurde. Diese 14 Schülerinnen und Schüler bearbeiteten zusätzlich zu den papierbasierten Materialien am Vormittag dann am Nachmittag einen computerbasierten Test, dessen Ergebnisse Gegenstand eines eigenen Berichtes sein werden.

Berücksichtigt man dieses erweiterte Forschungsdesign für PISA in Deutschland, dann wird ersichtlich, dass sich hieraus ein relativ komplexes mehrstufiges Ziehungsverfahren ergibt: Als erstes wurden – als primäre Stichprobeneinheiten – Schulen nach einem Gruppierungsverfahren gezogen. Da diese Ziehung vom internationalen PISA-Konsortium durchgeführt wurde, musste diesem zuvor eine vollständige Schulliste (*Sampling Frame*) bereitgestellt werden. Nach erfolgter Ziehung der Schulen wurde vom *IEA Data Processing and Research Center* (IEA DPC) mittels der vom internationalen PISA-Konsortium bereitgestellten Software *KeyQuest* die zufallsbasierte Klassenziehung (nur für allgemeinbildende Schulen und Förderschulen) sowie die zufallsbasierte Schülerziehung durchgeführt. Diese Schritte der Stichprobenziehung können der Abbildung 10.1 entnommen werden.

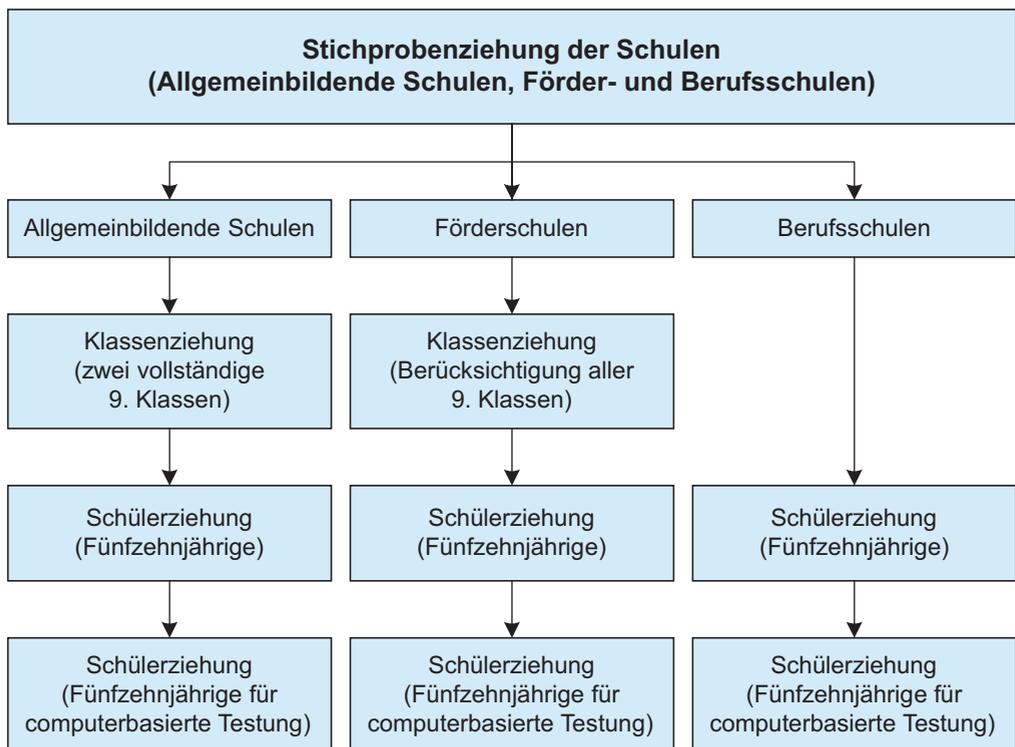


Abbildung 10.1: Schritte der PISA-Stichprobenziehung

Dabei ist zu beachten, dass in der Abbildung ausschließlich die einzelnen Schritte beziehungsweise die Reihenfolge der Stichprobenziehung dargestellt wird – nicht jedoch das genaue *Stichprobendesign*. Das bedeutet, dass die Fünfzehnjährigen, die für die Teilnahme an PISA 2012 gezogen wurden, nicht unbedingt zu einer der gezogenen 9. Klassen gehören müssen; allerdings kann dies umgekehrt durchaus der Fall sein. Die Klassenziehung (nur an allgemeinbildenden Schulen und Förderschulen) ist lediglich der Schülerziehung vorgeschaltet.

In den folgenden Abschnitten werden die Schritte der Schul- und Schülerstichprobenziehung näher betrachtet.

Schritt 1: Vorbereitung der Stichprobenziehung – vollständige Schulliste

Als Grundlage für die Stichprobenziehung der Schulen dient der sogenannte *Sampling Frame* (Stichprobenrahmen). Bei diesem handelt es sich um eine Liste aller Schulen in Deutschland, welche die Zielpopulation der Fünfzehnjährigen (und Neuntklässler) vollständig und korrekt abbildet. Dies bedeutet, dass im *Sampling Frame* gemäß der PISA-Konventionen *alle* Schulen aufgelistet werden müssen, an denen sich in Deutschland potentiell fünfzehnjährige Schülerinnen und Schüler befinden. Schulen, die potentiell 9. Klassen haben, sind nicht extra hinzugefügt, da das Kriterium hierfür diejenige Klassenstufe ist, die von den meisten Fünfzehnjährigen besucht wird (sogenannte *modal grade*, in Deutschland die 9. Jahrgangsstufe). Außerdem dürfen keine Schulen mehrfach, fehlerhaft oder unvollständig gelistet werden. Dabei werden als *Schulen*, gemäß der internationalen PISA-Standards, in der Regel *ganze Schulen* verstanden, weniger dagegen einzelne Schulprogramme oder Bildungsgänge.

Für Deutschland orientiert sich die Auflistung sämtlicher Schulen im *Sampling Frame* an der KMK-Definition von Schule als „schulartspezifische Einrichtung“: Demnach müssen Schulen „keine verwaltungsrechtlich eigenständigen Organisationseinheiten“ bilden (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2013, S. 8). „Vielmehr werden schulartspezifische Schulteile, die im Hinblick auf Lehrpläne, Bildungsziel beziehungsweise Qualifikationsniveau einen eigenständigen Charakter haben, als Schulen gezählt“ (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2013, S. 8). Gemäß dieser Definition ist es beispielsweise möglich, dass eine Schule doppelt im *Sampling Frame* gelistet wurde, wenn an der betreffenden Schule zwei Schularten vorliegen: Zum Beispiel dann, wenn es sich bei einer Schule um eine nominelle Haupt- und Realschule oder um eine kooperative Gesamtschule mit gymnasialer Oberstufe handelt; nicht aber, wenn es um eine Schule mit mehreren Bildungsgängen geht, denn diese zählen als eine einzige Schule mit Klassen, in denen Schülerinnen und Schüler unterschiedlicher Bildungsgänge gemeinsam unterrichtet werden.

Um den *Sampling Frame* nach den Vorgaben des internationalen PISA-Konsortiums erstellen zu können, wurden sämtliche hierfür benötigten Schulinformationen bei den

existierenden 14 Statistischen Landesämtern eingeholt.³ Bei diesen Informationen handelt es sich um Listen der Schulen (allgemeinbildende Schulen, Förderschulen, Berufsschulen) im jeweiligen Land, die folgende Angaben enthalten:

- die offizielle Schulnummer (die später im Sampling Frame pseudonymisiert wurde),
- die Schulart,
- die Anzahl der Schülerinnen und Schülern in den Geburtsjahrgängen 1993, 1994, 1995 und 1996,
- die Anzahl der Schülerinnen und Schüler in den Klassenstufen 8 bis 10,
- die Anzahl der Klassen 8 bis 10,
- die Trägerschaft (öffentlich oder privat),
- Informationen über Veränderungen der Schulart, Schulzusammenlegungen und Schulschließungen sowie
- für Förderschulen die Informationen über die Förderschwerpunkte gemäß der Kultusministerkonferenz (KMK), wobei in Anlehnung an die letzten PISA-Erhebungen die Förderschwerpunkte Lernen, Sprache sowie emotionale und soziale Entwicklung berücksichtigt wurden.

Als Datengrundlage für sämtliche der oben genannten Schulinformationen diente die Schulstatistik des jeweiligen Landes für das Schuljahr 2010/2011. Falls diese für eine oder mehrere Informationen nicht verfügbar war, wurde auf die jeweils aktuell vorhandenen Daten zurückgegriffen.⁴ In den *Sampling Frame* selbst wurden später allerdings nicht alle der bei den Statistischen Landesämtern erfragten Informationen mit aufgenommen. So dienten einige Informationen lediglich der Sicherung der Datenqualität.

Um die Zufallsauswahl treffen zu können, ist es notwendig, die Gesamtanzahl aller fünfzehnjährigen Schülerinnen und Schüler in Deutschland zu kennen. Als Schätzer für die 2012 zu erwartende Anzahl fünfzehnjähriger Schülerinnen und Schüler (PISA-Alterskohorte) – eine der wichtigsten Informationen auf Schulebene, die in den *Sampling Frame* aufgenommen werden musste – wurde die jeweilige Schüleranzahl des Geburtsjahrgangs 1995 verwendet.⁵ Diese Schüleranzahl des Geburtsjahrgangs 1995 wurde dann mit den Schüleranzahlen der Geburtsjahrgänge 1993, 1994 und 1996 abgeglichen, um Schwankungen in den Geburtenzahlen zu identifizieren. Darüber hinaus

3 Es gibt in Deutschland 14 Statistische Landesämter, da für Berlin und Brandenburg sowie für Hamburg und Schleswig-Holstein jeweils ein Landesamt zuständig ist.

4 Dies betrifft lediglich die Datengrundlagen für die Geburtsjahrgänge. Diesbezüglich wurde für Berufsschulen in Baden-Württemberg auf die Schulstatistik 2008/2009 zurückgegriffen und für die allgemeinbildenden und beruflichen Schulen in Nordrhein-Westfalen musste die Schulstatistik 2009/2010 verwendet werden.

5 Der Geburtsjahrgang 1995 wurde als Schätzer für die Anzahl der 2012 erwarteten Fünfzehnjährigen gewählt, da man die Anzahl der Schülerinnen und Schüler an den allgemeinbildenden Schulen überschätzen würde, wenn man auf den Geburtsjahrgang 1996 zurückgreift: Da in einigen Bundesländern viele Fünfzehnjährige die Hauptschule verlassen, um berufsbildende Schulen zu besuchen, dürfen nicht die Vierzehnjährigen als Schätzer für die Fünfzehnjährigen des nächsten Schuljahres genommen werden.

war es aufgrund des erweiterten Forschungsdesigns in Deutschland zusätzlich erforderlich, die Anzahl der 2012 erwarteten Neuntklässler im *Sampling Frame* zu listen. Hierfür wurde die Anzahl der jeweiligen Neuntklässler im Schuljahr 2010/2011 als der diesbezügliche Schätzer verwendet, wobei die Schülerzahlen der Klassenstufen 8 bis 10 sowie die Anzahl der dazugehörigen Klassen überprüft wurden, um mögliche Inkonsistenzen aufzudecken.

Um die Qualität der notwendigen Informationen für den *Sampling Frame* – und damit für die Schulstichprobenziehungen – zu optimieren, wurden sämtliche Schuldaten der Statistischen Landesämter noch einmal mit den Daten des Statistischen Bundesamts (den Daten der Fachserie 11) abgeglichen. Sämtliche Auffälligkeiten wurden dann, in Rücksprache mit dem entsprechenden Statistischen Landesamt geklärt.

Schritt 2: Ziehung der Schulstichprobe

Wie bereits oben ausgeführt, wurde die Schulstichprobenziehung durch das internationale PISA-Konsortium realisiert. Bevor dieses jedoch die Stichprobenziehung – auf Grundlage des übermittelten *Sampling Frames* – durchführen konnte, war es notwendig, die Schulen im *Sampling Frame* nach bestimmten Kriterien, den sogenannten Stratifizierungsvariablen, zu ordnen beziehungsweise zu schichten (vgl. Daniel, 2012; Kalton, 1983; Kish, 1995; Levy & Lemeshaw, 2008). Dies bedeutet, dass mittels der Stratifizierung sämtliche der gelisteten Schulen (gemäß ihrer Ähnlichkeit) in Gruppen aufgeteilt wurden, was (bei PISA) unter anderem folgenden Zwecken dient: Zunächst wird die Effizienz des *Sampling Designs* verbessert, wodurch die Schätzer für die unbekannt Parameter (z. B. Kompetenzen) in der Grundgesamtheit (fünfzehnjährige Schülerinnen und Schüler in Deutschland) verlässlicher werden. Das bedeutet, dass man mit einer möglichst geringen Stichprobengröße möglichst genaue Populationsschätzungen vornehmen kann. Darüber hinaus ist eine Stratifizierung notwendig, um verschiedene *Sampling-Prozeduren* durchführen zu können, beispielsweise ein *Oversampling* einzelner Schularten. Ein solches *Oversampling* könnte beispielsweise beinhalten, dass in einem Teilnehmerstaat mehr Gymnasien für die Teilnahme an PISA ausgewählt werden, als es ihrem Anteil an der Schulpopulation in diesem Staat entspricht. Diese erweiterte Stichprobe ermöglicht vertiefende wissenschaftliche Analysen über die fokussierte Zielgruppe (wie eben z. B. Gymnasien). Ein weiterer Grund für eine Stratifizierung ist eine bessere Kontrolle darüber, dass für alle Strata der Grundgesamtheit auch in der Stichprobe Beobachtungen/Fälle vorliegen, dass also die Verteilung der Schulen innerhalb der Strata der Stichprobe die Verteilung der Strata in der Grundgesamtheit widerspiegelt und somit zugleich spezifische Teilpopulationen (z. B. bestimmte Schularten) in der Stichprobe in adäquater Weise repräsentiert werden.

Die Stratifizierung selbst kann explizit und implizit erfolgen. Bei der expliziten Stratifizierung werden die Schulen in einzelne Gruppen beziehungsweise Strata aufgeteilt, die unabhängig voneinander behandelt werden und aus denen jeweils eine separate Zufallsauswahl getroffen wird. Die implizite Stratifizierung dient dazu, die Schulen noch ein-

mal innerhalb der expliziten Strata zu sortieren, um eine näherungsweise proportionale Verteilung der Schulen über sämtliche Strata innerhalb der Stichprobe zu gewährleisten.

Für Deutschland kamen bei PISA 2012 zwei explizite und zwei implizite Stratifizierungsvariablen zum Einsatz:

1. Einteilung der Schulen in allgemeinbildende Schulen, Förderschulen und Berufsschulen (erste explizite Stratifizierungsvariable)
2. Gruppierung der *allgemeinbildenden Schulen* nach Ländern (zweite explizite Stratifizierungsvariable)
3. Gruppierung der *Förder- und Berufsschulen* nach Ländern (erste implizite Stratifizierungsvariable)
4. Einteilung der *allgemeinbildenden Schulen* in die Schularten Hauptschule, Realschule, Schule mit mehreren Bildungsgängen, Gymnasium und Integrierte Gesamtschule (zweite implizite Stratifizierungsvariable)

Damit lässt sich das Stichprobendesign für Deutschland folgendermaßen beschreiben: Zunächst werden die zu ziehenden allgemeinbildenden Schulen in die 16 Länder eingeteilt (= explizite Stratifizierung). Dann erfolgt die spätere Stichprobenziehung so, dass sich die gezogenen allgemeinbildenden Schulen innerhalb der 16 einzelnen Strata näherungsweise proportional zur Gesamtzahl der Schülerinnen und Schüler je Schulart auf die verschiedenen Schularten verteilen (= *implizite Stratifizierung*). Förder- und Berufsschulen bilden zwei weitere explizite Strata. Das heißt, dass diese beiden Schularten stichprobentechnisch gesondert behandelt werden und daher auch unabhängig von den allgemeinbildenden Schulen gezogen werden. Damit wird die Population der Schulen in Deutschland, die potentiell von fünfzehnjährigen Schülerinnen und Schülern besucht werden, in insgesamt 18 explizite Strata unterteilt.

Durch die implizite Stratifizierung nach Ländern wird dann gesichert, dass sich die zu ziehenden Förder- und Berufsschulen ebenfalls näherungsweise proportional zur Gesamtzahl der Schülerinnen und Schüler dieser beiden Schularten je Land verteilen.

Im Anschluss an die Stratifizierung wurde die Schulstichprobe für Deutschland vom internationalen PISA-Konsortium gezogen. Hierbei ist wichtig, dass die Verteilung der Gesamtzahl der gezogenen Schulen auf die einzelnen Strata so erfolgt, dass der erwartete Anteil der Schülerinnen und Schüler in jedem dieser Strata in der Stichprobe näherungsweise dem proportionalen Anteil der Schülerinnen und Schüler (PISA-Alterskohorte) in jedem Stratum der Grundgesamtheit entspricht. Insgesamt wurden schließlich vom internationalen PISA-Konsortium 247 Schulen in Deutschland gezogen. Inwieweit sich diese Schulen dabei auf die einzelnen expliziten Strata verteilen, kann Tabelle 10.1 entnommen werden.

Tabelle 10.1: Brutto-Stichprobe der Schulen nach Bundesland und Schultart

	Haupt- schule	Inte- grierte Gesamt- schule	Schule mit mehreren Bildungs- gängen	Real- schule	Gymna- sium	Förder- schule	Berufs- schule	Gesamt
Baden- Württemberg	2	1	5	13	12	2	4	39
Bayern	10	1	0	13	12	1	6	43
Berlin	0	0	4	0	3	0	0	7
Brandenburg	0	0	2	0	2	1	0	5
Bremen	0	0	1	0	1	0	0	2
Hamburg	0	0	2	0	2	0	1	5
Hessen	2	3	0	5	6	1	1	18
Mecklenburg- Vorpommern	0	0	2	0	1	0	0	3
Niedersachsen	4	2	0	9	8	2	1	26
Nordrhein- Westfalen	10	10	0	15	18	3	2	58
Rheinland-Pfalz	0	2	4	1	4	1	0	12
Saarland	0	0	1	0	1	0	1	3
Sachsen	0	0	3	0	3	1	1	8
Sachsen-Anhalt	0	0	2	0	2	0	0	4
Schleswig- Holstein	0	4	1	0	3	1	0	9
Thüringen	0	0	3	0	1	0	1	5
Gesamt	28	23	30	56	79	13	18	247

Anmerkung: Hervorgehobene Schulzahlen entsprechen den expliziten Strata.

Hierbei zeigt sich, dass von den 18 Strata 16 auf die Länder entfallen, während sich jeweils ein weiteres Stratum auf die Förderschulen sowie die Berufsschulen bezieht. Dabei muss mit Nachdruck darauf hingewiesen werden, dass es durchaus möglich ist, dass – insbesondere in kleinen Strata – nicht alle Schularten gezogen werden und somit in der Schulstichprobe vertreten sind. Sämtliche Schulen hatten aber eine von null verschiedene Ziehungswahrscheinlichkeit.⁶

Schritt 3: Ziehung der Schülerstichprobe

Nachdem die Schulen ermittelt wurden, an denen die PISA-Studie durchgeführt werden sollte, wurden mittels einer weiteren Stichprobenziehung die einzelnen Schülerinnen

6 Ausführliche Beispiele zur Schulstichprobenziehung sowie die dazugehörigen Ziehungsalgorithmen können den *PISA Technical Reports* unter <http://www.oecd.org/pisa/pisaproducts/> entnommen werden.

nen und Schüler innerhalb dieser Schulen ausgewählt. Laut Stichprobendesign war es hier vorgesehen, in jeder Schule 25 Fünfzehnjährige per Zufall zu ziehen. Darüber hinaus wurden – wie bereits oben ausgeführt – an allen allgemeinbildenden Schulen zwei vollständige 9. Klassen zufällig ausgewählt, während an den Förderschulen eine Vollerhebung aller Neuntklässler durchgeführt werden konnte. Aufgrund des in Deutschland zusätzlich vorgesehenen CBA wurde außerdem aus der Gruppe der 25 Fünfzehnjährigen eine weitere Teilstichprobe von 14 Schülerinnen und Schülern gezogen, die dann neben der papierbasierten Testung am CBA teilnahmen. Dieses Stichprobendesign ist in Abbildung 10.2 dargestellt.

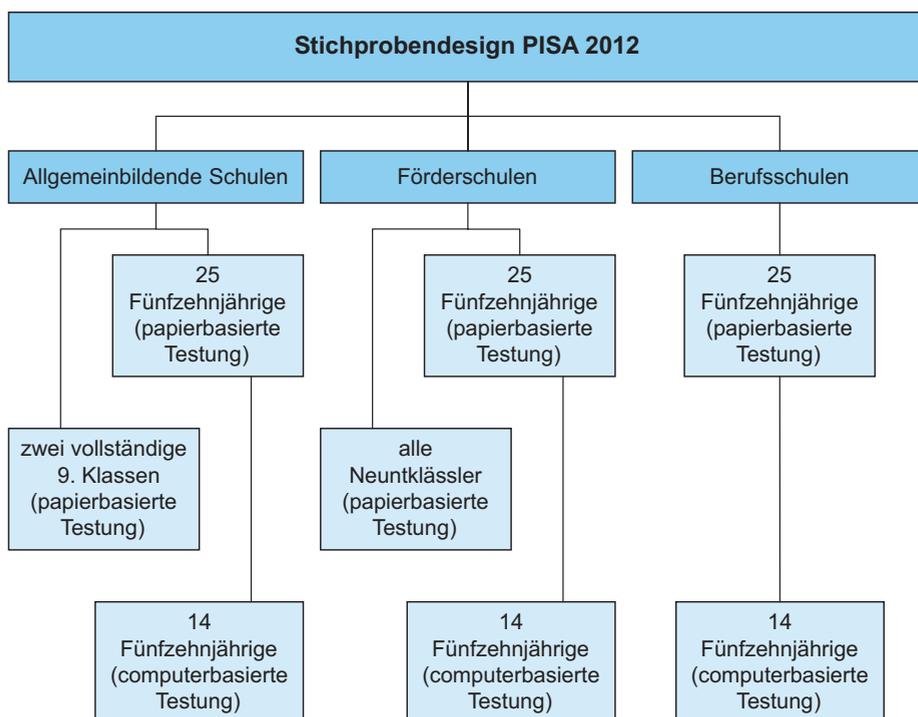


Abbildung 10.2: Stichprobendesign PISA 2012 Deutschland

Dabei sei darauf hingewiesen, dass an den allgemeinbildenden Schulen und Förderschulen von den 25 gezogenen Fünfzehnjährigen (Alterskohorte: Geburtsjahrgang 1996) vereinzelte Schülerinnen und Schüler auch zu einer der gezogenen 9. Klassen gehören können (aber der Testgruppe der Fünfzehnjährigen zugerechnet werden), während Schülerinnen und Schüler, die zu einer gezogenen 9. Klasse gehören, verschiedenen Geburtsjahrgängen angehören können, aber eben auch dem Geburtsjahrgang 1996, womit diese Schülerinnen und Schüler dann ebenfalls fünfzehnjährig sind (aber zur Testgruppe der Neuntklässler gehören). Sobald nämlich 25 Fünfzehnjährige für die Test-

gruppe der Fünfzehnjährigen gezogen waren, wurden die restlichen Fünfzehnjährigen – die zu einer gezogenen 9. Klasse gehören – nur der Testgruppe der entsprechenden 9. Klasse zugeordnet.

Um die beschriebene Schülerstichprobe umsetzen zu können, mussten in den für PISA 2012 gezogenen Schulen einige Informationen zusammengestellt werden. Bei diesen handelt es sich um den Vornamen, den Nachnamen, das Geschlecht, den Geburtsmonat, das Geburtsjahr, die Klassenbezeichnung sowie Informationen über einen möglicherweise vorliegenden sonderpädagogischen Förderbedarf.⁷ Das heißt, die für PISA gezogenen Schulen mussten eine vollständige Schülerliste erstellen, in der alle Schülerinnen und Schüler gelistet wurden, die für die Schülerstichprobe in Frage kamen. Bei diesen handelt es sich um die Leistung aller Fünfzehnjährigen (Geburtsjahrgang 1996) sowie um die Leistung aller Neuntklässler (ohne Geburtsjahrgang 1996).⁸

Für die Durchführung der Schülerleistung wurde den Schulen vom IEA DPC die Software *IEA PISA-W3S* zur Verfügung gestellt. In dieses Programm konnten die benötigten Schülerinformationen entweder manuell eingegeben oder automatisch importiert werden, wenn zuvor eine Dateivorlage mit den benötigten Informationen aus den Datenbanken der jeweiligen Schulverwaltung gefüllt wurde. Ein Vorteil der Schülerleistung mittels der Software *IEA PISA-W3S* besteht darin, dass bereits während der Dateneingabe beziehungsweise nach einem Datenimport sämtliche Schülerangaben automatisch auf Plausibilität überprüft werden können. So wurden fehlerhafte Einträge (z. B. Geburtsjahrgang 1899 oder Klassenstufe 3) in der Schülerliste im Programm farblich hervorgehoben und mussten vom jeweiligen PISA-Schulkoordinator beziehungsweise der jeweiligen PISA-Schulkoordinatorin in den Schulen korrigiert werden. Im Anschluss an die vollständige – und möglicherweise korrigierte – Auflistung aller Schülerinnen und Schüler wurden die Anzahl der gelisteten Fünfzehnjährigen sowie die Anzahl der gelisteten Neuntklässler mit den entsprechenden Schülerzahlen aus den Listen der Statistischen Landesämter abgeglichen. Sollte es dabei vorkommen, dass offensichtliche Abweichungen zwischen den gelisteten Schülerzahlen und den Zahlen der amtlichen Statistik vorlagen, so wurden diese Diskrepanzen in Rücksprache mit den Schulen geklärt. Nach dieser Datenprüfung konnten mit Hilfe von *IEA PISA-W3S* zwei Schülerlisten erstellt werden: Zum einen eine vollständige Schülerliste, die – aus Datenschutzgründen – in der Schule verblieb, zum anderen eine pseudonymisierte Schülerliste, die keine Schülernamen mehr enthielt und die an das IEA DPC für die Durchführung der Schülerstichprobenziehung gesendet wurde.⁹

7 Beim sonderpädagogischen Förderbedarf wird laut internationalen PISA-Standards zwischen Förderbedarf im Bereich körperliche und motorische Entwicklung, Förderbedarf im Bereich geistige, emotionale und soziale Entwicklung sowie Förderbedarf im Bereich Sprache (z. B. Legasthenie) unterschieden. Schülerinnen und Schüler mit einem dieser Förderbedarfe können später möglicherweise von der Testung ausgeschlossen werden.

8 Aus Datenschutzgründen mussten die Namen der Schülerinnen und Schüler in der Schule verbleiben.

9 Die Kommunikation zwischen den Schulen und dem IEA DPC über die für PISA gezogenen Schülerinnen und Schüler verlief dann später nur noch über Ordnungsnummern, welche in beiden Schülerlisten enthalten waren.

Nach Eingang der Schülerlisten beim IEA DPC wurde nochmals eine Qualitätskontrolle durchgeführt, indem die Listen ein weiteres Mal auf Plausibilität und mögliche Unstimmigkeiten überprüft wurden und im Zweifelsfall erneut Rücksprache mit den einzelnen Schulen gehalten wurde. Auf Grundlage der Schülerlisten wurde dann zunächst die Klassenziehung vorbereitet und durchgeführt. Hierzu musste in einem ersten Schritt eine umfassende Klassenliste aller allgemeinbildenden und Förderschulen erstellt werden. Diese enthielt für jede Schule einen Eintrag über alle 9. Klassen (Klassenname und Schüleranzahl) sowie eine Angabe darüber, wie viele Klassen gezogen werden sollten. Dies waren, wie bereits ausgeführt, zwei 9. Klassen an den allgemeinbildenden Schulen sowie alle 9. Klassen an Förderschulen.

In das vom internationalen PISA-Konsortium zur Verfügung gestellte Softwareprogramm *KeyQuest* wurde dann die Klassenliste eingelesen und die Klassenstichprobenziehung durchgeführt. Nach erfolgter Ziehung konnten schließlich die Informationen über die gezogenen Klassen bei der Schülerstichprobenziehung berücksichtigt werden. Hierzu wurden die gezogenen Klassen je Schule in den bereits vorliegenden Schülerlisten vermerkt, die Schülerlisten ebenfalls in *KeyQuest* importiert und zuletzt die Schülerstichprobe für jede einzelne PISA-Schule gezogen.

10.1.2 Realisierte Stichprobe

Wurde bisher auf die Schul- und Schülerstichprobe eingegangen, so handelt es sich bei diesen um die sogenannten Brutto-Stichproben. Das heißt, es sind die Ausgangsstichproben, die zunächst sämtliche für die PISA-Studie ausgewählten Schulen und Schüler enthalten, auch diejenigen, die – aus unterschiedlichen Gründen – nicht an der Studie teilnehmen konnten. Hiervon zu unterscheiden sind die Netto-Stichproben. Diese enthalten schließlich Informationen über alle Schulen sowie Schülerinnen und Schüler, für die auch tatsächlich auswertbare Daten vorliegen.

Auf Schulebene wurden durch die Stichprobenziehung 247 Schulen für die Mitwirkung an der PISA-Studie ermittelt. Allerdings nahmen nicht alle dieser Schulen an PISA teil. So kam es in fünf Fällen vor, dass sich private Schulen gegen eine Teilnahme an PISA entschieden, da diese Schulen nicht zur Teilnahme an der PISA-Studie verpflichtet werden können. Diese konnten allerdings durch entsprechende Ersatzschulen substituiert werden.¹⁰

Darüber hinaus kam es zu 17 Ausfällen von Schulen, bei denen entweder gemäß der internationalen PISA-Standards keine Ersetzung vorgesehen war oder bei denen keine Ersatzschulen mehr zur Verfügung standen. Hierzu gehören 11 Fälle von Berufsschulen, an denen es keine fünfzehnjährigen Schülerinnen und Schüler gab; in zwei Fällen handelte es sich um private Schulen, die nicht an PISA teilnehmen konnten und selbst

¹⁰ Gemäß des internationalen Stichprobendesigns waren für jede der gezogenen 247 Originalschulen zwei Ersatzschulen vorgesehen. In zwei der oben genannten fünf Fälle wurde bereits die erste Ersatzschule ersetzt.

bereits zweite Ersatzschule waren; zwei Schulen waren bereits aufgelöst und existierten nicht mehr; eine private Schule sagte ihre Teilnahme so kurzfristig vor Testbeginn ab, dass keine Ersetzung mehr möglich war und eine dänische Schule, an der lediglich vier Fünfzehnjährige zu erwarten gewesen wären, wurde ohne Ersatz ausgeschlossen, da der zeitliche und organisatorische Aufwand für die Übersetzung sämtlicher deutscher Testmaterialien ins Dänische – verglichen mit der geringen Schüleranzahl – zu groß gewesen wäre. Damit verblieben letztlich 230 Schulen, an denen die PISA-Studie durchgeführt wurde. Die Teilnahmequote auf Schulebene beträgt somit 98,3 Prozent. Diese Zahl bezieht sich auf die ursprünglich gezogenen Schulen und zeigt auf, wie viele dieser Schulen nach eventuellen Ersetzungen tatsächlich an PISA 2012 teilgenommen haben. Bei den Schülerinnen und Schülern beläuft sich die Teilnahmequote auf 93,2 Prozent. Beide Werte erfüllen die von der OECD gesetzten Mindeststandards. Die deutsche PISA-2012-Stichprobe umfasst daher $n = 5001$ Fünfzehnjährige beziehungsweise $n = 9998$ Neuntklässlerinnen und Neuntklässler.

10.1.3 Gewichtung

Für die Analyse der aus PISA gewonnenen Daten muss berücksichtigt werden, dass die Stichprobe in mehreren Stufen gezogen wird und nicht jede gezogene Schule beziehungsweise jede gezogene Schülerin oder jeder gezogene Schüler auch tatsächlich an der Studie teilnehmen kann. Die Ziehung der Stichprobe in mehreren Stufen beeinflusst die Ziehungswahrscheinlichkeit jeder einzelnen Schule sowie jeder einzelnen Schülerin und jedes einzelnen Schülers, so dass nicht jede Schule beziehungsweise jede Schülerin und jeder Schüler mit derselben Wahrscheinlichkeit in die Stichprobe gelangt. Dies wird durch ein Verfahren der Gewichtung ausgeglichen, das in den folgenden Abschnitten beschrieben wird. Neben der unterschiedlichen Ziehungswahrscheinlichkeit gehen fünf Korrekturfaktoren in die Gewichtung ein. Auch diese werden nachfolgend beschrieben.

Schulbasisgewichte

Die Stichprobe für PISA wird in mehreren Stufen gezogen. Der Sampling Frame wird hierzu in 18 explizite Strata unterteilt (16 Bundesländer sowie Förder- und Berufsschulen), die dann nach Schulformen sortiert werden. Nachdem die Schulen gezogen worden sind, erfolgt die zufällige Ziehung der Schülerinnen und Schüler beziehungsweise der 9. Klassen innerhalb der Schulen. Somit haben in PISA weder alle Schulen noch alle Schülerinnen und Schüler dieselbe Ziehungswahrscheinlichkeit für die Teilnahme an der Studie: Die Ziehungswahrscheinlichkeit der Schulen hängt davon ab, wie viele Schulen derselben Schulform sich in einem Stratum befinden. Beispielsweise hat ein Gymnasium in einem Bundesland mit sehr vielen Gymnasien eine geringere Wahrscheinlichkeit, in die PISA-Stichprobe zu gelangen als in Bundesländern mit sehr wenigen Gymnasien. Die Gewichtung wird aus diesem Grund umgekehrt proportional zur Ziehungswahr-

scheinlichkeit vorgenommen, um im Nachhinein eine gleich große Ziehungswahrscheinlichkeit für alle Schulen herzustellen.

Schülerbasisgewichte und Schülergesamtgewicht

Innerhalb gezogener Schulen macht es einen Unterschied, ob es sich um eine große oder eine kleine Schule handelt. In Schulen mit vielen Fünfzehnjährigen hat die einzelne Schülerin oder der einzelne Schüler eine wesentlich geringere Ziehungswahrscheinlichkeit als in Schulen, die nur sehr wenige fünfzehnjährige Schülerinnen und Schüler haben. Dasselbe gilt für die gezogenen 9. Klassen: In einer dreizügigen Schule hat jede 9. Klasse eine Ziehungswahrscheinlichkeit von 33 Prozent, während diese in fünfzügigen Schulen lediglich 20 Prozent beträgt. Um trotz dieser Unterschiede in den Ziehungswahrscheinlichkeiten gültige Aussagen auf der Ebene der Gesamtpopulation der Fünfzehnjährigen machen zu können, müssen die Daten gemäß den unterschiedlichen Ziehungswahrscheinlichkeiten gewichtet werden. Analog zur gestuften Stichprobenziehung erfolgt die Gewichtung schrittweise und teilt jedem Schüler *post hoc* innerhalb ihres Stratum dieselbe Ziehungswahrscheinlichkeit zu. Damit erfolgt die Gewichtung, ebenso wie auf der Ebene der Schulen, umgekehrt proportional zur Ziehungswahrscheinlichkeit. Diese Gewichtung erzeugt ein sogenanntes Schul- beziehungsweise Schülerbasisgewicht. Aus dem Schul- und dem Schülerbasisgewicht wird dann ein *Schülergesamtgewicht* berechnet, welches den Analysen der PISA-Daten zu Grunde gelegt wird.

Notwendigkeit einer Gewichtung

Darüber hinaus kann es weitere Gründe für die Notwendigkeit einer Gewichtung geben. Beispielsweise muss berücksichtigt werden, dass es sowohl auf Schul- als auch auf Schü- lerebene vereinzelt *Ausfälle* gibt. Etwa wenn Schülerinnen und Schüler aufgrund von Krankheit am Testtag abwesend oder wegen eines Ausschlussgrunds nicht zum PISA-Test zugelassen sind, aber auch für ganze Schulen kann es Gründe geben, die eine Teilnahme an PISA unmöglich machen (vgl. Kapitel 1). Diese sogenannten Ausfälle bergen die Gefahr einer verzerrten Schätzung bei der Auswertung der Tests, weil die abwesenden Jugendlichen innerhalb der Stichprobe unterrepräsentiert wären. Für diese Ausfälle geht daher ein Korrekturfaktor in die Gewichtung ein. Ein weiterer Grund für eine nachträgliche Korrektur kann die *Größe einer Schule* sein, wenn diese von der beim *Sampling* erwarteten Schulgröße abweicht (OECD, 2012). Hatte eine Schule bei der Schulstichprobenziehung etwa 50 fünfzehnjährige Schülerinnen und Schüler, so war die Hälfte dieser Fünfzehnjährigen für die PISA-Stichprobe vorgesehen. Zum Zeitpunkt der Schülerziehung stellt sich in diesem Szenario heraus, dass tatsächlich nur noch 40 Fünfzehnjährige an dieser Schule sind. Die Ziehungswahrscheinlichkeit des einzelnen Schü- lers erhöht sich somit von 50 auf 62.5 Prozent, was im Vergleich zu den meisten anderen Schülerinnen und Schülern in der PISA-Stichprobe wesentlich höher ist. Auch die geographische Lage von Schulen oder ein gezieltes *Oversampling* von Schularten kann

ein Grund dafür sein, dass die Ziehungswahrscheinlichkeit von Schulen in einem Staat bewusst nicht gleich ist.

Korrekturfaktoren für die Gewichtung

In die Gewichtung der PISA-Daten gehen neben dem Schul- und dem Schülerbasisgewicht fünf Korrekturfaktoren ein (OECD, 2012). Zunächst muss der Ausfall von Schulen kompensiert werden, indem andere Schulen entsprechend höher gewichtet werden, die den ausgefallenen Schulen so ähnlich wie möglich sind. Ein zweiter Korrekturfaktor betrifft Staaten, in denen nur diejenigen Fünfzehnjährigen an PISA teilnehmen, welche sich in der Klassenstufe befinden, in der die meisten fünfzehnjährigen Schülerinnen und Schüler im jeweiligen Staat zu erwarten sind. Der dritte Korrekturfaktor gleicht die Nicht-Teilnahme von Schülerinnen und Schülern aus, die eigentlich in der Stichprobe gewesen wären. Damit wird vermieden, dass fehlende Schülerinnen und Schüler unterrepräsentiert sind. Auch hier wird auf eine möglichst große Ähnlichkeit der abwesenden (und damit auszugleichenden) Jugendlichen mit denjenigen geachtet, die entsprechend stärker gewichtet werden. Als vierter und fünfter Korrekturfaktor werden zwei sogenannte *Trimming Factors* eingesetzt, die zum einen auf Schulebene ungewöhnlich hohe Werte des Schulbasisgewichts ausgleichen und zum anderen auf Schülerenebene ungewöhnlich hohe Werte beim Produkt aus allen Gewichtungskomponenten zurechtschneiden (OECD, 2012). Für die detaillierte Berechnung der Gewichte und ihrer Komponenten sei auf den *Technical Report* zu PISA verwiesen (z. B. OECD, 2012).

10.2 Skalierung der Leistungstests zur Modellierung von Kompetenzen

Bei der Entwicklung von Messinstrumenten (Tests) für groß angelegte Untersuchungen im bildungswissenschaftlichen Forschungskontext ergibt sich meist ein Konflikt zwischen den psychometrischen Anforderungen aus messtheoretischer Perspektive einerseits und der praktischen Durchführbarkeit von Tests andererseits. Tests dienen in der Regel dazu, bestimmte latente Eigenschaften von Personen zu erfassen, wie etwa Einstellungen, Überzeugungen oder auch Kompetenzen. Dabei ist es seitens der Psychometrie im Allgemeinen wünschenswert, die Erfassung solcher latenten Eigenschaften auf eine möglichst breite Basis zu stellen – also möglichst eine große Anzahl von Fragen beziehungsweise Aufgaben (Items) einzusetzen. Dadurch sollen die latenten Personeneigenschaften (*bei PISA die Kompetenz der Schülerinnen und Schüler in den drei Bereichen Mathematik, Naturwissenschaften und Lesen*) in einem breiten Ausprägungsbereich möglichst genau erfasst werden. Andererseits ist die zur Verfügung stehende Testzeit, nicht zuletzt aus Rücksichtnahme auf die teilnehmenden Schülerinnen und Schüler, begrenzt, wobei aber dennoch meist gleichzeitig mehrere unterschiedliche latente Eigenschaften erfasst werden sollen. Um diesem Dilemma der widersprüchlichen Anforderungen an

die zu entwickelnden Messinstrumente zu begegnen, werden in *Large-Scale-Assessments* zur Gestaltung der Fragebögen üblicherweise sogenannte „*Balanced Incomplete Block Designs*“ (BIBD) zu Grunde gelegt (vgl. z.B. van der Linden, Veldkamp & Carlson, 2004). Bei diesen BIBD handelt es sich letztlich um randomisierte, balancierte, aber unvollständige Versuchspläne beziehungsweise Designs (vgl. Bortz & Schuster, 2010, S. 471ff.). Das grundlegende Prinzip solcher Designs besteht darin, dass nicht jede teilnehmende Person sämtliche zu Verfügung stehenden Items (Fragen, Aufgaben) des Messinstrumentes vorgelegt bekommt (Johnson & Lord, 1958). Aus dieser Grundüberlegung und den resultierenden Anforderungen heraus werden dann mehrere Testhefte entwickelt, welche eine unterschiedliche Auswahl von Aufgabengruppen pro Testheft enthalten. Die nach solchen Prinzipien gestalteten Messinstrumente werden im bildungswissenschaftlichen Forschungskontext dann als Multi-Matrix-Design bezeichnet (Rost, 2004, S. 328; Johnson & Lord, 1958; Lord, 1962; Kapitel 1 dieses Buches). Aufgrund der unvollständigen, aber balancierten Struktur bei gleichzeitig breiter Basis von manifesten Indikatoren beziehungsweise Fragen (Items) solcher *Multi-Matrix-Designs*, eignen sich die damit erhobenen Daten sehr gut zur Bestimmung von Populationsschätzwerten auf Basis der Erhebung großer Stichproben aus der betreffenden Population (Gressard & Loyd, 1991).

In der PISA-Erhebungsrunde 2012 wurden erstmalig nicht nur die Aufgaben zur Erfassung der Kompetenzen in den drei Bereichen (Mathematik, Lesen und Naturwissenschaften), sondern auch einzelne Fragen aus dem Schülerfragebogen (Hintergrundvariablen) auf insgesamt drei unterschiedliche Schülerfragebogenhefte in Form eines Multi-Matrix-Designs verteilt. So war es in dieser PISA-Erhebungsrunde möglich, neben der Erhebung einer Reihe von Hintergrundinformationen, die in allen drei Fragebogenheften enthalten waren, ein breiteres Spektrum von zusätzlichen Schülermerkmalen zu erfassen. Für eine detaillierte Darstellung der einzelnen Frage- beziehungsweise Konstruktbereiche im gemeinsamen Teil des Schülerfragebogens sowie der in Form des Multi-Matrix-Designs erfassten Frage- beziehungsweise Konstruktbereiche sei an dieser Stelle auf die theoretische Rahmenkonzeption zu PISA 2012 verwiesen (OECD, 2013, S. 191ff.).

10.2.1 Das PISA-Multi-Matrix-Design und IRT-Skalierung

Beschreibung der PISA-Items als manifeste Indikatoren kognitiver Kompetenz

In der PISA-Erhebungsrunde 2012 wurden im Rahmen der Erfassung der kognitiven Kompetenz die drei Themenbereiche Mathematik, Lesen und Naturwissenschaften getestet, wobei in dieser PISA-Erhebungsrunde der Bereich Mathematik den Schwerpunkt bildet. Die einzelnen Items (Aufgaben) zu diesen drei Bereichen gruppieren sich dabei um einen gemeinsamen Aufgabenstamm (genannt *Testlet* oder *Unit*), welcher entweder aus Textelementen, Grafiken, Tabellen oder einer Kombination daraus besteht. Die

Anzahl der Aufgaben (Items) pro Unit variiert dabei zwischen einer und sieben. Insgesamt standen für die drei Kompetenzbereiche 87 solcher Units zur Verfügung, was insgesamt 207 einzelnen Aufgaben (Items) entspricht. Diese Units beziehungsweise die darin enthaltenen Items verteilen sich dabei wie folgt auf die drei erfassten Kompetenzbereiche (Tabelle 10.2):

Tabelle 10.2: Anzahl der in PISA 2012 eingesetzten Units und Items nach Domäne

	Units gesamt	Items gesamt	Link-Units	Link-Items
Mathematik	56	110	25	36
Lesen	13	44	13	44
Naturwissenschaften	18	53	18	53
Gesamt	87	207	56	133

Die sogenannten *Link-Units* sind Aufgaben, die aus vorangegangenen PISA-Erhebungsrunden unverändert übernommen worden sind. Wie in PISA üblich, kamen dabei in den beiden Nebendomänen (2012: Lesen und Naturwissenschaften) ausschließlich Link-Units zum Einsatz, während für die Schwerpunktdomäne Mathematik mehr als die Hälfte der Aufgaben neu entwickelt worden waren.

Aufbau der Testhefte

Die oben beschriebenen Units für die drei Domänen wurden in einem ersten Schritt insgesamt 15 Aufgabengruppen, sogenannten *Clustern*, zugeordnet. Für den Schwerpunktbereich Mathematik waren es neun Cluster (PM1 bis PM5, PM6a und PM6b sowie PM7a und PM7b), wobei Cluster PM6a und PM7a in all jenen Staaten eingesetzt wurden, die in vergangenen PISA-Erhebungsrunden im Kompetenzbereich Mathematik mindestens 480 Punkte erreicht hatten. In allen übrigen Staaten erhielten die Jugendlichen die etwas leichteren Cluster PM6b und PM7b in ihren Testheften, so dass die Reliabilität auch in diesen Staaten durch möglichst der Kompetenz der Schülerschaft entsprechende Aufgaben so hoch wie möglich ist. Für die beiden Bereiche Lesen und Naturwissenschaften waren jeweils drei Cluster (PR1 bis PR3 bzw. PS1 bis PS3) vorgesehen. Jedes Cluster ist dabei so gestaltet, dass seine Bearbeitung in etwa 30 Minuten von den insgesamt 120 Minuten der zur Verfügung stehenden Testzeit in Anspruch nimmt. Für die Haupterhebung wurden die 15 Cluster (neun Mathematik, drei Lesen und drei Naturwissenschaft) auf dreizehn Standardtesthefte verteilt. Jedes der 13 Testhefte enthält dabei vier Cluster, was einer Bearbeitungszeit von ca. 2 Stunden entspricht.

Eine Übersicht über die Zusammensetzung der in Deutschland eingesetzten Testhefte gibt Tabelle 10.3.

Tabelle 10.3: Testheftdesign von PISA 2012

Testheft													
Dauer	1	2	3	4	5	6	7	8	9	10	11	12	13
30	M5	N3	L3	M6	M7	M1	M2	N2	L2	M3	M4	N1	L1
30	N3	L3	M6	M7	N1	M2	N2	L2	M3	M4	M5	L1	M1
30	M6	M7	N1	L1	M1	L1	M3	M4	M5	N3	L3	M2	N2
30	N2	L2	M3	M4	M5	M6	M7	N1	L1	M1	M2	N3	L3
45	Fragebogen												

Anmerkung: M = Mathematik, N = Naturwissenschaften, L = Lesen.

Vergleichbarkeit der Schülerleistungen untereinander

Wie im vorangegangenen Abschnitt beschrieben, erhält jede einzelne Schülerin und jeder einzelne Schüler im Rahmen der PISA-2012-Erhebung ein unterschiedliches Testheft, welches lediglich eine unterschiedliche Auswahl aus dem gesamten zur Verfügung stehenden Aufgabenpool enthält. Obwohl bei der Auswahl der Aufgaben und der Zusammenstellung der Testhefte weitestgehend darauf geachtet wird, dass die Aufgabenschwierigkeiten über die Testhefte hinweg gleichmäßig verteilt sind, sind die Testhefte beziehungsweise die darin enthaltenen Aufgaben dennoch unterschiedlich schwierig. Dies bedeutet, dass für einen Vergleich der Kompetenz zwischen einzelnen Schülerinnen oder Schülern, oder auch zwischen unterschiedlichen Gruppen, nicht einfach die reine Anzahl der gelösten beziehungsweise nicht gelösten Aufgaben als Vergleichswert herangezogen werden kann. Es ist, auf Basis der Unterschiedlichkeit der Testhefte, unmittelbar einsichtig, dass für einen solchen Vergleich, soll er fair und objektiv sein, auch die jeweilige Schwierigkeit der individuell unterschiedlich vorgelegten Aufgaben berücksichtigt werden muss. Genau ein solches Einbeziehen der Aufgabenschwierigkeit bei der Berechnung beziehungsweise Schätzung eines individuellen Kompetenzwertes wird durch die bei PISA verwendete Methode der Skalierung (vgl. Abschnitt zum Rasch-Modell) gewährleistet.

Latente (kontinuierliche) Personenvariable

Im Gegensatz zu einem einfachen Aufaddieren der gelösten Aufgaben basiert bei den PISA-Untersuchungen die Auswertung auf Antwortmodellen aus der *Item Response Theory* (IRT) (vgl. einführend Fischer & Molenaar, 1995; Rost, 2004). Die psychologisch plausible Kernannahme solcher Antwortmodelle besteht dabei in der Annahme einer latenten, also einer nicht direkt beobachtbaren, individuellen Personeneigenschaft (Personenvariable), welche ursächlich mit dem manifest zu beobachtenden Antwortverhalten einer jeden einzelnen Person verbunden ist. Ganz konkret besteht am Beispiel der

Domäne Mathematik die Annahme, dass die Ausprägung einer Schülerin oder eines Schülers auf der latenten Variable „Mathematikkompetenz“ dafür verantwortlich ist, wie wahrscheinlich es ist, dass er oder sie eine Aufgabe mit bestimmter Schwierigkeit löst. Eine einfache Aufgabe wird von sehr fähigen Schülerinnen oder Schülern sehr wahrscheinlich gelöst, während eine schwierige Aufgabe von weniger fähigen Schülerinnen oder Schülern nur sehr unwahrscheinlich gelöst wird. Ist die Schülerin oder der Schüler genauso fähig, wie die Aufgabe schwierig ist, beträgt die Lösungswahrscheinlichkeit nach diesem Modell 50 Prozent. Diesen Zusammenhang zwischen der Ausprägung auf der angenommenen latenten Variablen und der manifest beobachteten Lösung beziehungsweise Nicht-Lösung von Aufgaben (unterschiedlicher Schwierigkeit) lässt sich im Rahmen der IRT mit dem *Rasch-Modell* (Rasch, 1960) und dessen Erweiterung, dem *Partial-Credit-Modell* (Masters, 1982), mathematisch-formal modellieren. Die Grundzüge dieser beiden Modelle werden in den folgenden beiden Abschnitten kurz erläutert.

Das Rasch-Modell

Das von dem dänischen Mathematiker Georg Rasch (1960) beschriebene Modell, welches die Grundlage der psychometrischen Auswertung der PISA-Untersuchungen bildet, basiert zunächst auf der Annahme zweier grundlegender Parameter, welche bei der Analyse und Auswertung von psychologischen Messinstrumenten (Fragebögen oder Tests), die eine bestimmte latente Personeneigenschaft beziehungsweise -kompetenz erfassen, berücksichtigt werden müssen. Diese beiden Parameter sind einerseits die Fähigkeit beziehungsweise die Ausprägung der Person auf der latenten Eigenschaft – der *Personenparameter* – und andererseits die Schwierigkeit der Aufgabe – der *Itemparameter*. Das Besondere an dem Modell von Rasch ist dabei, dass es neben der Annahme dieser beiden Parameter eine psychologisch plausible, stochastisch-probabilistische Beziehung der beiden Modellparameter und der Wahrscheinlichkeit zur (richtigen) Lösung einer gegebenen Testaufgabe (Item) beschreibt. Diese Beziehung ist in dem Modell als logistische Funktion formalisiert. Eine graphische Veranschaulichung dieser stochastischen Beziehung beziehungsweise Funktion, stellt die *Item Characteristic Curve*, kurz ICC, dar. Dabei nimmt mit einer zunehmend positiven Differenz der beiden Modellparameter (Personenparameter – Itemparameter) die Wahrscheinlichkeit der Lösung eines betreffenden Items oberhalb der 50-Prozent-Lösungswahrscheinlichkeit zu. Mit einer zunehmend negativen Differenz der beiden Modellparameter nimmt dagegen die Wahrscheinlichkeit der Lösung eines betreffenden Items unterhalb der 50-Prozent-Lösungswahrscheinlichkeit ab (vgl. Abbildung 10.3). Anders ausgedrückt bedeutet dies: Wenn der Personenparameter größer ist als der Itemparameter, dann ist die Lösungswahrscheinlichkeit für das betreffende Item größer als 50 Prozent; wenn der Itemparameter größer ist als der Personenparameter, dann ist sie niedriger als 50 Prozent.

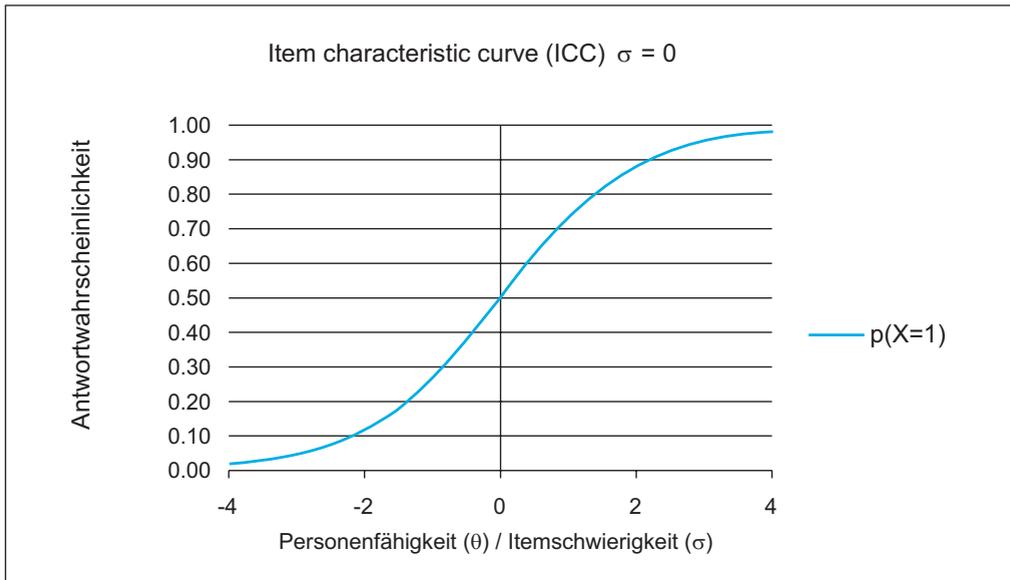


Abbildung 10.3: Item Characteristic Curve (ICC)

Mathematisch-formal ist das Modell mit den beiden Parametern σ (Itemparameter) und θ (Personenparameter) sowie der Lösungswahrscheinlichkeit $p(X)$ daher durch die folgende Modellgleichung definiert:

$$p(X_{vi}) = \frac{\exp(X_{vi}(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i)}, X \in \{0,1\}$$

Wobei die Laufindizes (v und i) jeweils für die Antwort der Person v auf das Item i stehen. Dabei ist zu beachten, dass die Kodierung der Antworten so zu erfolgen hat, dass $X_{vi} = 0$ für die falsche Antwort auf ein Item steht und $X_{vi} = 1$ für die Lösung. Dabei wird deutlich, dass das Modell in dieser grundlegenden Formulierung die Kodierung von teilweise richtigen Antworten nicht zulässt. In PISA werden allerdings neben einer solchen Richtig-Falsch-Kodierung der Antworten auch teilrichtige Antworten berücksichtigt, was die Verwendung einer Erweiterung des Rasch-Modells nötig macht, das sogenannte Partial-Credit-Modell.

Das Partial-Credit-Modell

Das von Masters (1982) formulierte *Partial-Credit-Modell* greift die grundsätzliche Idee des oben beschriebenen *Rasch-Modells* auf und generalisiert dieses auf mehrstufige Antwortformate. Dadurch wird eine Berücksichtigung von teilrichtigen Antworten möglich. Innerhalb der ursprünglichen Modellgleichung des *Rasch-Modells* wird dabei der Itemparameter in einzelne Schwellen-Parameter zerlegt, welche dann das mehrstufige, ordinale (hierarchisch geordnete) Antwortformat der Items abbilden.

$$p(X_{vi} = x) = \frac{\exp((x \cdot \theta_v) - \sigma_{ix})}{\sum_{s=0}^m \exp((s \cdot \theta_v) - \sigma_{is})}, x \in \{0, 1, \dots, m\}$$

Die Modellgleichung des Partial-Credit-Modells formalisiert so die Wahrscheinlichkeit, eine bestimmte Antwortkategorie x zu wählen, als Funktion der Ausprägung der Person auf der latenten Eigenschaft θ , der Anzahl der vorgegebenen Antwortkategorien beziehungsweise Schwellen s , der Summe aller (möglichen) Schwellenparameter σ_{is} und der Summe aller (realisierten) Schwellenparameter σ_{ix} bis zur gewählten Kategorie x , aus $m+1$ möglichen Kategorien. Zu beachten ist dabei, dass die Antwortkategorien eines Items in der Datenmatrix aufsteigend von 0 bis m kodiert werden müssen. Aus dieser formalen Darstellung des Partial-Credit-Modells (und der Kodierung der Antwortkategorien) wird deutlich, dass das Modell insofern eine Generalisierung des *Rasch-Modells* darstellt, als dass für den Spezialfall einer lediglich zweistufigen Richtig-Falsch-Kodierung die Modellgleichung des *Partial-Credit-Modells* in die Modellgleichung des *Rasch-Modells* überführt werden kann. Dies ist für die Auswertung der PISA-Erhebungen von Bedeutung, weil es dadurch ohne Probleme möglich ist, sowohl zweistufig („richtig/falsch“) als auch mehrstufig („richtig/teilrichtig/falsch“) kodierte Items innerhalb der Skalierung gleichzeitig zu verwenden.

Voraussetzung(en) für die Existenz und Schätzbarkeit von Personenparametern

In PISA werden die Personenparameter, das heißt die Kompetenz der teilnehmenden Schülerinnen und Schüler, im Rahmen der oben vorgestellten probabilistischen Testmodelle anhand der beobachteten Daten geschätzt. Schätzen bedeutet hierbei, dass in einem iterativen Verfahren diejenigen Parameter bestimmt werden, die nach dem zugrunde gelegten Test- beziehungsweise Antwortmodell die Wahrscheinlichkeit für das Auftreten der beobachteten Datenmatrix maximieren. Eine notwendige Voraussetzung dafür, dass für einen gegebenen Datensatz überhaupt eindeutige Modellparameter existieren und diese damit im Rahmen der probabilistischen Testmodelle schätzbar sind, besteht nach Fischer (1981) in bestimmten strukturellen Eigenschaften der erhobenen Daten

(vgl. Rost, 2004). Demnach muss gewährleistet sein, dass sich die beobachtete Datenmatrix nicht durch einfaches Umsortieren der Personen und Items in zwei (unabhängige) Teildatensätze aufteilen lässt. Dies wäre dann der Fall, wenn eine Teilgruppe der Personen n_1 eine Teilmenge der Items i_1 komplett nicht löst und gleichzeitig eine Teilgruppe der Personen n_2 eine Teilmenge der Items i_2 komplett nicht löst. Wobei zu beachten ist, dass im Prinzip mindestens zwei Ursachen, einzeln oder in Kombination, für das Nichtlösen der beiden Item-Teilgruppen von den beiden Personen Teilgruppen vorliegen können. So kann im einfachsten Fall bereits ein fehlerhaftes Design (vgl. Abschnitt zum *Multi-Matrix-Design*) der Erhebung in Verbindung mit mehreren Testheften dazu führen, dass bestimmte Personengruppen bestimmte Aufgabengruppen erst gar nicht zur Bearbeitung vorgelegt bekommen – und damit nicht lösen können. Andererseits kann aber auch die unbeabsichtigte, systematische Zuordnung von Aufgaben bestimmter Schwierigkeiten zu Personen mit einer bestimmten Eigenschaftsausprägung dazu führen, dass die resultierende Datenmatrix in zwei Teildatensätze zerfällt.

Durch das in PISA verwendete Testheft-Design wird sichergestellt, dass dieser Problematik einerseits durch eine geeignete Verteilung der einzelnen Items beziehungsweise Cluster auf die 13 Testhefte wirksam begegnet wird. Andererseits wird durch eine kontrollierte Ziehung der Personenstichprobe (vgl. Abschnitt Stichprobe) und eine geeignete Verteilung der 13 Testhefte innerhalb dieser Stichprobe dafür gesorgt, dass die oben beschriebene Voraussetzung zur Schätzbarkeit der Personenparameter gegeben ist.

Methoden zur Parameterschätzung

Seit der grundlegenden Formulierung des probabilistischen Testmodells durch den dänischen Mathematiker Georg Rasch im Jahre 1960 und dessen Erweiterung auf mehrstufige, ordinale Antwortformate durch Masters (1982) sind im Zusammenhang mit der praktischen Anwendung des Modells eine ganze Reihe unterschiedlicher Methoden zur Bestimmung der darin formulierten Personen- und Itemparameter entwickelt und vorgeschlagen worden. Jedes dieser unterschiedlichen Verfahren hat dabei bestimmte Vor- und Nachteile, welche jeweils in Abhängigkeit der verschiedenen Zielsetzungen empirischer Untersuchungen und deren Designs gegeneinander abgewogen und entsprechend ausgewählt werden müssen. Im folgenden Abschnitt sollen drei dieser Verfahren zur Parameterbestimmung, welche in der aktuellen bildungswissenschaftlichen Forschung und Literatur vorherrschen, kurz skizziert werden.

Wie bereits erwähnt, gilt bei probabilistischen Testmodellen die Annahme, dass die beiden grundlegenden Parameter, die Itemschwierigkeit (Itemparameter σ) sowie die Ausprägung der Person auf einer Eigenschaftsdimension (z. B. ihre Fähigkeit, Personenparameter θ), das Antwortverhalten der Schülerinnen und Schüler und damit deren Reaktion auf die einzelnen Aufgaben erklären. Technisch gesprochen lässt sich also die Realisierung der beobachteten Daten, im Rahmen der Erfassung von latenten Personeneigenschaften, als Resultat bestimmter vorliegender Werte dieser beiden Parameter erklären, welche im Rahmen der Skalierung anhand dieser empirisch vorgefunde-

nen Datenmatrix ermittelt werden können. Die im Folgenden erläuterten Verfahren zur Bestimmung dieser beiden Parameter stützen sich, ausgehend von diesen Überlegungen, auf die sogenannte Likelihood der Daten. Diese Likelihood drückt in einem Wert aus, wie gut bestimmte Werte der beiden Parameter σ und θ zu den empirisch beobachteten Daten (der Datenmatrix), unter der Annahme der Gültigkeit des psychometrischen Antwortmodells, passen. Das Ziel der Parameterbestimmung besteht also darin, diejenigen Werte für die beiden Parameter zu finden, welche die empirisch beobachteten Daten am besten erklären – beziehungsweise die Wahrscheinlichkeit dieser Datenmatrix maximiert (vgl. auch Bortz & Schuster, 2010, S. 92). Dieses Prinzip im Vorgehen wird mit dem Begriff *Maximum Likelihood (Schätzung)* bezeichnet.

Joint-Maximum-Likelihood-Schätzung (JML)

Bei der *Joint-Maximum-Likelihood-Methode* (JML) werden die beiden Parameter σ und θ gleichzeitig (daher „joint“) bestimmt. Ausgehend von den weiter oben dargestellten Modellgleichungen wird dabei zunächst die *Likelihood-Funktion* aufgestellt und dann deren erste Ableitung gleich Null gesetzt, um das Funktionsmaximum zu bestimmen. Allerdings resultiert dabei ein Gleichungssystem, das mehrere Unbekannte (Größen) beinhaltet, welches daher nicht explizit gelöst werden kann, sondern iterativ angenähert (approximiert) werden muss (daher auch Schätzung der Parameter). Für diese Methode der Parameterbestimmung konnte Habermann (1977) zeigen, dass sie unter bestimmten Bedingungen zu konsistenten Parameterschätzungen führt. Nähert sich die Anzahl der Personen in der Stichprobe dem Umfang der Population an, so nähern sich die geschätzten Itemparameter den wahren Itemparametern an. Die gleiche Bedingung gilt analog für die Annäherung der geschätzten Personenparameter an die wahren Personenparameter bei einer zunehmenden Anzahl von Items.¹¹ Diese beiden Konsistenzbedingungen für die Parameter führen in der Praxis allerdings dazu, dass bei Anwendung der JML-Methode die Schätzung der Itemparameter stets genauer ausfällt als die Schätzung der Personenparameter. Die beiden Parameter sind nämlich, wie die Datenmatrix, nicht symmetrisch zueinander. Während die Anzahl der Personen in der untersuchten Stichprobe zur Steigerung der Schätzgenauigkeit (der Itemparameter) relativ leicht erhöht werden kann, ist dies bei der Anzahl der Items, insbesondere bei der Erfassung mehrerer latenter Personeneigenschaften, nicht ohne Weiteres möglich. Da die Struktur (Anzahl der Items) eines psychologischen Messinstruments (Fragebogen oder Test) mit der Datenerhebung festgelegt ist, wohingegen die Anzahl der Personen welche diese Items bearbeiten, leicht veränderbar ist, unterscheidet man zwischen *strukturellen* (Item-) und *inzidentellen* (Personen-)Parametern. Das Problem der JML-Methode mit dem Prinzip der symmetrischen beziehungsweise gleichzeitigen Parameterschätzungen

¹¹ Bei dieser Art der Betrachtung geht man implizit davon aus, dass die untersuchten Items, also die in der betreffenden Skala verwendeten Items, lediglich eine Auswahl aller möglichen Items zur Erfassung der latenten Variable sind – es handelt sich also gewissermaßen um eine (repräsentative) Item-Stichprobe aus einer Population (unbekannter Größe) von Items.

besteht bei vergleichsweise kleiner Anzahl von Items und großen Stichprobenumfängen darin, dass jede hinzukommende Person einen zu schätzenden Personenparameter mitbringt, ohne dass die Genauigkeit der Schätzung (dieser Personenparameter) durch Hinzunahme von Items gesteigert werden könnte. Diese Eigenschaften führen dazu, dass die Schätzung der Personenparameter bei gleichzeitiger Schätzung der Itemparameter zu inkonsistenten Schätzwerten führen können (Gustafsson, 1980; Neyman & Scott, 1948). Insbesondere bei Verwendung relativ weniger Items für jede der zu erfassenden latenten Eigenschaften können die Parameterschätzungen mit der JML-Methode verzerrt ausfallen.

Conditional-Maximum-Likelihood-Schätzung (CML)

Bei der *Conditional-Maximum-Likelihood*-Methode (CML) wird dieser Problematik der Asymmetrie der beiden Parameter dadurch begegnet, dass diese in zwei Schritten getrennt geschätzt werden. Dabei nutzt man eine besondere Eigenschaft der Rasch-Modellgleichung aus, welche dazu führt, dass sich der Personenparameter in der *Likelihood*-Funktion zur Schätzung der Itemparameter, unter der Bedingung (daher *conditional*) der Anzahl der gelösten Items (*Scoregruppe*), herauskürzen lässt. Auf diese Weise ist es mit der CML-Methode möglich, in einem ersten Schritt zunächst die Itemparameter zu schätzen, um dann auf dieser Basis die Personenparameter jeweils für alle die Personen zu schätzen, welche dieselbe Anzahl von Items gelöst haben. Damit wird die mit der JML-Schätzung verbundene Problematik, dass mit jeder zusätzlichen Person in der Stichprobe ein zusätzlicher Personenparameter geschätzt werden muss, umgangen.

Da also bei diesem Verfahren bei der Parameterschätzung auf die *Scoregruppe* (Anzahl der gelösten Items), also den Summenwert der Personen auf der Skala konditioniert wird, eignet sich dieses Verfahren dann sehr gut zur Parameterbestimmung, wenn vom Design der Untersuchung her sämtliche Personen der Stichprobe dieselbe Itemauswahl bearbeiten können beziehungsweise vorgelegt bekommen. In einem solchen Fall bietet dieses Verfahren gegenüber anderen Verfahren dann den Vorteil konsistenter und nicht verzerrter Item- und Personenparameter-Schätzungen. Darüber hinaus müssen keinerlei A-priori-Annahmen über die Verteilung der latenten Personeneigenschaft gemacht werden, da diese ja im ersten Schritt bei der Schätzung der Itemparameter herauskonditioniert wurden.

Marginal-Maximum-Likelihood-Schätzung (MML)

Wie oben dargestellt eignet sich das Prinzip der getrennten Schätzung der Modellparameter durch Herausconditionieren der Personenparameter dazu, unverzerrte Parameterschätzer zu erhalten. Wird bei der Auswertung von Tests eine mehrdimensionale Skalierung angestrebt (vgl. Abschnitt *Mehrdimensionales Rasch-Modell*), so würde ein Herausconditionieren der Personenparameter über die Anzahl der gelösten Aufgaben,

insbesondere bei langen Skalen beziehungsweise vielen Items, zu einem massiven Ansteigen der Anzahl der zu schätzenden Modellparameter führen. Es müssen bei solch einer mehrdimensionalen Skalierung nämlich nicht nur alle möglichen Anzahlen aller gelösten Items – also *Scoregruppen* – für jede einzelne Skala berücksichtigt werden, sondern auch alle möglichen Kombinationen aus diesen *Scoregruppen* (vgl. Rost, 2004). Bei der *Marginal-Maximum-Likelihood-Methode* (MML), welche in ihrer ursprünglichen Form auf Bock und Lieberman (1970) und deren Erweiterung auf Bock und Aitkin (1981) zurückgeht, wird diesem Problem des Anwachsens der (zu schätzenden) Modellparameter begegnet. Das Prinzip besteht darin, dass bezüglich der zu schätzenden Personenparameter a priori angenommen wird, dass deren Verteilung der Normalverteilung folgt, welche mit lediglich zwei Parametern (Mittelwert und Streuung) zu beschreiben ist. Da sich diese Annahme der Normalverteilung auf die (anhand der Stichprobe) zu beschreibende Population (aller fünfzehnjährigen Schüler) bezieht, spricht man in diesem Zusammenhang auch vom sogenannten *Populationsmodell* (vgl. dazu auch Abschnitt *Latente Regression im Hintergrundmodell*). Auch bei der MML-Methode werden (wie bei der CML-Methode) Item- und Personenparameter in zwei Schritten getrennt geschätzt, um unverzerrte Schätzwerte zu erhalten. Im Gegensatz zur weiter oben beschriebenen CML-Methode werden die Personenparameter allerdings nicht über die *Scoregruppen*, sondern über die Verteilungsannahme der Personenparameter herauskonditioniert. Der Vorteil dieser Methode besteht also darin, dass insbesondere bei komplexeren Modellen (Multimatrix-Design, mehrere erfasste latente Konstrukte und viele Items pro Skala) die Anzahl der zu schätzenden Modellparameter nicht so extrem anwächst und diese damit stabil geschätzt werden können. Die Verringerung der Anzahl der zu schätzenden Modellparameter fällt bei der in PISA 2012 vorgenommenen multidimensionalen Skalierung (vgl. Abschnitt *Mehrdimensionales Rasch-Modell*) besonders ins Gewicht. Da die Verteilung der latenten Variablen als normalverteilt angesehen wird, können auch deren Interkorrelationen (also z.B. der Zusammenhang zwischen den Kompetenzen Mathematik, Lesen und Naturwissenschaften) direkt im Rahmen der Skalierung ermittelt werden (vgl. Rost, 2004, S. 263f.). Als Fazit zur MML-Methode lässt sich festhalten, dass sich dieses Verfahren, aufgrund des dargelegten Prinzips, sehr gut zur Bestimmung der Modell-Parameter im Rahmen der mehrdimensionalen Skalierung auf der Basis von Daten, welche mit *Multi-Matrix-Designs* an großen Stichproben aus der betreffenden Population erhoben wurden, eignet.

Weighted-Likelihood-Schätzer

Als Ergebnis der oben beschriebenen (Schätz-)Verfahren zur Bestimmung der Modellparameter im Rahmen von IRT-Modellen erhält man üblicherweise die *Maximum-Likelihood-Schätzer* (Schätzer = engl. *estimates*, daher MLE) als sogenannte Punkt-Schätzwerte. Durch einen iterativen Algorithmus werden diejenigen Werte für die Modell-Parameter ermittelt, welche die Wahrscheinlichkeit für die beobachteten Daten (daher der englische Begriff *Maximum Likelihood*) unter der Annahme eines bestimm-

ten Antwortmodells am wahrscheinlichsten werden lassen (vgl. Rost, 2004). Diese Punkt-Schätzwerte und die dazugehörigen Standardfehler ergeben sich im Rahmen der Parameterschätzung als Funktion der Randsummen der beobachteten Datenmatrix und können, betrachtet man die Personenparameter, direkt als Wert für das Ausmaß der Eigenschaftsausprägung einer bestimmten Person interpretiert werden. Entsprechend stellt deren Standardfehler ein Maß für die erzielte Genauigkeit dieser Punktschätzung dar. Für einen Teil der in PISA erfassten (latenten) Personenvariablen werden als Punktschätzer sogenannte *Weighted-Likelihood-Schätzer* (WLE – Warm, 1989) bestimmt, welche insofern eine Erweiterung beziehungsweise Abwandlung des MLE-Prinzips darstellen, als dass bei diesen Schätzern die Bestimmung von Parametern für Personen ermöglicht wird, welche alle oder keines der Items einer Skala richtig beantwortet haben. Abschließend lässt sich feststellen, dass sich diese Punktschätzungen für die Personeneigenschaften dabei ausschließlich auf die Informationen beziehen, die sich aus der Bearbeitung jener Aufgaben ergeben, welche der Person vorgelegt wurden.

Plausible Values

Neben diesen eben beschriebenen Punktschätzungen für die Eigenschaftsausprägung der untersuchten Schülerinnen und Schüler werden in PISA zur Bestimmung der Populationsschätzwerte für die drei Kompetenzbereiche (Mathematik, Lesen und Naturwissenschaften) sogenannte *Plausible Values* (PV) eingesetzt (Mislevy, 1991; Mislevy, Beaton, Kaplan & Sheehan, 2005). Die grundlegende Idee besteht im Vergleich zu der oben beschriebenen Punktschätzung darin, zunächst eine A-posteriori-Verteilung des erfassten Kompetenzbereiches auf Basis der beobachteten Daten zu modellieren. Ist eine solche Verteilung bestimmt, wird aus dieser zufällig eine bestimmte Anzahl von Werten gezogen – die *Plausible Values* (PVs). Im Gegensatz zu den oben beschriebenen WL-Schätzern erhält man also für einen zu schätzenden Parameter nicht lediglich einen Schätzwert, sondern eine (Wahrscheinlichkeits-)Verteilung aller möglichen, plausiblen (Schätz-)Werte für diesen Parameter. In der PISA-2012-Studie wurden aus diesen Verteilungen der Parameterschätzungen jeweils fünf Werte zufällig gezogen. Der Vorteil dieses Vorgehens liegt darin begründet, dass auf Basis der Plausible Values, Populationsstatistiken, wie zum Beispiel Mittelwertunterschiede zwischen zwei Gruppen, unverfälscht geschätzt werden können, da diese nicht durch die Messfehler auf Personenebene verfälscht sind.

Die auf diesen Parameterschätzungen aufbauenden Analysen (wie z. B. Bestimmung der durchschnittlichen Kompetenz aller Schülerinnen und Schüler in Deutschland oder Vergleiche von Mittelwerten zwischen verschiedenen Teilstichproben) werden mit jedem einzelnen der fünf PVs durchgeführt und die Ergebnisse daraus gemittelt.

Mehrdimensionales Rasch-Modell

Sowohl das *Rasch-Modell* in seiner ursprünglichen Formulierung durch Georg Rasch als auch dessen Erweiterung für mehrstufige Antwortformate (Masters, 1982) gehen davon aus, dass es sich bei den untersuchten Items um manifeste Indikatoren für eine einzige latente Personeneigenschaft handelt, die sich lediglich hinsichtlich ihrer Schwierigkeit unterscheiden. Sollen, wie in PISA 2012, mehrere latente Personeneigenschaften beziehungsweise Kompetenzbereiche untersucht werden, müssten zur Skalierung jeweils die entsprechenden Items für einen Kompetenzbereich getrennt analysiert werden. Eine vorteilhaftere Alternative zur getrennten Skalierung der einzelnen Kompetenzbereiche stellt die sogenannte *Mehrdimensionale Rasch-Skalierung* dar. Bei diesem Ansatz wird, analog zum Prinzip der Faktorenanalyse im Rahmen der klassischen Testtheorie, zunächst die gesamte Itemmenge für alle untersuchten Kompetenzbereiche gemeinsam analysiert. Diese gemeinsame Skalierung wird durch eine mehrdimensionale Erweiterung der Modellgleichung des *Rasch-Modells* beziehungsweise des *Partial-Credit-Modells* möglich (Reckase, 2007). In PISA 2012 wurde die mehrdimensionale (Wu, Adams, Wilson & Haldane, 2007) Rasch-Skalierung der drei Kompetenzbereiche (Mathematik, Naturwissenschaften und Lesen) mit dem in der Software *ConQuest* (Adams, Wu & Wilson, 2012) implementierten *Mixed-Coefficients-Multinomial-Logit-Modells* (MCML – Adams & Wu, 2007) vorgenommen (vgl. auch Adams, Wilson & Wang, 1997). Das grundlegende Prinzip einer solchen mehrdimensionalen Erweiterung soll hier, der besseren Übersicht und einfacheren Darstellung wegen, am Beispiel des *Faceted-Design-Ansatzes* von Rost und Carstensen (2002) vorgestellt werden. Es kann gezeigt werden, dass dieser hier beispielhaft vorgestellte Ansatz wiederum als Spezialfall der allgemeineren, in PISA 2012 angewandten Modellparametrisierung dargestellt werden kann. Eine vertiefende Darstellung und vergleichende Diskussion solcher mehrdimensionalen Erweiterungen ist zum Beispiel bei Rost und Carstensen (2002) zu finden.

Analog zur Formalisierung des Modells der gemeinsamen Faktoren in der klassischen Testtheorie wird bei einer mehrdimensionalen Erweiterung des Rasch- beziehungsweise des Partial-Credit-Modells ein zusätzlicher Parameter zu der ursprünglichen Modellgleichung hinzugefügt, welcher gewissermaßen als Gewichtung angibt, wie die Items jeweils den zu skalierenden Dimensionen zuzuordnen sind. Dieser in der unten abgebildeten Modellgleichung für das dichotome Rasch-Modell mit q bezeichnete, zusätzliche Parameter entspricht, im übertragenen Sinne, den Faktor-Ladungen in der Faktorenanalyse (vgl. auch Rost, 2004).

$$p(X_{vi}) = \frac{\exp\left(X_{vi} \sum_{j=1}^h q_{ij} (\theta_{vj} - \sigma_i)\right)}{1 + \exp\left(\sum_{j=1}^h q_{ij} \theta_{vj} - \sigma_i\right)}, X \in \{0,1\}$$

Im Gegensatz zum explorativen Vorgehen bei der Faktorenanalyse, bei der die Anzahl der Dimensionen anhand bestimmter Kriterien aus den empirischen Daten erst ermittelt wird, muss bei der Skalierung nach dem mehrdimensionalen *Rasch-Modell* eine bestimmte Anzahl von Dimensionen theoriegeleitet festgelegt werden. Der in der obigen Modellgleichung ergänzte Parameter q wird also nicht geschätzt, sondern muss (in Form einer Matrix) a priori vorgegeben werden. In einer solchen, sogenannten Q-Matrix wird letztlich spezifiziert, welches Item als manifester Indikator einer jeweiligen latenten Eigenschaft angesehen wird. Ein einfaches Beispiel einer solchen Q-Matrix ist in Tabelle 10.4 gegeben.

Tabelle 10.4: Beispiel einer Q-Matrix für 9 Items und 3 Kompetenzbereiche

Items	Mathematik	Naturwissenschaften	Lesen
I 1	1	0	0
I 2	1	0	0
I 3	1	0	0
I 4	0	1	0
I 5	0	1	0
I 6	0	1	0
I 7	0	0	1
I 8	0	0	1
I 9	0	0	1

Neben der Möglichkeit die Gültigkeit der a priori getroffenen Annahmen hinsichtlich der Zuordnung der Items zu den latenten Eigenschaften über Aussagen zur Modellpassung zu überprüfen, ergeben sich bei diesem Skalierungsansatz noch weitere Vorteile bezüglich der weiteren Analysen, welche im Folgenden kurz skizziert werden sollen.

10.2.2 Latente Korrelationen als mehrdimensionale Modellierung

MCML, MML-Schätzung

Zur Bestimmung der Populations-Schätzwerte für die Stichprobe in PISA 2012 wurde die weiter oben beschriebene MML-Methode eingesetzt. Nachdem für die drei in PISA erfassten Kompetenzbereiche (Mathematik, Naturwissenschaften und Lesen) die plausible, und auch empirisch untermauerte Annahme (vgl. Prenzel et al., 2004) besteht, dass diese untereinander korreliert sind, erfolgt die Skalierung dieser drei Bereiche gemeinsam im Rahmen eines mehrdimensionalen Antwortmodells – dem *Mixed-Coefficients-Multinomial-Logit-Modell* (MCML – Adams & Wu, 2007, S. 57ff.) –, welches, wie bereits beschrieben, eine mehrdimensionale Erweiterung des Eingangs skizzierten *Rasch-*

Modells beziehungsweise des *Partial-Credit-Modells* darstellt. Diese Vorgehensweise bietet, neben den oben beschriebenen Vorteilen in Bezug auf das Design der PISA-Studie, weitere Vorteile, die sich positiv auf die Genauigkeit (Reliabilität) bei der Erfassung der drei Kompetenzbereiche sowie deren Zusammenhänge untereinander auswirken. Diese Vorteile sollen im Folgenden kurz skizziert werden.

Ganz allgemein ist bei der Erfassung von latenten Variablen stets mit einer gewissen Unschärfe beziehungsweise einem Messfehler zu rechnen. Das Ausmaß der Messgenauigkeit wird in der Psychometrie dabei über die Reliabilität der Messung ausgedrückt, welche theoretisch als Quotient der wahren (latenten) Varianz und der Varianz der beobachteten (fehlerbehafteten) Messwerte definiert ist. Für den hypothetischen Fall einer perfekt reliablen Messung ohne Fehler würde dieser Quotient den Wert $r = 1$ annehmen. In der Praxis der psychometrischen Messung ist aber stets mit geringeren Werten für die Reliabilität (zwischen 0 und 1) zu rechnen. Diese Ungenauigkeit der Messung ist insofern von Bedeutung, als dass sie die, auf Basis der empirischen Daten, berechneten Zusammenhänge zwischen zwei oder mehreren gemessenen latenten Variablen mindert. Bei der in PISA eingesetzten mehrdimensionalen Modellierung im Rahmen des MCML-Modells werden die korrelativen Beziehungen der drei Kompetenzbereiche (und anderer Variablen) untereinander direkt als (zusätzliche) Modellparameter geschätzt und dabei die Minderung dieser korrelativen Beziehungen aufgrund etwaiger Messungenauigkeiten korrigiert. Neben dieser Reliabilitätskorrektur der korrelativen Beziehungen der latenten Variablen können diese Zusammenhänge zwischen den erfassten Variablen dazu genutzt werden, genauere Schätzungen der Kompetenzwerte in den einzelnen Bereichen für die Schülerinnen und Schüler zu erhalten.

10.2.3 Latente Regression im Hintergrundmodell

Neben den Aufgaben zu den drei Kompetenzbereichen (*Mathematik, Lesen* und *Naturwissenschaften*) bearbeiten die Schülerinnen und Schüler im Rahmen der PISA-Erhebung einen weiteren Fragebogen. Darin werden unterschiedliche Einstellungen zu schul- und unterrichtsbezogenen Fragestellungen erfasst. Darüber hinaus wurden die Eltern der an PISA teilnehmenden Schülerinnen und Schüler aufgefordert im *Elternfragebogen* Angaben zu schulbezogenen Einstellungen sowie Angaben zu ihrer persönlichen (Lebens-)Situation zu machen. Über entsprechende *Schlüsselvariablen* ist es möglich, die Daten aus diesen zusätzlichen Fragebögen, unter Wahrung der anonymisierten Datenverarbeitung, mit den Ergebnissen aus der Kompetenzmessung in den drei Kompetenzbereichen in Bezug zu setzen. Diese zusätzlich erhobenen *Hintergrundvariablen* werden zum einen in den einzelnen Teilkapiteln dazu genutzt, vertiefende Analysen der Kompetenzverteilung in spezifischen Schülergruppen durchzuführen. Für die Skalierung der Antworten der Schülerinnen und Schüler nach dem weiter oben beschriebenen Rasch-Modell kommt den *Hintergrundvariablen* noch eine weitere Rolle zu.

Bezogen auf die Schätzung der Kompetenzverteilung auf Populationsebene hat es sich gezeigt, dass die Schätzgenauigkeit der Personenparameter mit der weiter oben beschriebenen MML-Methode verbessert werden kann, wenn die zusätzlich erhobenen *Hintergrundvariablen* bei der Schätzung der Personenparameter mit berücksichtigt werden (Mislevy & Sheehan, 1989). Die grundlegende Idee besteht darin, die untersuchte Stichprobe (anhand der Hintergrundinformationen) in Gruppen zu unterteilen und den jeweiligen Mittelwertparameter zu schätzen. Ähnlich wie bei den im Abschnitt 10.2.2 beschriebenen latenten Korrelationen sind diese Gruppenparameter als Populationschätzer nicht von der Messungsgenauigkeit auf Personenebene beeinflusst.

Die Analysen zur Kompetenzverteilung in den unterschiedlichen Teilkapiteln des vorliegenden Bandes basieren auf diesen Populationsschätzern, welche unter Berücksichtigung derjenigen *Hintergrundvariablen* bestimmt wurden, die im Rahmen der Internationalen Erhebung in allen teilnehmenden Staaten erfasst wurden. Dieses Vorgehen wurde gewählt, um für die Ergebnisse in allen Teilkapiteln des vorliegenden Bandes eine Vergleichbarkeit sowohl zwischen den Kapiteln als auch zu den von der OECD, im Rahmen der Internationalen Skalierung, veröffentlichten Ergebnissen sicherzustellen. Für eine detaillierte Aufstellung der in diesem ‚internationalen‘ Hintergrundmodell verwendeten Variablen sei an dieser Stelle auf den zu diesem PISA-Zyklus erscheinenden Technischen Bericht der OECD verwiesen (in Vorbereitung).

10.3 Reliabilität und Stichprobenfehler

Wie bereits im Abschnitt 10.1 kurz angesprochen, ist bei der Messung von latenten Variablen im Kontext bildungswissenschaftlicher Fragestellungen immer mit einer gewissen Unschärfe zu rechnen. Die Reliabilität einer Messung, also deren Genauigkeit, wird dabei, allgemein formuliert, als Verhältnis von verschiedenen Varianzquellen operationalisiert. Das Ziel dabei ist einzuschätzen, wie groß der Fehleranteil in der Variabilität (also der Varianz) der beobachteten Messwerte ist. Es geht also darum abzuschätzen, in welchem Verhältnis die unerwünschten Varianzquellen (Fehlerquellen bei der Messung) zu den inhaltlich begründeten und im Rahmen der Fragestellung interessierenden Varianzquellen (reale Unterschiede in den Kompetenzen) stehen. In diesem Abschnitt sollen diese unerwünschten Varianzquellen spezifisch für die PISA-Studie, sowie die sich daraus ergebenden Implikationen, diskutiert werden.

Im Rahmen der bei PISA verwendeten Messinstrumente zur Erfassung der drei Kompetenzbereiche sowie vor dem Hintergrund der Methodik der Stichprobenziehung (vgl. Abschnitt 10.1) sind zunächst zwei Hauptquellen unerwünschter Varianz beziehungsweise Ungenauigkeit, bezogen auf die in diesem Band berichteten Parameter (wie z. B. der Mittelwert der Kompetenz der Schülerinnen und Schüler im Bereich Mathematik), zu unterscheiden. Zum einen besteht eine potentielle Quelle von Unsicherheit darin, dass (wie bei fast allen sozialwissenschaftlichen Untersuchungen) zur Gewinnung der statistischen Kennwerte eine Stichprobe aus einer definierten Grundgesamtheit (bei

PISA alle fünfzehnjährigen Schülerinnen und Schüler) herangezogen wird. Die zweite Quelle von Unsicherheit bezieht sich auf das eingesetzte Messinstrument und resultiert aus der zunächst trivial erscheinenden Tatsache, dass man den teilnehmenden Schülerinnen und Schülern zum Beispiel zur Erfassung ihrer Mathematikkompetenz, nicht unendlich viele Mathematikaufgaben, sondern eben nur eine Auswahl aller theoretisch denkbaren Aufgaben vorlegen kann. Diese beiden Quellen der Unsicherheit werden dieser Darstellung entsprechend als *Stichprobenfehler* und *Messfehler* bezeichnet.

10.3.1 Designeffekte und Berechnung von Stichprobenfehlern

Im Rahmen der klassischen Testtheorie lässt sich bezogen auf einfache Zufallsstichproben (*simple random sample*) zeigen, dass sich der Stichprobenfehler für unterschiedliche statistische Kennwerte proportional mit dem Ansteigen des Stichprobenumfangs n verringert. Für derartige einfache Zufallsstichproben lassen sich (relativ einfache) explizite Formeln zur Berechnung des als Standardfehler des jeweiligen statistischen Kennwertes bezeichneten Stichprobenfehlers herleiten (Bortz, 2005). Eine Voraussetzung für diese relativ einfache Bestimmung der Standardfehler von statistischen Kennwerten besteht dabei darin, dass die an den einzelnen Schülerinnen oder Schülern erhobenen Daten statistisch unabhängig voneinander sind. Diese Voraussetzung ist allerdings bei dem aus praktischen Überlegungen hinsichtlich der Durchführbarkeit resultierenden Ziehungsverfahren für die PISA-Stichprobe (vgl. Abschnitt 10.1.1) nicht gegeben. Da bei PISA zunächst die Schulen und in den Schulen die einzelnen Schülerinnen oder Schüler zufällig gezogen werden, ist davon auszugehen, dass für Schülerinnen oder Schüler innerhalb einer Schule im statistischen Sinne Abhängigkeiten bestehen. Es ist unmittelbar einsichtig, dass zum Beispiel die Unterschiede in der Mathematikkompetenz zwischen zwei Schülerinnen oder Schülern aus derselben Schulart (z. B. Gymnasium) vergleichsweise geringer ausfallen als die Unterschiede zwischen zwei Schülerinnen oder Schülern aus unterschiedlichen Schularten. Die Variabilität der Kompetenz der Schülerinnen oder Schüler in der PISA-Gesamtstichprobe lässt sich so zu einem gewissen Teil auf die Variabilität zwischen den (unterschiedlichen) Schulen zurückführen, die, und das ist hier der entscheidende Unterschied zur eben erwähnten einfachen Zufallsstichprobe, bei der Ziehung der Stichprobe bereits als Schichtungsmerkmal bei der Ziehung mit eingegangen ist. Das Ausmaß des Verhältnisses der Variabilität zwischen den Schulen und zwischen den Schülerinnen oder Schülern (innerhalb einer Schule) lässt sich a priori nicht bestimmen – sondern ist oftmals erst Gegenstand einer empirischen Fragestellung (vgl. dazu die die Zerlegung der Varianz innerhalb und zwischen Schulen in Kapitel 5). Darüber hinaus ist davon auszugehen, dass dieses Verhältnis zum Beispiel bezogen auf die OECD-Gesamtstichprobe durchaus unterschiedlich (je nach OECD-Staat und Schulsystem) ausfallen wird (vgl. auch OECD, 2012, S. 168ff.). Damit ergeben sich aber zum Beispiel bezogen auf einen Vergleich der einzelnen Kompetenz-Mittelwerte der OECD-Staaten unterschiedlich ausgeprägte statistische Abhängigkeiten zwischen den Schülerinnen

und Schülern, die je nach OECD-Staat variieren. Dieser Effekt, der sich aus der Art der Stichprobenziehung ergibt, wird auch als *Designeffekt* bezeichnet (Adams, 2005). Für die im Rahmen der Datenauswertung zu bestimmenden Standardfehler der einzelnen Koeffizienten würde dies bedeuten, dass diese abhängig vom Varianzverhältnis (zwischen den Schulen und Schülerinnen sowie Schülern) ausfällt und somit je nach PISA-Teilnehmerstaat in unterschiedlichem Ausmaß unterschätzt werden. Aufgrund der mehrstufigen Stichprobenziehung (Schülerebene in Schulen) und den damit einhergehenden unterschiedlichen Ziehungswahrscheinlichkeiten der Untersuchungsobjekte (Schülerinnen und Schüler), besteht also die Gefahr der Unterschätzung der Varianz der verschiedenen statistischen Kennwerte (vgl. Berger & Tillé, 2009). Da sich das Ausmaß des Verhältnisses der Variabilität zwischen den Schulen und den Schülerinnen sowie Schülern nicht a priori bestimmen lässt, wird bei PISA der Stichprobenfehler über ein Replikationsverfahren berechnet, um die sich aus der Art der Stichprobenziehung ergebende Unsicherheit angemessen zu berücksichtigen. Die dahinter liegende Idee besteht darin, aus der vorliegenden Stichprobe nach bestimmten Kriterien Teilstichproben (*Replikationen*) zu ziehen und mit diesen den jeweils interessierenden statistischen Kennwert (z. B. den Mittelwert) zu berechnen. Die sich aus solchen wiederholten Berechnungen ergebende Variabilität (Varianz) des jeweiligen statistischen Kennwerts drückt somit den Fehleranteil aus, der auf die Art und Weise der Stichprobenziehung zurückzuführen ist (Rust & Rao, 1996). Für eine detailliertere Darstellung der bei PISA eingesetzten Methodik bei der Berechnung der Replikationen sei aus Platzgründen hier auf den entsprechenden Abschnitt im *Technical Report* verwiesen (z. B. OECD, 2012 und der zeitnah erscheinende *Technical Report* zu PISA 2012).

10.4 Bestimmung von Kompetenzstufen

Die Kompetenz der Schülerinnen und Schüler in den in PISA erfassten Domänen Mathematik, Lesen und Naturwissenschaften wird, wie bereits in den vorangegangenen Abschnitten dargestellt, als kontinuierliche latente Variable angesehen. Zur besseren Interpretation und zum leichteren Vergleich der Kompetenzen wird diese latente Variable jeweils auf einer standardisierten Skala abgetragen. Diese Skala ist dabei so definiert, dass sie bei ihrem ersten Auftreten als Hauptdomäne über alle teilnehmenden OECD-Staaten hinweg einen Mittelwert von 500 bei einer Standardabweichung von 100 aufweist. Betrachtet man die über diese beiden Koeffizienten definierte Kompetenzverteilung, so befinden sich in dem Intervall -1 Standardabweichung bis $+1$ Standardabweichung ungefähr 68 Prozent aller teilnehmenden Fünfzehnjährigen. Gut zwei Drittel der untersuchten Jugendlichen erreichen also einen Kompetenzwert im Bereich von 400 bis 600 Punkten.

Zur besseren inhaltlichen Bewertung solcher Kompetenzverteilungen werden diese in sogenannte *Kompetenzstufen* unterteilt. Die inhaltliche Beschreibung dieser Kompetenzstufen ergibt sich dabei jeweils aus den kognitiven Anforderungen der Aufga-

ben, welche bei einem bestimmten Kompetenzniveau mit einer bestimmten, festgelegten Wahrscheinlichkeit von den Schülerinnen und Schülern gelöst werden können. Die Zuordnung der Schülerinnen und Schüler zu einer bestimmten Kompetenzstufe gibt dann wiederum an, welche Arten von Aufgaben mit hoher Wahrscheinlichkeit gelöst werden können und an welchen die Jugendlichen wahrscheinlich scheitern.

Die Konstruktion beziehungsweise Definition der Kompetenzstufen der drei Kompetenzbereiche wurde dabei jeweils in denjenigen der vorangegangenen PISA-Zyklen durchgeführt, in denen der jeweilige Kompetenzbereich den Schwerpunkt der Erhebung bildete. So wurden zunächst fünf Kompetenzstufen für den Bereich Lesen in der PISA-Erhebungsrunde 2000 gebildet, welche 2009 erweitert wurden (OECD, 2001; OECD, 2012) und jeweils sechs Kompetenzstufen für die Bereiche Naturwissenschaften und Mathematik in den PISA-Erhebungsrunden 2003 und 2006 (OECD, 2004; OECD, 2007).

10.4.1 Kompetenzstufen und Lösungswahrscheinlichkeiten

Neben einer Darstellung von erzielten Punkten auf Kompetenzskalen liefert PISA regelmäßig auch eine inhaltlich zu interpretierende Beschreibung der Kompetenzen in Form von Kompetenzstufen. Diese Kompetenzstufen geben an, wozu die Schülerinnen und Schüler auf einer Stufe jeweils typischerweise in der Lage sind. Es geht also nicht darum, aufzulisten, welche PISA-Aufgaben die Schülerinnen und Schüler am Testtag jeweils lösen konnten und welche nicht, sondern um die Formulierung einer generellen Kompetenz (OECD, 2012). In Bezug auf die Wahrscheinlichkeit, mit der eine Schülerin oder ein Schüler eine Aufgabe lösen kann, wird eine Kompetenzstufe immer dann erreicht, wenn Aufgaben auf dieser Stufe mit einer Wahrscheinlichkeit von mindestens 50 Prozent gelöst werden können. Dies entspricht dem unteren Ende der Kompetenzstufe. Am oberen Ende einer Kompetenzstufe können die Fünfzehnjährigen die Aufgaben mit einer Wahrscheinlichkeit von etwa 70 Prozent richtig lösen (OECD, 2012).

10.4.2 Kompetenzstufen für Mathematik, Lesen und Naturwissenschaften

Die drei Kapitel zu den kognitiven Domänen Mathematik, Naturwissenschaften und Lesekompetenz enthalten eine differenzierte Beschreibung der Kompetenzstufen, welche zur inhaltlichen Beschreibung der Kompetenzen der Schülerinnen und Schüler dienen. In Tabelle 10.5 wird überblicksartig dargestellt, welches jeweils die unteren *Cut-Off*-Werte der Kompetenzstufen der drei Domänen sind.

Tabelle 10.5: *Cut-Off*-Werte der Kompetenzstufen in PISA 2012

Kompetenzstufe	Kompetenzbereich		
	Mathematik	Lesen	Naturwissenschaften
6	669	698	708
5	607	626	633
4	545	553	559
3	482	480	484
2	420	407	409
1	358	1a: 335 1b: 262	335

10.5 Skalierung von Veränderungen zwischen PISA 2000 und PISA 2012

Jede PISA-Erhebung liefert querschnittliche Indikatoren für Merkmale in verschiedenen Bildungssystemen. Über diese Momentaufnahmen hinaus besteht ein Ziel der PISA-Erhebungen in der längsschnittlichen Betrachtung von Veränderungen dieser Indikatoren, die sich möglicherweise als Ausdruck von Veränderungen der Bildungssysteme und sozialen Rahmenbedingungen der jeweils teilnehmenden Staaten zeigen. Seit Beginn der PISA-Erhebungen im Jahre 2000 bildet jeweils einer der drei Kompetenzbereiche den Schwerpunkt der Untersuchung. Um einerseits jeden der drei Bereiche einmal als Schwerpunkt zu erheben und andererseits die längsschnittlichen Veränderungen mit eindeutigem Bezug zu den vorangegangenen Erhebungsrounden zu erfassen, erfolgt die längsschnittliche Zuordnung der Schwerpunktsetzung in PISA seit dem Jahr 2000 nach dem in Abbildung 10.6 dargestellten Schema. Ganz allgemein ist es bei längsschnittlichen Betrachtungen notwendig, die einzelnen Messzeitpunkte über eine gemeinsame (Teil-)Menge von Aufgaben (sogenannte *Link-Units*, siehe oben) miteinander zu verbinden. Link-Units sind geheim gehaltene Aufgaben, die den Jugendlichen in unveränderter Form in mehreren aufeinanderfolgenden PISA-Erhebungsrounden vorgelegt werden. Darüber hinaus besteht bei der Interpretation etwaiger Veränderungen die Annahme, dass diese Link-Units zu allen Zeitpunkten die gleiche Kompetenz erfassen. Wie in Abbildung 10.4 dargestellt, liegen inzwischen fünf Messzeitpunkte mit Kohorten fünfzehnjähriger Schülerinnen und Schüler für die drei Kompetenzbereiche Mathematik, Lesen und Naturwissenschaften vor. Eine eindeutige Vergleichbarkeit der Ergebnisse der einzelnen Erhebungsrounden im Sinne einer längsschnittlichen Veränderung in einem der jeweiligen Kompetenzbereiche ist immer zwischen denjenigen Erhebungsrounden uneingeschränkt möglich, in denen der jeweilige Kompetenzbereich den Schwerpunkt bildete – sowie in den darauf folgenden Erhebungsrounden. Daraus ergibt sich, dass für PISA 2012, bei längsschnittlicher Betrachtung der Ergebnisse, die wichtigste Referenz die PISA-Erhebung im Jahre 2003 ist.

Domäne	Erhebungsrunde				
	2000	2003	2006	2009	2012
Lesen	●	●	●	●	●
Mathematik	●	●	●	●	●
Naturwissenschaften	●	●	●	●	●

Abbildung 10.4: Längsschnittliche Zuordnung der Schwerpunktdomänen in PISA seit 2000

Die in PISA 2000 erstmals berichteten Skalen zur Beschreibung von Kompetenzen in den Bereichen Mathematik, Lesen und Naturwissenschaften sind lineare Transformationen der natürlichen Logarithmen, die sich aus der Skalierung ergeben (OECD 2012; Adams & Wu, 2002). Für die Naturwissenschaften, die als Kompetenzbereich erstmals in PISA 2006 den Schwerpunkt bildeten, gilt dasselbe (OECD, 2007). Dies bedeutet, dass die jeweilige Hauptdomäne in PISA einen Anfangspunkt markiert, ab dem die Kompetenzskala so differenziert und ausführlich erfasst wird, dass statistisch haltbare Rückbezüge möglich und sinnvoll sind. Für Lesen war dieser Anfangspunkt PISA 2000, für Mathematik PISA 2003 und für die Naturwissenschaften PISA 2006.

Rückblickend sind im Bereich Mathematik, einschließlich PISA 2012, vier Erhebungsrunden auf einer gemeinsamen Skala zu veranschaulichen (2003, 2006, 2009 und 2012). Entsprechend gilt für die Naturwissenschaften, dass ein direkter Vergleich seit 2006 möglich ist und damit bisher drei Messzeitpunkte unmittelbar miteinander in Beziehung gesetzt werden können. Die Lesekompetenz wurde hingegen in allen fünf bisherigen PISA-Erhebungsrunden auf einer gemeinsamen Skala abgebildet (OECD, 2012). Beispielsweise sind die 484 Punkte, die deutsche Schülerinnen und Schüler im Mittel beim Lesen in PISA 2000 erreicht hatten (Baumert et al., 2001), unmittelbar vergleichbar mit dem Mittelwert im Lesen von 491 Punkten in PISA 2003 (Prenzel et al., 2004) beziehungsweise den 495 oder 497 Punkten in PISA 2006 (Prenzel et al., 2007) und 2009 (Klieme et al., 2010) sowie schließlich in PISA 2012 mit den durchschnittlich erreichten 508 Punkten.

Um die Ergebnisse aus verschiedenen PISA-Erhebungsrunden miteinander vergleichen zu können, besteht ein Teil der Aufgaben aus sogenannten Link-Units. Diese werden in unveränderter Form in mehreren Erhebungsrunden eingesetzt. Unter der Annahme, dass die Link-Units in jeder PISA-Erhebungsrunde dieselbe Kompetenz erfassen, wird in einem mehrstufigen Verfahren die jeweils aktuelle Kompetenzskala (z. B. Mathematik in PISA 2012) auf die gewünschte Referenzskala (z. B. Mathematik in PISA 2003) überführt. Dieses Verfahren nennt man *Equating* und unterscheidet dabei Item- und Personen-*Equating* (OECD, 2012). Für Daten, die wie in PISA nach der *Item*

Response Theory (IRT) skaliert sind, können Itemparameter direkt auf der Fähigkeitskala von Personen abgebildet werden. Trotz der 13 in PISA eingesetzten unterschiedlichen Testhefte ist es nicht notwendig, Vergleichbarkeit über eine Referenzkala anhand der Anzahl korrekt gelöster Aufgaben zu erzeugen (Yamamoto & Mazzeo, 1992). Für die Verlinkung mehrerer PISA-Erhebungsrunden müssen also die Kompetenzen aus unterschiedlichen PISA-Jahren auf einer gemeinsamen Skala repräsentiert werden, die durch die Anwendung des Antwortmodells auf der Fähigkeitsskala (Personenparameter) gleichgesetzt werden kann. In PISA findet dieses *Equating* für die Schwerpunktdomäne in sechs Schritten statt, wobei die ersten beiden Schritte zum Personen-*Equating* und die übrigen vier Schritte zum Item-*Equating* gezählt werden (OECD, 2012). Für die Nebendomänen wird ausschließlich Item-*Equating* eingesetzt, welches in vier Schritte unterteilt ist.

Im Kompetenzbereich Mathematik wurden in PISA 2012 folgende sechs Schritte für die Verlinkung zu früheren PISA-Erhebungsrunden seit PISA 2003 durchgeführt. In Schritt 1 erfolgte eine Kalibrierung der Itemschwierigkeit anhand der Kalibrierungstichprobe aus PISA 2012 (Verfahren analog zu PISA 2009 für die Schwerpunktdomäne Lesen, vgl. OECD, 2012). Der zweite Schritt transformierte die erhaltenen Itemschwierigkeiten, indem eine Konstante addiert wurde. Dadurch wurde der Mittelwert für Mathematik 2012 gleich demjenigen aus der letzten PISA-Erhebungsrunde, also 2009, gesetzt. In Schritt 3 wurde der Datensatz für alle OECD-Staaten in PISA 2012 zweimal skaliert; einmal mit allen Mathematik-Items und einmal nur mit den Link-Items. Schritt 4 berechnete für die OECD-Staaten die Differenz zwischen beiden Skalierungen, die als weitere Konstante zur Transformation hinzugefügt wurde. Anschließend erfolgte in Schritt 5 die Schätzung der Personenparameter (Fähigkeit) für PISA 2012, die an den Itemschwierigkeiten von 2012 verankert sind. Im letzten Schritt wurden die Personenparameter schließlich anhand der berechneten Verschiebung (Schritte 2 und 4) transformiert.

Eine solche Verlinkung ist stets auch mit einer gewissen Unsicherheit verknüpft. Diese Unsicherheit geht auf eine möglicherweise unterschiedliche Schwierigkeit der eingesetzten Aufgaben zwischen mehreren PISA-Erhebungsrunden zurück. Folglich kann die errechnete Transformation von der Auswahl der Link-Units abhängen und wäre minimal anders, wenn andere Link-Units eingesetzt worden wären. Ein Maß für diese Unsicherheit ist der sogenannte *Linking-Error* (OECD, 2012), der für Berechnungen mit PISA-Daten aus verschiedenen Erhebungsrunden berücksichtigt werden muss. Der *Linking-Error* kann als Wert nicht exakt bestimmt werden, jedoch ist eine Schätzung von einem durch den Standardfehler definierten Wertebereich möglich. Einzelheiten zum Umgang mit dem *Linking-Error* können dem *Technical Report* für PISA 2012 entnommen werden.

Literatur

- Adams, R. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31 (2-3), 162–172.
- Adams, R. J., Wilson, M. & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21 (1), 1–23.
- Adams, R. J. & Wu, M. (2002). *PISA 2000 Technical Report*. Paris: OECD.
- Adams, R. & Wu, M. (2007). A generalized form of the Rasch Model. In M. von Davier & C. H. Carstensen (Hrsg.), *Multivariate and mixture distribution Rasch Models: Extensions and applications* (1. Aufl., S. 57–75). New York: Springer.
- Adams, R., Wu, M. L. & Wilson, M. R. (2012). ACER ConQuest 3.0 [Computer software]. Melbourne: Australian Council for Educational Research (ACER).
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W. et al. (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Berger, Y. G. & Tillé, Y. (2009). Sampling with unequal probabilities. In D. Pfeffermann & C. R. Rao (Hrsg.), *Sample Surveys* (S. 39–54). Amsterdam: Elsevier.
- Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46 (4), 443–459.
- Bock, R. D. & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35 (2), 179–197.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (Springer-Lehrbuch, 6. Aufl.). Berlin: Springer.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.
- Daniel, J. (2012). *Sampling essentials – practical guidelines for making sampling choices*. Thousand Oaks: Sage.
- Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch Model. *Psychometrika*, 46 (1), 59–77.
- Fischer, G. H. & Molenaar, I. (1995). *Rasch Models: Foundations, recent developments, and applications*. New York: Springer.
- Gressard, R. P. & Loyd, B. H. (1991). A comparison of item sampling plans in the application of multiple matrix sampling. *Journal of Educational Measurement*, 28 (2), 119–130.
- Gustafsson, J.-E. (1980). A solution of the conditional estimation problem for long tests in the Rasch Model for dichotomous items. *Educational and Psychological Measurement*, 40 (2), 377–385.
- Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, 5 (5), 815–841.
- Häder, M. (2010). *Empirische Sozialforschung – Eine Einführung* (2. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Johnson, M. C. & Lord, F. M. (1958). An empirical study of the stability of a group mean in relation to the distribution of test items among students. *Educational and Psychological Measurement*, 18 (2), 325–329.
- Kalton, G. (1983). *Introduction to survey sampling*. Newbury Park: Sage.
- Kish, L. (1995). *Survey sampling*. New York: Wiley & Sons.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M. et al. (Hrsg.). (2010). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.

- KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2013). *Definitionenkatalog zur Schulstatistik 2012*. Zugriff am 27.05.2013. Verfügbar unter <http://www.kmk.org/statistik/schule/statistische-veroeffentlichungen/definitionenkatalog-zur-schulstatistik.html>.
- Levy, P. S. & Lemeshaw, S. (2008). *Sampling of populations – methods and applications* (4. Aufl.). Hoboken, NJ: Wiley & Sons.
- Lord, F. M. (1962). Estimating norms by item-sampling. *Educational and Psychological Measurement*, 22 (2), 259–267.
- Masters, G. (1982). A Rasch Model for partial credit scoring. *Psychometrika*, 47 (2), 149–174.
- Mislevy, R. J. & Sheehan, K. M. (1989). The role of collateral information about examinees in item parameter estimation. *Psychometrika*, 54 (4), 661–679.
- 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. & Sheehan, K. M. (2005). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29 (2), 133–161.
- Neyman, J. & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, 16, 1–5.
- OECD. (2001). *Knowledge and skills for life*. Paris: OECD.
- OECD. (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.
- OECD. (2007). *PISA 2006: Science competencies for tomorrow's world*. Paris: OECD.
- OECD. (2012). *PISA 2009 technical report*. Paris: OECD.
- OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. et al. (Hrsg.). (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M. et al. (Hrsg.). (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Rasch, G. W. (1960). *Probabilistic models for some intelligence and attainment tests* (Studies in mathematical psychology). Chicago: The University of Chicago Press.
- Reckase, M. D. (2007). Multidimensional item response theory. In C. R. Rao & S. Sinharay (Hrsg.), *Psychometrics* (Handbook of statistics, Bd. 26, 1. Aufl., S. 607–642). Amsterdam, Boston: Elsevier.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Berlin: Hans Huber.
- Rost, J. & Carstensen, C. H. (2002). Multidimensional Rasch measurement via item component models and faceted designs. *Applied Psychological Measurement*, 26 (1), 42–56.
- Rust, K. & Rao, J. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5 (3), 283–310.
- Thompson, S. K. (2012). *Sampling* (3. Aufl.). Hoboken, NJ: Wiley & Sons.
- van der Linden, J. W., Veldkamp, B. P. & Carlson, J. E. (2004). Optimizing balanced incomplete block designs for educational assessments. *Applied Psychological Measurement*, 28 (5), 317–331.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54 (3), 427–450.

- Wu, M., Adams, R., Wilson, M. & Haldane, S. (2007). *ACER ConQuest 2.0*. Generalised item response modelling software. Camberwell: ACER Press.
- Yamamoto, K. & Mazzeo, J. (1992). Chapter 4: Item response theory scale linking in NAEP. *Journal of Educational and Behavioral Statistics*, 17 (2), 155–173.

Anhang

Tabelle 1: Mittelwerte, Streuungen und Perzentile mathematischer Kompetenz

OECD-Staaten	Perzentile									
	M	(SE)	SD	(SE)	5%	10%	25%	75%	90%	95%
Korea	554	(4.6)	99	(2.1)	386	425	486	624	679	710
Japan	536	(3.6)	94	(2.2)	377	415	473	603	657	686
Schweiz	531	(3.0)	94	(1.5)	374	408	466	597	651	681
Niederlande	523	(3.5)	92	(2.1)	367	397	457	591	638	665
Estland	521	(2.0)	81	(1.2)	389	417	465	576	626	657
Finnland	519	(1.9)	85	(1.2)	376	409	463	577	629	657
Kanada	518	(1.8)	89	(0.8)	370	402	457	580	633	663
Polen	518	(3.6)	90	(1.9)	373	402	454	580	636	669
Belgien	515	(2.1)	102	(1.4)	343	378	444	589	646	677
Deutschland	514	(2.9)	96	(1.6)	353	385	447	583	637	667
Österreich	506	(2.7)	92	(1.7)	353	384	440	572	624	654
Australien	504	(1.6)	96	(1.2)	348	382	437	571	630	663
Irland	501	(2.2)	85	(1.3)	359	391	445	559	610	640
Slowenien	501	(1.2)	92	(1.0)	357	384	434	566	624	655
Dänemark	500	(2.3)	82	(1.3)	363	393	444	556	607	635
Neuseeland	500	(2.2)	100	(1.2)	340	371	428	570	632	665
Tschechische Republik	499	(2.9)	95	(1.6)	344	377	432	566	621	653
Frankreich	495	(2.5)	97	(1.7)	330	365	429	565	621	652
Vereinigtes Königreich	494	(3.3)	95	(1.7)	336	371	429	560	616	648
Island	493	(1.7)	92	(1.3)	339	372	431	557	612	641
Norwegen	489	(2.7)	90	(1.3)	341	373	428	552	604	638
Portugal	487	(3.8)	94	(1.4)	333	363	421	554	610	640
Luxemburg	490	(1.1)	95	(0.9)	334	363	422	558	613	644
Italien	485	(2.0)	93	(1.1)	333	366	421	550	607	639
Spanien	484	(1.9)	88	(0.7)	339	370	424	546	597	626
Slowakische Republik	482	(3.4)	101	(2.5)	314	352	413	553	613	647
Vereinigte Staaten	481	(3.6)	90	(1.3)	339	368	418	543	600	634
Schweden	478	(2.3)	92	(1.3)	329	360	415	543	596	627
Ungarn	477	(3.2)	94	(2.4)	327	358	411	540	603	637
Israel	466	(4.7)	105	(1.8)	292	328	393	541	603	639
Griechenland	453	(2.5)	88	(1.3)	308	338	393	513	567	597
Türkei	448	(4.8)	91	(3.1)	313	339	382	507	577	614
Chile	423	(3.1)	81	(1.5)	299	323	365	476	532	563
Mexiko	413	(1.4)	74	(0.7)	295	320	362	462	510	539
OECD-Durchschnitt	494	(0.5)	92	(0.3)	343	375	430	568	614	645
OECD-Partnerstaaten										
Shanghai (China)	613	(3.3)	101	(2.3)	435	475	546	685	737	765
Singapur	573	(1.3)	105	(0.9)	393	432	501	650	707	737
Hongkong (China)	561	(3.2)	96	(1.9)	391	430	499	629	679	709
Chinesisch Taipeh	560	(3.3)	116	(1.9)	363	402	478	645	703	738
Macao (China)	538	(1.0)	94	(0.9)	379	415	476	605	657	685
Liechtenstein	535	(4.0)	95	(3.7)	370	403	470	606	656	680
Vietnam	511	(4.8)	86	(2.7)	371	401	454	568	623	654
Lettland	491	(2.8)	82	(1.5)	360	387	434	546	597	626
Russische Föderation	482	(3.0)	86	(1.6)	341	371	423	540	595	626
Litauen	479	(2.6)	89	(1.4)	334	364	418	540	596	627
Kroatien	471	(3.5)	88	(2.5)	334	360	408	531	589	623
Serbien	449	(3.4)	91	(2.2)	306	335	386	508	567	603
Rumänien	445	(3.8)	81	(2.2)	322	344	386	497	553	588
Zypern	440	(1.1)	93	(0.8)	287	320	376	503	561	595
Bulgarien	439	(4.0)	94	(2.2)	290	320	372	503	565	597
Vereinigte Arabische Emirate	434	(2.4)	90	(1.2)	297	323	370	494	555	591
Kasachstan	432	(3.0)	71	(1.8)	319	343	383	478	527	554
Thailand	427	(3.4)	82	(2.1)	302	328	372	476	535	575
Malaysia	421	(3.2)	81	(1.6)	294	319	363	474	530	562
Montenegro	410	(1.1)	83	(1.1)	280	306	352	465	520	552
Uruguay	409	(2.8)	89	(1.7)	267	297	347	470	526	558
Costa Rica	407	(3.0)	68	(1.8)	301	323	361	449	496	525
Albanien	394	(2.0)	91	(1.4)	236	278	338	454	510	540
Brasilien	391	(2.1)	78	(1.6)	275	298	337	440	495	530
Argentinien	388	(3.5)	77	(1.7)	264	292	337	440	488	514
Tunesien	388	(3.9)	78	(3.1)	267	292	334	437	488	523
Jordanien	386	(3.1)	78	(2.7)	263	290	335	435	485	514
Kolumbien	376	(2.9)	74	(1.7)	262	285	326	423	474	506
Katar	376	(0.8)	100	(0.7)	230	257	306	440	514	560
Indonesien	375	(4.0)	71	(3.3)	266	288	327	418	469	501
Peru	368	(3.7)	84	(2.2)	237	264	311	421	478	517

□ signifikant über dem OECD-Durchschnitt

□ nicht signifikant verschieden vom OECD-Durchschnitt

□ signifikant unter dem OECD-Durchschnitt

Tabelle 2: Mittelwerte, Streuungen und Perzentile für den mathematischen Inhaltsbereich *Veränderung und Beziehungen*

OECD-Staaten	Perzentile									
	M	(SE)	SD	(SE)	5%	10%	25%	75%	90%	95%
Korea	559	(5.2)	107	(2.7)	382	422	488	633	692	727
Japan	542	(4.0)	107	(2.4)	362	404	470	618	680	715
Estland	530	(2.3)	84	(1.1)	394	422	472	587	639	669
Schweiz	530	(3.4)	103	(1.6)	359	396	459	602	661	695
Kanada	525	(2.0)	94	(0.9)	367	403	461	591	647	679
Finnland	520	(2.6)	97	(2.3)	363	400	458	584	643	677
Niederlande	518	(3.9)	103	(3.2)	345	388	453	593	642	669
Deutschland	516	(3.8)	114	(3.4)	321	368	443	597	656	688
Belgien	513	(2.6)	116	(3.2)	312	362	443	596	653	684
Polen	509	(4.1)	100	(2.1)	347	380	440	578	641	677
Australien	509	(1.7)	104	(1.2)	339	375	437	581	645	680
Österreich	506	(3.4)	109	(2.7)	326	365	433	584	643	677
Irland	501	(2.6)	87	(1.5)	355	389	443	561	613	642
Neuseeland	501	(2.5)	112	(1.6)	319	356	422	578	646	686
Tschechische Republik	499	(3.5)	112	(3.3)	317	364	430	576	636	674
Slowenien	499	(1.1)	100	(1.0)	338	372	429	570	632	667
Frankreich	497	(2.7)	107	(2.4)	313	355	425	572	632	667
Vereinigtes Königreich	496	(3.4)	99	(1.8)	333	368	429	565	626	659
Dänemark	494	(2.7)	91	(1.3)	345	377	432	557	613	643
Vereinigte Staaten	488	(3.5)	95	(1.4)	339	368	421	552	614	649
Luxemburg	488	(1.0)	102	(1.0)	317	352	415	562	619	652
Island	487	(1.9)	100	(1.5)	318	355	420	557	614	647
Portugal	486	(4.1)	98	(1.4)	323	356	417	556	615	645
Spanien	482	(2.0)	93	(0.8)	326	361	420	547	600	630
Ungarn	481	(3.5)	100	(2.7)	320	352	411	550	614	651
Norwegen	478	(3.1)	102	(1.3)	306	346	409	547	608	644
Italien	477	(2.1)	100	(1.3)	310	348	410	546	604	638
Slowakische Republik	474	(4.0)	114	(2.9)	282	327	401	553	617	655
Schweden	469	(2.8)	107	(1.6)	291	331	397	544	606	641
Israel	462	(5.3)	117	(2.4)	266	308	382	545	613	651
Türkei	448	(5.0)	92	(3.1)	310	336	383	508	575	612
Griechenland	446	(3.2)	101	(1.6)	278	317	378	515	574	609
Chile	411	(3.5)	95	(1.6)	263	293	345	475	537	574
Mexiko	405	(1.6)	87	(0.8)	264	295	347	462	516	549
OECD-Durchschnitt	493	(0.6)	101	(0.4)	325	362	424	563	622	657
OECD-Partnerstaaten										
Shanghai (China)	624	(3.6)	112	(2.4)	431	473	547	704	764	797
Singapur	580	(1.5)	114	(0.9)	387	428	502	662	725	759
Hongkong (China)	564	(3.6)	103	(2.2)	380	426	497	636	691	723
Chinesisch Taipeh	561	(3.5)	121	(2.2)	355	398	476	648	714	752
Macao (China)	542	(1.2)	100	(1.1)	375	413	478	612	667	700
Liechtenstein	542	(4.0)	104	(3.6)	363	400	469	621	675	703
Vietnam	509	(5.1)	94	(2.7)	355	389	445	572	631	664
Lettland	496	(3.4)	92	(4.0)	347	381	434	558	613	642
Russische Föderation	491	(3.4)	93	(1.8)	338	371	428	553	611	644
Litauen	479	(3.2)	92	(1.6)	330	364	417	542	599	632
Kroatien	468	(4.2)	103	(2.8)	301	336	395	539	602	640
Rumänien	446	(3.9)	89	(2.4)	307	336	382	504	566	602
Vereinigte Arabische Emirate	442	(2.6)	95	(1.2)	294	325	376	505	570	607
Serbien	442	(4.1)	104	(2.7)	274	311	371	512	578	618
Zypern	440	(1.2)	102	(1.0)	272	310	371	509	572	608
Bulgarien	434	(4.5)	109	(2.5)	263	299	358	507	579	620
Kasachstan	433	(3.2)	84	(1.9)	298	327	375	489	541	573
Thailand	414	(3.9)	93	(2.3)	269	300	350	471	535	576
Costa Rica	402	(3.5)	81	(1.9)	273	300	348	454	506	538
Uruguay	401	(3.2)	105	(2.2)	230	267	331	472	537	576
Malaysia	401	(4.0)	92	(2.1)	258	287	337	461	524	561
Montenegro	399	(1.3)	93	(1.0)	253	282	333	462	521	556
Albanien	388	(2.1)	98	(1.4)	217	263	327	453	510	543
Jordanien	387	(3.7)	87	(2.7)	246	279	330	447	499	529
Tunesien	379	(4.5)	91	(3.0)	234	264	318	438	496	531
Argentinien	379	(4.2)	90	(1.9)	231	263	318	440	495	525
Brasilien	372	(2.7)	99	(1.9)	217	250	304	435	500	542
Indonesien	364	(4.3)	79	(3.4)	240	267	311	414	468	501
Katar	363	(0.9)	110	(0.7)	197	230	285	434	514	562
Kolumbien	357	(3.7)	91	(1.8)	214	244	295	415	475	513
Peru	349	(4.5)	101	(2.6)	191	224	280	415	482	525

□ signifikant über dem OECD-Durchschnitt

□ nicht signifikant verschieden vom OECD-Durchschnitt

□ signifikant unter dem OECD-Durchschnitt

Tabelle 3: Mittelwerte, Streuungen und Perzentile für den mathematischen Inhaltsbereich *Quantität*

OECD-Staaten	Perzentile									
	M	(SE)	SD	(SE)	5%	10%	25%	75%	90%	95%
Korea	537	(4.1)	94	(2.0)	377	416	477	604	654	682
Niederlande	532	(3.6)	97	(2.3)	365	398	463	604	653	682
Schweiz	531	(3.1)	96	(1.4)	369	404	467	598	652	684
Finnland	527	(1.9)	87	(1.0)	382	415	469	586	638	669
Estland	525	(2.2)	86	(1.2)	382	415	466	583	636	667
Belgien	519	(2.0)	104	(1.4)	341	381	447	594	650	681
Polen	519	(3.5)	89	(1.6)	375	406	457	579	634	664
Japan	518	(3.6)	94	(2.2)	359	395	456	584	638	670
Deutschland	517	(3.1)	100	(1.9)	348	384	449	588	643	674
Kanada	515	(2.2)	99	(1.0)	349	386	448	585	643	676
Österreich	510	(2.9)	91	(1.7)	358	391	446	576	627	656
Irland	505	(2.6)	92	(1.4)	350	386	443	569	624	653
Tschechische Republik	505	(3.0)	101	(2.0)	336	373	438	576	633	668
Slowenien	504	(1.2)	94	(1.0)	351	382	438	570	629	661
Dänemark	502	(2.4)	91	(1.3)	354	387	441	565	619	648
Australien	500	(1.9)	104	(1.3)	330	367	429	572	634	669
Neuseeland	499	(2.4)	103	(1.3)	331	365	426	572	634	667
Island	496	(1.9)	102	(1.5)	322	362	429	567	627	661
Frankreich	496	(2.6)	103	(1.8)	324	362	425	570	628	661
Luxemburg	495	(1.0)	100	(0.9)	326	362	424	567	623	656
Vereinigtes Königreich	494	(3.8)	102	(1.9)	325	362	424	567	625	658
Norwegen	492	(2.9)	95	(1.6)	335	372	429	556	613	648
Spanien	491	(2.3)	101	(1.0)	321	360	423	562	618	651
Italien	491	(2.0)	101	(1.0)	321	360	423	561	619	652
Slowakische Republik	486	(3.5)	105	(2.2)	312	350	414	560	621	658
Schweden	482	(2.5)	97	(1.3)	320	357	417	549	607	639
Portugal	481	(4.0)	96	(1.5)	321	355	415	550	604	636
Israel	480	(5.2)	116	(2.1)	284	327	398	563	629	667
Vereinigte Staaten	478	(3.9)	99	(1.7)	322	354	408	545	610	646
Ungarn	476	(3.4)	99	(2.2)	314	350	406	545	606	641
Griechenland	455	(3.0)	97	(1.6)	295	330	388	523	579	613
Türkei	442	(5.0)	97	(3.0)	295	324	373	506	576	613
Chile	421	(3.3)	90	(1.6)	280	310	359	482	541	575
Mexiko	414	(1.5)	87	(0.9)	271	304	355	472	526	559
OECD-Durchschnitt	495	(0.5)	97	(0.3)	334	369	429	563	620	653
OECD-Partnerstaaten										
Shanghai (China)	591	(3.2)	98	(2.4)	419	460	528	658	710	741
Singapur	569	(1.2)	104	(0.9)	390	428	500	642	699	731
Hongkong (China)	566	(3.4)	101	(2.0)	383	430	501	637	688	718
Chinesisch Taipeh	543	(3.1)	108	(1.8)	357	396	470	622	677	707
Liechtenstein	538	(4.1)	100	(3.6)	364	398	467	615	660	686
Macao (China)	531	(1.1)	92	(1.0)	375	411	469	595	646	675
Vietnam	509	(5.5)	93	(2.7)	354	391	446	571	629	662
Lettland	487	(2.9)	84	(1.5)	350	381	430	546	596	624
Litauen	483	(2.8)	93	(1.4)	331	363	420	547	605	637
Kroatien	480	(3.7)	93	(2.5)	332	363	414	543	603	637
Russische Föderation	478	(3.0)	93	(1.6)	326	360	417	540	598	632
Serbien	456	(3.7)	97	(2.6)	303	334	390	521	582	619
Rumänien	443	(4.5)	94	(2.5)	298	327	376	505	567	605
Bulgarien	443	(4.3)	102	(2.8)	280	313	373	513	576	612
Zypern	439	(1.1)	100	(1.1)	276	310	370	508	568	604
Vereinigte Arabische Emirate	431	(2.7)	101	(1.2)	273	304	360	500	567	603
Kasachstan	428	(3.5)	79	(2.1)	305	331	373	479	533	564
Thailand	419	(3.7)	88	(2.2)	282	311	359	473	534	573
Uruguay	411	(3.2)	98	(1.9)	250	284	344	478	539	572
Malaysia	409	(3.6)	94	(1.9)	263	291	343	471	536	572
Montenegro	409	(1.2)	88	(1.1)	269	298	349	467	523	556
Costa Rica	406	(3.6)	81	(2.4)	278	306	353	457	509	544
Brasilien	393	(2.5)	91	(1.6)	250	280	330	452	513	552
Argentinien	391	(3.7)	84	(2.2)	251	284	336	448	499	529
Albanien	386	(2.7)	101	(1.7)	206	257	326	453	511	543
Tunesien	378	(4.6)	91	(3.4)	233	264	316	437	493	530
Kolumbien	375	(3.4)	90	(2.1)	232	264	315	434	491	527
Katar	371	(0.9)	105	(0.7)	212	244	298	437	514	559
Jordanien	367	(3.4)	90	(2.3)	223	255	307	425	483	518
Peru	365	(4.1)	97	(2.4)	211	245	301	427	490	532
Indonesien	362	(4.7)	83	(3.5)	235	261	307	414	471	507

□ signifikant über dem OECD-Durchschnitt

□ nicht signifikant verschieden vom OECD-Durchschnitt

□ signifikant unter dem OECD-Durchschnitt

Tabelle 4: Mittelwerte, Streuungen und Perzentile für den mathematischen Inhaltsbereich *Raum und Form*

OECD-Staaten	Perzentile									
	M	(SE)	SD	(SE)	5%	10%	25%	75%	90%	95%
Korea	573	(5.2)	112	(2.4)	388	428	495	653	716	753
Japan	558	(3.7)	100	(2.4)	393	429	489	627	688	723
Schweiz	544	(3.1)	101	(1.7)	375	413	475	614	675	711
Polen	524	(4.2)	101	(2.2)	370	398	450	593	660	697
Estland	513	(2.5)	94	(1.1)	364	395	449	575	634	671
Kanada	510	(2.1)	95	(0.9)	355	388	444	576	636	670
Belgien	509	(2.4)	108	(1.5)	330	368	434	585	649	684
Niederlande	507	(3.5)	94	(2.3)	350	385	442	573	628	660
Deutschland	507	(3.2)	98	(1.9)	346	379	440	575	633	667
Finnland	507	(2.1)	90	(1.3)	361	393	446	567	624	658
Slowenien	503	(1.4)	99	(1.2)	345	379	433	572	636	671
Österreich	501	(3.1)	98	(2.2)	340	375	432	569	627	662
Tschechische Republik	499	(3.4)	102	(1.9)	331	369	428	569	630	666
Dänemark	497	(2.5)	84	(1.2)	357	388	441	553	604	633
Australien	497	(1.8)	102	(1.4)	334	368	425	564	630	669
Portugal	491	(4.2)	109	(1.9)	318	351	414	568	633	669
Neuseeland	491	(2.4)	100	(1.7)	334	366	421	558	624	663
Slowakische Republik	490	(4.1)	109	(2.7)	311	351	416	564	632	670
Frankreich	489	(2.7)	99	(1.9)	326	360	418	558	619	652
Island	489	(1.5)	88	(1.3)	339	373	430	549	604	634
Italien	487	(2.5)	106	(1.4)	316	354	415	559	627	665
Luxemburg	486	(1.0)	96	(1.1)	332	364	418	554	612	645
Norwegen	480	(3.3)	102	(1.4)	312	351	412	548	610	647
Irland	478	(2.6)	94	(1.4)	323	357	415	542	598	631
Spanien	477	(2.0)	94	(0.9)	324	357	412	542	599	631
Vereinigtes Königreich	475	(3.5)	99	(1.8)	313	347	407	542	605	641
Ungarn	474	(3.4)	96	(2.7)	325	354	406	536	604	643
Schweden	469	(2.5)	94	(1.6)	313	348	405	533	590	623
Vereinigte Staaten	463	(4.0)	96	(1.5)	314	342	396	527	591	631
Israel	449	(4.8)	105	(1.9)	278	314	376	522	586	622
Türkei	443	(5.5)	109	(3.8)	280	312	365	512	597	641
Griechenland	436	(2.6)	90	(1.4)	290	324	375	497	552	585
Chile	419	(3.2)	86	(1.5)	288	313	358	475	533	569
Mexiko	413	(1.6)	82	(0.9)	280	309	358	466	519	550
OECD-Durchschnitt	490	(0.5)	98	(0.3)	331	365	422	556	618	653
OECD-Partnerstaaten										
Shanghai (China)	649	(3.6)	114	(2.5)	445	493	575	728	787	822
Chinesisch Taipeh	592	(3.8)	136	(2.3)	362	407	494	693	764	803
Singapur	580	(1.5)	117	(1.1)	380	423	500	664	727	764
Hongkong (China)	567	(4.0)	107	(2.3)	382	422	495	642	701	734
Macao (China)	558	(1.4)	109	(1.0)	375	416	485	635	697	732
Liechtenstein	539	(4.5)	99	(4.3)	373	406	475	611	667	695
Vietnam	507	(5.1)	99	(2.8)	346	382	439	573	637	674
Lettland	497	(3.3)	88	(1.5)	356	386	437	556	611	645
Russische Föderation	496	(3.9)	95	(2.1)	344	376	430	560	622	657
Litauen	472	(3.1)	98	(1.7)	313	347	404	539	600	637
Kroatien	460	(3.9)	88	(3.4)	328	354	399	516	575	615
Kasachstan	450	(3.9)	85	(2.3)	317	344	391	506	562	595
Rumänien	447	(4.1)	91	(2.6)	306	335	383	505	567	607
Serbien	446	(3.9)	98	(2.5)	293	324	377	510	576	616
Bulgarien	442	(4.3)	95	(2.2)	291	321	376	506	569	604
Zypern	436	(1.1)	92	(1.0)	289	320	373	498	555	592
Malaysia	434	(3.4)	86	(1.8)	300	327	373	492	550	583
Thailand	432	(4.1)	95	(2.5)	287	316	367	490	558	603
Vereinigte Arabische Emirate	425	(2.4)	97	(1.4)	274	304	356	490	553	591
Albanien	418	(2.6)	106	(1.4)	240	283	351	486	551	590
Uruguay	413	(3.1)	94	(2.1)	262	292	347	477	536	569
Montenegro	412	(1.1)	80	(1.1)	287	313	357	464	518	552
Costa Rica	397	(3.2)	72	(2.2)	289	310	348	442	489	524
Jordanien	385	(3.1)	81	(2.9)	258	286	332	437	488	520
Argentinien	385	(3.5)	78	(1.7)	259	287	334	436	485	514
Indonesien	383	(4.2)	82	(2.8)	252	281	328	435	487	524
Tunesien	382	(3.9)	85	(3.0)	252	278	324	436	491	530
Brasilien	381	(2.0)	81	(1.8)	255	282	327	431	485	521
Katar	380	(1.0)	101	(0.7)	229	259	310	443	517	563
Peru	370	(4.1)	93	(2.4)	221	256	309	429	489	528
Kolumbien	369	(3.5)	81	(1.9)	241	269	315	420	474	508

□ signifikant über dem OECD-Durchschnitt

□ nicht signifikant verschieden vom OECD-Durchschnitt

■ signifikant unter dem OECD-Durchschnitt

Tabelle 5: Mittelwerte, Streuungen und Perzentile für den mathematischen Inhaltsbereich *Unsicherheit und Daten*

OECD-Staaten	Perzentile									
	M	(SE)	SD	(SE)	5%	10%	25%	75%	90%	95%
Korea	538	(4.2)	97	(1.9)	374	413	473	606	661	690
Niederlande	532	(3.8)	99	(2.6)	367	399	461	606	659	687
Japan	528	(3.5)	90	(2.0)	376	410	468	591	642	671
Schweiz	522	(3.2)	97	(1.6)	357	396	457	589	644	677
Finnland	519	(2.4)	91	(1.4)	367	403	460	580	634	664
Polen	517	(3.5)	87	(1.9)	374	403	456	578	630	660
Kanada	516	(1.8)	90	(0.9)	367	401	456	579	632	661
Estland	510	(2.0)	81	(1.1)	378	408	456	565	615	645
Deutschland	509	(3.0)	101	(1.8)	340	376	439	581	639	669
Irland	509	(2.5)	88	(1.4)	361	395	450	569	619	648
Belgien	508	(2.5)	110	(2.3)	323	366	435	585	647	681
Australien	508	(1.5)	97	(1.1)	349	384	441	575	633	666
Neuseeland	506	(2.6)	106	(1.6)	332	370	432	580	644	680
Dänemark	505	(2.4)	84	(1.3)	363	396	448	564	613	641
Vereinigtes Königreich	502	(3.0)	97	(1.6)	341	378	436	570	626	659
Österreich	499	(2.7)	95	(1.9)	339	374	433	567	618	647
Slowenien	496	(1.2)	92	(0.9)	347	378	430	562	619	648
Norwegen	497	(3.0)	91	(2.1)	345	381	437	558	613	644
Island	496	(1.8)	98	(1.7)	329	365	430	565	620	652
Frankreich	492	(2.7)	103	(1.8)	317	355	421	567	622	653
Tschechische Republik	488	(2.8)	92	(2.0)	338	371	426	551	606	638
Vereinigte Staaten	488	(3.5)	89	(1.5)	344	374	426	551	604	635
Spanien	487	(2.3)	94	(1.1)	329	367	425	552	605	635
Portugal	486	(3.8)	91	(1.5)	334	366	422	550	604	632
Luxemburg	483	(1.0)	100	(1.0)	319	352	411	555	613	645
Schweden	483	(2.5)	93	(1.3)	327	363	420	547	603	634
Italien	482	(2.0)	96	(1.1)	321	359	418	549	605	637
Ungarn	476	(3.3)	94	(2.5)	318	353	412	541	599	632
Slowakische Republik	472	(3.6)	100	(2.5)	305	343	405	541	599	633
Israel	465	(4.7)	108	(2.0)	283	323	391	542	605	641
Griechenland	460	(2.6)	87	(1.4)	312	347	402	519	572	602
Türkei	447	(4.6)	91	(2.7)	307	336	383	506	573	610
Chile	430	(2.9)	76	(1.4)	309	335	378	481	531	561
Mexiko	413	(1.2)	67	(0.7)	303	328	368	457	499	524
OECD-Durchschnitt	493	(0.5)	93	(0.3)	338	373	430	558	613	644
OECD-Partnerstaaten										
Shanghai (China)	592	(3.0)	96	(1.9)	427	464	528	660	712	741
Singapur	559	(1.5)	104	(0.8)	384	421	487	634	692	725
Hongkong (China)	553	(3.0)	91	(1.8)	392	430	494	617	666	694
Chinesisch Taipeh	549	(3.2)	108	(2.1)	364	403	474	627	684	716
Liechtenstein	526	(3.9)	97	(3.3)	359	390	456	599	648	679
Macao (China)	525	(1.1)	89	(0.9)	374	409	467	587	637	666
Vietnam	519	(4.5)	79	(2.4)	385	416	466	574	619	646
Lettland	478	(2.8)	79	(1.2)	350	378	424	533	581	607
Litauen	474	(2.7)	91	(1.3)	324	357	412	536	593	624
Kroatien	468	(3.5)	90	(2.2)	324	354	405	529	587	619
Russische Föderation	463	(3.3)	85	(1.5)	323	355	406	521	572	601
Serbien	448	(3.3)	86	(1.9)	310	341	391	505	559	592
Zypern	442	(1.1)	90	(1.1)	292	326	381	504	557	589
Rumänien	437	(3.3)	76	(1.8)	314	340	384	487	536	567
Thailand	433	(3.1)	77	(2.1)	312	339	383	480	531	565
Vereinigte Arabische Emirate	432	(2.4)	86	(1.1)	296	324	372	489	546	581
Bulgarien	432	(3.9)	90	(2.4)	285	318	370	493	549	581
Malaysia	422	(3.0)	81	(1.6)	287	318	367	476	526	557
Montenegro	415	(1.0)	85	(1.0)	279	308	357	470	526	559
Costa Rica	414	(2.9)	63	(1.5)	315	336	372	455	496	521
Kasachstan	414	(2.6)	58	(1.3)	318	339	374	453	490	511
Uruguay	407	(2.7)	81	(1.9)	279	305	351	460	515	548
Brasilien	402	(2.0)	71	(1.4)	289	314	355	447	495	524
Tunesien	399	(3.6)	71	(2.7)	287	311	352	444	489	520
Jordanien	394	(3.2)	77	(2.8)	266	297	346	443	489	517
Argentinien	389	(3.5)	81	(1.9)	258	287	335	443	492	521
Kolumbien	388	(2.4)	67	(1.5)	280	303	344	431	473	501
Albanien	386	(2.4)	96	(1.7)	219	264	327	450	505	537
Indonesien	384	(3.9)	68	(3.0)	276	300	339	427	470	501
Katar	382	(0.8)	100	(0.7)	234	263	311	445	518	565
Peru	373	(3.3)	75	(2.0)	252	279	323	422	470	501

□ signifikant über dem OECD-Durchschnitt

□ nicht signifikant verschieden vom OECD-Durchschnitt

□ signifikant unter dem OECD-Durchschnitt

Tabelle 6: Mittelwerte, Streuungen und Perzentile naturwissenschaftlicher Kompetenz

OECD-Staaten	Perzentile									
	M	(SE)	SD	(SE)	5%	10%	25%	75%	90%	95%
Japan	547	(3.6)	96	(2.2)	379	421	485	614	664	693
Finnland	545	(2.2)	93	(1.2)	386	424	486	609	662	692
Estland	541	(1.9)	80	(1.1)	409	439	487	597	645	672
Korea	538	(3.7)	82	(1.8)	396	431	485	595	639	664
Polen	526	(3.1)	86	(1.5)	382	415	467	584	637	668
Kanada	525	(1.9)	91	(0.9)	370	407	467	588	639	670
Deutschland	524	(3.0)	95	(2.0)	361	397	461	592	642	671
Niederlande	522	(3.5)	95	(2.2)	357	393	458	591	641	667
Irland	522	(2.5)	91	(1.6)	366	404	462	586	637	666
Australien	521	(1.8)	100	(1.0)	353	391	453	592	650	682
Neuseeland	516	(2.1)	105	(1.4)	339	377	444	591	649	682
Schweiz	515	(2.7)	91	(1.1)	358	394	455	579	630	658
Slowenien	514	(1.3)	91	(1.2)	364	397	451	578	631	661
Vereinigtes Königreich	514	(3.4)	100	(1.8)	344	384	448	584	639	672
Tschechische Republik	508	(3.0)	91	(2.1)	356	392	449	572	622	650
Belgien	505	(2.1)	101	(1.4)	326	369	439	579	630	658
Österreich	506	(2.7)	92	(1.6)	350	383	442	571	623	650
Frankreich	499	(2.6)	100	(2.2)	323	366	433	570	622	651
Dänemark	498	(2.7)	93	(1.7)	338	378	438	563	615	644
Vereinigige Staaten	497	(3.8)	94	(1.5)	344	377	431	563	619	652
Spanien	496	(1.8)	86	(0.9)	349	384	440	557	605	632
Norwegen	495	(3.1)	100	(1.9)	325	365	429	564	620	651
Ungarn	494	(2.9)	90	(1.9)	345	376	432	558	610	639
Italien	494	(1.9)	93	(1.1)	336	371	431	559	611	641
Luxemburg	491	(1.3)	103	(1.0)	318	355	419	566	624	655
Portugal	489	(3.7)	89	(1.6)	337	372	430	551	602	630
Schweden	485	(3.0)	100	(1.5)	314	354	419	554	611	643
Island	478	(2.1)	99	(1.5)	310	348	413	548	603	635
Slowakische Republik	471	(3.6)	101	(2.8)	300	339	403	542	599	632
Israel	470	(5.0)	108	(2.1)	286	328	396	548	608	640
Griechenland	467	(3.1)	88	(1.5)	317	352	408	528	578	608
Türkei	463	(3.9)	80	(1.9)	339	363	407	518	573	602
Chile	445	(2.9)	80	(1.5)	317	343	388	500	552	581
Mexiko	415	(1.3)	71	(0.9)	300	325	368	462	505	532
OECD-Durchschnitt	501	(0.5)	93	(0.3)	344	380	439	566	619	648
OECD-Partnerstaaten										
Shanghai (China)	580	(3.0)	82	(1.8)	435	472	527	639	681	704
Hongkong (China)	555	(2.6)	83	(1.8)	403	446	505	613	655	679
Singapur	551	(1.5)	104	(1.2)	374	412	480	627	681	714
Vietnam	528	(4.3)	77	(2.3)	398	428	478	580	625	652
Liechtenstein	525	(3.5)	86	(4.1)	383	408	464	588	635	658
Chinesisch Taipeh	523	(2.3)	83	(1.4)	379	411	469	582	626	652
Macao (China)	521	(0.8)	79	(0.7)	383	416	469	575	619	643
Lettland	502	(2.8)	79	(1.4)	370	400	449	557	603	628
Litauen	496	(2.6)	86	(1.7)	352	383	438	555	605	634
Kroatien	491	(3.1)	85	(1.8)	350	380	433	551	602	630
Russische Föderation	486	(2.9)	85	(1.3)	347	377	428	544	596	627
Vereinigte Arabische Emirate	448	(2.8)	94	(1.1)	299	328	382	512	572	605
Bulgarien	446	(4.8)	102	(2.5)	280	315	374	519	580	612
Serbien	445	(3.4)	87	(1.9)	303	333	385	504	558	590
Thailand	444	(2.9)	76	(1.7)	323	349	392	494	544	575
Rumänien	439	(3.3)	79	(2.0)	316	340	383	492	543	573
Zypern	438	(1.2)	97	(1.1)	274	313	373	503	561	594
Costa Rica	429	(2.9)	71	(1.6)	315	341	382	476	520	546
Kasachstan	425	(3.0)	74	(1.5)	303	330	375	475	521	547
Malaysia	420	(3.0)	79	(1.4)	293	319	365	473	521	550
Uruguay	416	(2.8)	95	(1.7)	256	293	352	480	538	572
Montenegro	410	(1.1)	85	(1.0)	274	302	352	468	522	552
Jordanien	409	(3.1)	83	(2.0)	271	303	355	466	514	542
Argentinien	406	(3.9)	86	(2.2)	262	297	350	464	513	543
Brasilien	405	(2.1)	79	(1.4)	280	306	351	456	507	538
Kolumbien	399	(3.1)	76	(1.6)	273	302	347	449	497	525
Tunesien	398	(3.5)	79	(1.9)	267	296	345	452	497	527
Albanien	397	(2.4)	99	(1.8)	221	271	340	464	517	549
Katar	384	(0.7)	106	(0.7)	222	254	309	453	530	573
Indonesien	382	(3.8)	68	(2.3)	271	297	336	427	471	497
Peru	373	(3.6)	78	(1.9)	248	275	321	425	475	504

□ signifikant über dem OECD-Durchschnitt

□ nicht signifikant verschieden vom OECD-Durchschnitt

□ signifikant unter dem OECD-Durchschnitt

Tabelle 7: Mittelwerte, Streuungen und Perzentile der Lesekompetenz

OECD-Staaten	Perzentile									
	M	(SE)	SD	(SE)	5%	10%	25%	75%	90%	95%
Japan	538	(3.7)	99	(2.3)	364	409	475	607	658	689
Korea	536	(3.9)	87	(2.0)	382	424	483	596	640	665
Finnland	524	(2.4)	95	(1.3)	359	399	463	590	639	669
Irland	523	(2.6)	86	(1.7)	373	410	469	582	631	659
Kanada	523	(1.9)	92	(0.9)	363	403	464	587	638	667
Polen	518	(3.1)	87	(1.6)	366	404	461	579	626	655
Estland	516	(2.0)	80	(1.2)	381	412	463	571	618	645
Neuseeland	512	(2.4)	106	(1.6)	332	373	443	586	645	679
Australien	512	(1.6)	97	(1.0)	347	386	448	579	634	664
Niederlande	511	(3.5)	93	(3.0)	349	386	451	579	625	650
Belgien	509	(2.2)	103	(1.7)	324	372	444	583	635	663
Schweiz	509	(2.6)	90	(1.1)	352	388	451	573	622	648
Deutschland	508	(2.8)	91	(1.7)	346	384	447	574	621	646
Frankreich	505	(2.8)	109	(2.3)	312	358	435	584	639	669
Norwegen	504	(3.2)	100	(1.9)	330	375	442	573	627	658
Vereinigtes Königreich	499	(3.5)	97	(2.3)	330	372	438	567	619	650
Vereinigte Staaten	498	(3.7)	92	(1.6)	342	378	436	561	614	646
Dänemark	496	(2.6)	86	(2.2)	347	385	442	555	602	629
Tschechische Republik	493	(2.9)	89	(1.9)	344	378	434	554	604	634
Italien	490	(2.0)	97	(0.9)	317	359	427	559	609	636
Österreich	490	(2.8)	92	(1.8)	329	365	427	557	603	629
Ungarn	488	(3.2)	92	(1.9)	327	363	427	555	603	630
Spanien	488	(1.9)	92	(1.1)	327	367	430	552	601	630
Luxemburg	488	(1.5)	105	(1.0)	304	347	418	564	620	651
Portugal	488	(3.8)	94	(1.9)	320	362	429	554	604	631
Israel	486	(5.0)	114	(2.5)	282	329	414	568	624	656
Schweden	483	(3.0)	107	(1.8)	297	343	416	558	614	647
Island	483	(1.8)	98	(1.4)	308	352	422	551	602	631
Slowenien	481	(1.2)	92	(0.9)	324	362	420	548	598	626
Griechenland	477	(3.3)	99	(2.1)	302	346	416	545	597	626
Türkei	475	(4.2)	86	(2.4)	335	365	417	534	588	620
Slowakische Republik	463	(4.2)	104	(3.3)	274	321	396	538	591	621
Chile	441	(2.9)	78	(1.4)	310	339	388	496	541	567
Mexiko	424	(1.5)	80	(1.0)	288	319	370	479	525	552
OECD-Durchschnitt	496	(0.5)	94	(0.3)	332	372	435	563	613	642
OECD-Partnerstaaten										
Shanghai (China)	570	(2.9)	80	(1.8)	431	463	518	626	667	690
Hongkong (China)	545	(2.8)	85	(1.8)	391	430	493	604	648	672
Singapur	542	(1.4)	101	(1.2)	369	408	475	614	668	698
Chinesisch Taipeh	523	(3.0)	91	(1.8)	361	399	467	587	633	659
Liechtenstein	516	(4.1)	88	(4.2)	360	391	452	584	630	649
Macao (China)	509	(0.9)	82	(0.7)	366	400	457	566	611	637
Vietnam	508	(4.4)	74	(2.6)	379	411	462	559	599	623
Lettland	489	(2.4)	85	(1.7)	341	375	434	548	593	619
Kroatien	485	(3.3)	86	(2.1)	337	370	427	546	593	622
Litauen	477	(2.5)	86	(1.5)	331	363	419	538	585	612
Russische Föderation	475	(3.0)	91	(1.5)	323	359	415	537	592	623
Zypern	449	(1.2)	111	(1.3)	249	297	378	528	583	616
Serbien	446	(3.4)	93	(2.0)	290	325	384	509	566	596
Vereinigte Arabische Emirate	442	(2.5)	95	(1.1)	281	316	376	508	562	595
Thailand	441	(3.1)	78	(1.8)	310	341	389	494	541	569
Costa Rica	441	(3.5)	74	(1.6)	315	344	391	490	536	563
Rumänien	438	(4.0)	90	(2.0)	290	322	375	501	555	586
Bulgarien	436	(6.0)	119	(2.8)	233	275	353	523	585	619
Montenegro	422	(1.2)	92	(1.3)	267	301	360	487	540	571
Uruguay	411	(3.2)	96	(2.0)	248	285	348	477	534	564
Brasilien	410	(2.1)	85	(1.2)	271	302	353	468	520	552
Tunesien	404	(4.5)	88	(2.5)	252	286	346	466	515	543
Kolumbien	403	(3.4)	84	(1.9)	262	295	348	460	509	540
Jordanien	399	(3.6)	91	(2.5)	237	280	343	462	510	537
Malaysia	398	(3.3)	84	(1.5)	255	288	343	457	503	530
Indonesien	396	(4.2)	75	(2.7)	270	299	346	447	492	517
Argentinien	396	(3.7)	96	(2.3)	233	274	332	462	516	549
Albanien	394	(3.2)	116	(2.0)	189	247	325	473	536	572
Kasachstan	393	(2.7)	74	(1.4)	268	297	344	444	487	511
Katar	388	(0.8)	113	(0.8)	203	242	310	465	535	575
Peru	384	(4.3)	94	(2.3)	231	263	319	447	504	540

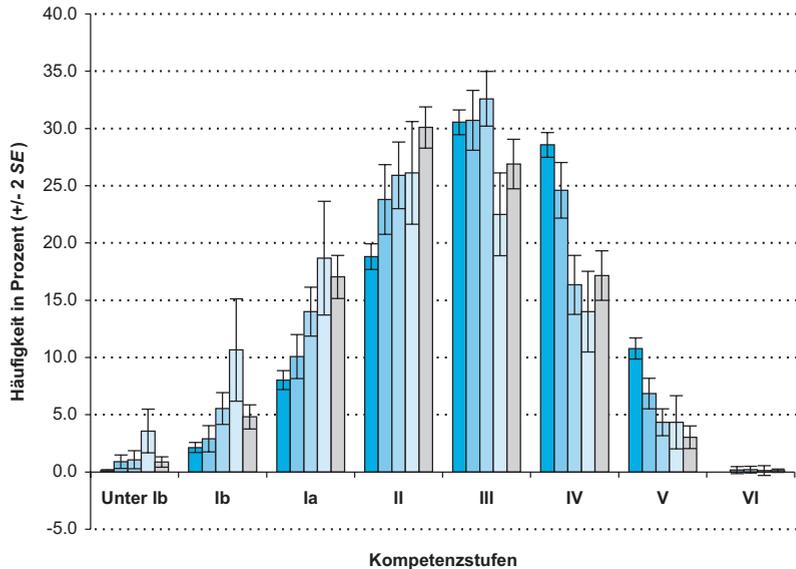
□ signifikant über dem OECD-Durchschnitt

□ nicht signifikant verschieden vom OECD-Durchschnitt

□ signifikant unter dem OECD-Durchschnitt

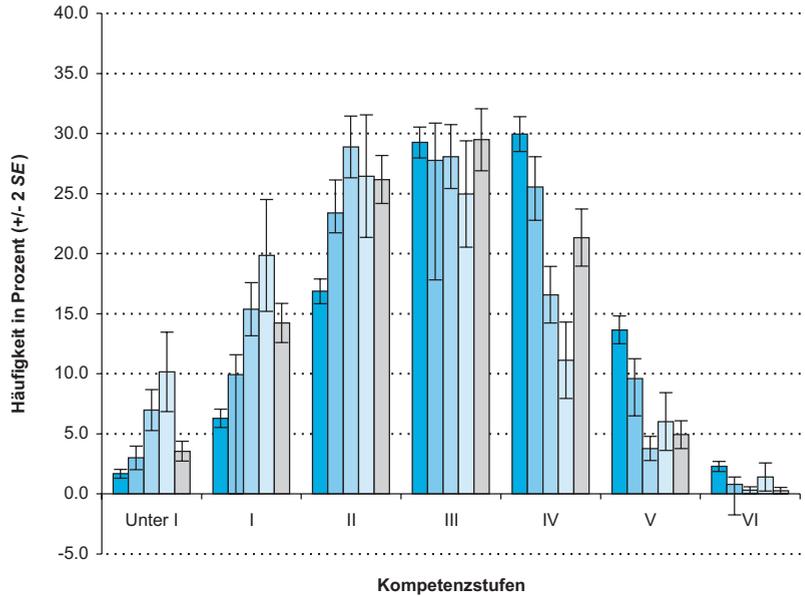
Tabelle 8: Mittlere Kompetenz von Jugendlichen mit Zuwanderungshintergrund für Mathematik, Lesen und Naturwissenschaften

	Ohne Zuwanderungshintergrund		Mit Zuwanderungshintergrund							
			Insgesamt Zuwanderung		Ein Elternteil im Ausland geboren		Zweite Generation		Erste Generation	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)
Mathematik	531	(3.4)	485	(4.1)	504	(5.4)	476	(5.3)	461	(9.0)
Lesen	525	(3.2)	486	(4.2)	504	(5.0)	481	(5.7)	453	(8.8)
Naturwissenschaften	543	(3.3)	492	(4.6)	517	(5.1)	478	(6.1)	468	(9.6)



	Unter Ib	Ib	Ia	II	III	IV	V	VI
	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)
Ohne Zuwanderungshintergrund	0.1 (0.1)	2.1 (0.4)	8.0 (0.8)	18.8 (1.1)	30.5 (1.1)	28.6 (1.1)	10.8 (0.9)	1.0 (0.4)
Generationsstatus								
Ein Elternteil im Ausland geboren	0.9 (0.6)	2.9 (1.1)	10.1 (1.9)	23.8 (3.0)	30.7 (2.6)	24.6 (2.4)	6.8 (1.3)	0.2 (0.3)
Zweite Generation	1.1 (0.8)	5.5 (1.4)	14.0 (2.1)	25.9 (2.9)	32.6 (2.4)	16.3 (2.6)	4.3 (1.2)	0.2 (0.3)
Erste Generation	3.6 (1.9)	10.7 (4.5)	18.7 (5.0)	26.1 (4.5)	22.5 (3.6)	14.0 (3.5)	4.3 (2.3)	0.1 (0.4)
Nicht zuzuordnen	0.9 (0.5)	4.8 (1.1)	17.0 (1.9)	30.1 (1.8)	26.9 (2.2)	17.1 (2.2)	3.0 (1.0)	0.1 (0.1)

Abbildung 1: Kompetenzstufen von fünfzehnjährigen Schülerinnen und Schülern mit Zuwanderungshintergrund für Lesen



	Unter I	I	II	III	IV	V	VI
	% (SE)	% (SE)					
Ohne Zuwanderungshintergrund	1.7 (0.4)	6.3 (0.8)	16.9 (1.0)	29.3 (1.3)	29.9 (1.4)	13.7 (1.2)	2.3 (0.4)
Generationsstatus							
Ein Elternteil im Ausland geboren	3.0 (1.0)	9.9 (1.7)	23.4 (2.7)	27.8 (3.1)	25.5 (2.5)	9.6 (1.7)	0.8 (0.6)
Zweite Generation	7.0 (1.7)	15.4 (2.2)	28.9 (2.6)	28.1 (2.7)	16.6 (2.3)	3.8 (1.0)	0.3 (0.3)
Erste Generation	10.2 (3.3)	19.9 (4.7)	26.4 (5.1)	25.0 (4.4)	11.1 (3.2)	6.0 (2.4)	1.4 (1.2)
Nicht zuzuordnen	3.6 (0.8)	14.2 (1.6)	26.2 (2.0)	29.5 (2.6)	21.3 (2.4)	4.9 (1.2)	0.3 (0.3)

Abbildung 2: Kompetenzstufen von fünfzehnjährigen Schülerinnen und Schülern mit Zuwanderungshintergrund für Naturwissenschaften

Abbildungsverzeichnis

Abbildung 1.1:	Perzentilband und Normalverteilungskurve	40
Abbildung 2.1:	Mathematischer Modellierungskreislauf (OECD, 2013, S. 26)	56
Abbildung 2.2:	Beispielaufgabe „MP3-Spieler“	62
Abbildung 2.3:	Beispielaufgabe „Die Radfahrerin Heike“	63
Abbildung 2.4:	Beispielaufgabe „Kabelfernsehen“	66
Abbildung 2.5:	Beispielaufgabe „Ölteppich“	68
Abbildung 2.6:	Perzentilbänder mathematischer Kompetenz in den OECD-Staaten	71
Abbildung 2.7:	Prozentuale Anteile von Schülerinnen und Schülern auf Kompetenzstufe I oder darunter beziehungsweise auf Kompetenzstufe V oder VI	74
Abbildung 2.8:	Mittelwerte mathematischer Kompetenz von Jungen und Mädchen in den OECD-Staaten	76
Abbildung 2.9:	Perzentilbänder für den Inhaltsbereich <i>Veränderung und Beziehungen</i> in den OECD-Staaten	78
Abbildung 2.10:	Perzentilbänder für den Inhaltsbereich <i>Quantität</i> in den OECD-Staaten	80
Abbildung 2.11:	Perzentilbänder für den Inhaltsbereich <i>Raum und Form</i> in den OECD-Staaten	81
Abbildung 2.12:	Perzentilbänder für den Inhaltsbereich <i>Unsicherheit und Daten</i> in den OECD-Staaten	83
Abbildung 2.13:	Mittelwerte mathematischer Kompetenz in OECD-Staaten in PISA 2003 und PISA 2012	85
Abbildung 2.14:	Perzentilbänder mathematischer Kompetenz in Deutschland nach Schulart	87
Abbildung 2.15:	Prozentuale Anteile von Schülerinnen und Schülern auf den Stufen der mathematischen Kompetenz in Deutschland	89
Abbildung 2.16:	Mathematische Kompetenz in Deutschland und an Gymnasien von PISA 2003 bis PISA 2012	91
Abbildung 2.17:	Mathematische Kompetenz von Jungen und Mädchen in Deutschland von PISA 2003 bis PISA 2012	92
Abbildung 3.1:	Emotionale und motivationale Orientierungen im internationalen Vergleich	107
Abbildung 3.2:	Mathematikbezogene Selbstbilder im internationalen Vergleich	110
Abbildung 3.3:	Mathematikbezogene Einstellungen und Verhaltensweisen im internationalen Vergleich	112
Abbildung 3.4:	Veränderung der Schülermerkmale in Deutschland zwischen den Jahren 2003 und 2012	114
Abbildung 4.1	Aufgabenbeispiele für ein Item zur Erfassung des Vorkommens außer- mathematischer Anwendungsaufgaben im Unterricht	132
Abbildung 4.2:	Basisdimensionen der Unterrichtsqualität im internationalen Vergleich	139
Abbildung 4.3:	Wahrgenommene Unterrichtspraktiken im internationalen Vergleich	140
Abbildung 4.4:	Mathematische Begriffe und Aufgabenstellungen im internationalen Vergleich	141
Abbildung 4.5:	Profillinien der latenten Klassen für die Unterrichtswahrnehmung der Fünfzehnjährigen in Deutschland	146
Abbildung 5.1:	Varianzanteile der Schülerkompetenz (Mathematik) zwischen den Schulen, nach Schulart	162
Abbildung 6.1:	Die PISA-Rahmenkonzeption naturwissenschaftlicher Grundbildung	193
Abbildung 6.2:	Beispielaufgabe „Saurer Regen“	196
Abbildung 6.3:	Perzentilbänder naturwissenschaftlicher Kompetenz in den OECD-Staaten	199

Abbildung 6.4:	Prozentuale Anteile von Schülerinnen und Schülern auf Kompetenzstufe I oder darunter beziehungsweise auf Kompetenzstufe V oder VI.....	202
Abbildung 6.5:	Mittelwerte naturwissenschaftlicher Kompetenz von Mädchen und Jungen in den OECD-Staaten	204
Abbildung 6.6:	Perzentilbänder naturwissenschaftlicher Kompetenz in Deutschland nach Schulart.....	207
Abbildung 6.7:	Prozentuale Anteile von Schülerinnen und Schülern auf den Stufen der naturwissenschaftlichen Kompetenz in Deutschland nach Schulart	208
Abbildung 6.8:	Naturwissenschaftliche Kompetenz in Deutschland von PISA 2006 bis PISA 2012.....	209
Abbildung 7.1:	Aufgabenbeispiel „Das Schauspiel sei das Werkzeug“	226
Abbildung 7.2:	Perzentilbänder der Lesekompetenz in den OECD-Staaten.....	228
Abbildung 7.3:	Prozentuale Anteile von Schülerinnen und Schülern auf Kompetenzstufe Ia, Kompetenzstufe Ib oder darunter sowie auf Kompetenzstufe V und Kompetenzstufe VI.....	231
Abbildung 7.4:	Mittelwerte der Lesekompetenz von Mädchen und Jungen in den OECD-Staaten.....	233
Abbildung 7.5:	Perzentilbänder für die Lesekompetenz in Deutschland nach Schulart und für die Gesamtstichprobe.....	236
Abbildung 7.6:	Prozentuale Anteile der Schülerinnen und Schüler in Deutschland auf den Stufen der Lesekompetenz nach Schulart	237
Abbildung 7.7:	Prozentuale Anteile der Mädchen und Jungen in Deutschland auf den Stufen der Lesekompetenz.....	238
Abbildung 7.8:	Prozentuale Anteile von Mädchen und Jungen in Deutschland bei PISA 2009 und 2012 auf Kompetenzstufe Ia, Ib oder darunter sowie auf Kompetenzstufe V und VI der Gesamtskala Lesen	239
Abbildung 7.9:	Veränderung der mittleren Lesekompetenz in Deutschland von 2000 bis 2012.....	240
Abbildung 7.10:	Veränderung der mittleren Lesekompetenz in Deutschland von 2000 bis 2012 für Mädchen und Jungen.....	241
Abbildung 8.1:	Verteilung des sozioökonomischen Status (HISEI) in den OECD-Staaten.....	251
Abbildung 8.2:	Mathematikkompetenz und Varianzaufklärung (R^2) durch den ökonomischen, sozialen und kulturellen Status (ESCS)	257
Abbildung 9.1:	Kompetenzstufen von fünfzehnjährigen Schülerinnen und Schülern mit Zuwanderungshintergrund für Mathematik.....	299
Abbildung 10.1:	Schritte der PISA-Stichprobenziehung	311
Abbildung 10.2:	Stichprobendesign PISA 2012 Deutschland.....	317
Abbildung 10.3:	Item Characteristic Curve (ICC)	327
Abbildung 10.4:	Längsschnittliche Zuordnung der Schwerpunktdomänen in PISA seit 2000.....	342

Anhang

Abbildung 1:	Kompetenzstufen von fünfzehnjährigen Schülerinnen und Schülern mit Zuwanderungshintergrund für Lesen.....	356
Abbildung 2:	Kompetenzstufen von fünfzehnjährigen Schülerinnen und Schülern mit Zuwanderungshintergrund für Naturwissenschaften	357

Tabellenverzeichnis

Tabelle 1.1	An PISA 2012 teilnehmende Staaten.....	22
Tabelle 1.2	Organisation der theoretischen Rahmenkonzeptionen der drei kognitiven Domänen in PISA 2012.....	23
Tabelle 1.3	Testheftdesign von PISA 2012	25
Tabelle 1.4	Zusammensetzung der Stichprobe in den an PISA 2012 teilnehmenden Schulen.....	28
Tabelle 1.5	Zielpopulation und Ausschöpfungsgrad bei PISA 2012 in Deutschland.....	29
Tabelle 1.6	Untersuchungsbeteiligung der Fünfzehnjährigen nach Schulart und Klassenstufe.....	31
Tabelle 2.1:	Zusammenhang zwischen mathematischen Prozessen und fundamentalen mathematischen Fähigkeiten.....	55
Tabelle 2.2:	Anteile der Aufgaben nach Inhalten, Prozessen und Kontexten an der maximal möglichen Punktezahl für Mathematik.....	58
Tabelle 2.3:	Stufen mathematischer Kompetenz in PISA 2012	61
Tabelle 2.4:	Prozentuale Anteile von Schülerinnen und Schülern auf den Stufen mathematischer Kompetenz in Deutschland nach Schulart.....	88
Tabelle 3.1:	Skalen der Schülermerkmale bei PISA 2012 mit Beispielitems.....	103
Tabelle 3.2:	Veränderung der mathematikbezogenen Selbstwirksamkeitserwartungen	115
Tabelle 4.1:	Skalen der Unterrichtsmerkmale bei PISA 2012 mit Beispielitems.....	131
Tabelle 4.2:	Prozentuale Antworthäufigkeiten für Beispielitems der Skalen zu den Basisdimensionen der Unterrichtsqualität	134
Tabelle 4.3:	Prozentuale Antworthäufigkeiten für Beispielitems der Skalen zu den Unterrichtspraktiken	136
Tabelle 4.4:	Prozentuale Antworthäufigkeiten für Beispielitems der Skalen zu den mathematischen Begriffen und Aufgabenstellungen.....	137
Tabelle 4.5:	Prozentuale Antworthäufigkeiten für die außermathematische Anwendungsaufgabe	142
Tabelle 4.6:	Prozentuale Verteilung der drei Unterrichtsmuster in Deutschland und den europäischen Vergleichsstaaten.....	147
Tabelle 4.7:	Zusammenhang von Unterrichtsmustern mit mathematischer Kompetenz, Freude und Interesse an Mathematik sowie mathematikbezogene Selbstwirksamkeitserwartungen am Gymnasium	149
Tabelle 5.1:	Varianz der mathematischen Kompetenz im internationalen Vergleich.....	161
Tabelle 5.2:	Klassenstufe der PISA-Kohorte in den OECD-Staaten.....	164
Tabelle 5.3:	Prozentualer Anteil der Schülerinnen und Schüler, die mindestens eine Klasse wiederholt haben	166
Tabelle 5.4:	Anteile der Fünfzehnjährigen in Prozent, Verteilung auf Klassenstufen und Klassenwiederholungen.....	168
Tabelle 5.5:	Anzahl der Zeitstunden pro Woche, die in OECD-Staaten mit Unterricht in Mathematik, Naturwissenschaften oder der Testsprache verbracht werden.....	169
Tabelle 5.6:	Durchschnittliche Klassengröße in der Sekundarstufe I.....	172
Tabelle 5.7:	Instrumente der Qualitätssicherung und -entwicklung an Schulen	174
Tabelle 5.8:	Beeinträchtigungen des Lernens durch Schülerabsenzen oder Zu-spät-Kommen	179
Tabelle 5.9:	Gefühl der Zugehörigkeit zur eigenen Schule in den OECD-Staaten.....	181
Tabelle 6.1:	Verteilung der Items des Naturwissenschaftstests auf die drei Teilkompetenzen und zwei Wissensaspekte.....	194
Tabelle 6.2:	Stufen naturwissenschaftlicher Kompetenz in PISA 2012.....	195
Tabelle 6.3:	Mittelwerte und Standardabweichungen der naturwissenschaftlichen Kompetenz in Deutschland nach Schulart	206

Tabelle 7.1:	Überblick über die typischen Anforderungen pro Kompetenzstufe (Ib–VI)	223
Tabelle 7.2:	Mittelwerte und Standardabweichungen der Lesekompetenz in Deutschland nach Schulart und für die Gesamtstichprobe.....	235
Tabelle 8.1:	Zusammenhang zwischen Mathematikkompetenz und sozioökonomischem Status im internationalen Vergleich (Prädiktorvariable: HISEI).....	254
Tabelle 8.2:	Zusammenhang zwischen Mathematikkompetenz und ökonomischem, kulturellem und sozialem Status im internationalen Vergleich (Prädiktorvariable: ESCS)	255
Tabelle 8.3:	Veränderungen in den sozialen Disparitäten des Kompetenzerwerbs über die Zeit im internationalen Vergleich des HISEI-Index und seines Zusammenhangs mit der Mathematikkompetenz zwischen PISA 2003 und 2012	259
Tabelle 8.4:	Veränderungen in den sozialen Disparitäten des Kompetenzerwerbs über die Zeit im internationalen Vergleich des ESCS-Index und seines Zusammenhangs mit der Mathematikkompetenz zwischen PISA 2003 und 2012	261
Tabelle 8.5:	Regressionsmodelle zur Schätzung von Disparitäten in der Mathematikkompetenz: Vorhersage durch den sozioökonomischen und soziokulturellen Hintergrund sowie weitere Merkmale der sozialen Lage	263
Tabelle 8.6:	Beispielhafte Beschreibung der EGP-Klassifikation (Quelle: Baumert & Schümer, 2001, S. 339)	265
Tabelle 8.7:	Merkmale der sozialen Lage und der Bildungswege differenziert nach EGP-Klassen.....	266
Tabelle 8.8:	Mittelwerte und Standardabweichungen der Lesekompetenz differenziert nach EGP-Klassen (Bezugsperson) zwischen PISA 2000 und 2012 in Deutschland.....	268
Tabelle 8.9:	Prozentuale Anteile von Schülerinnen und Schülern, deren Lesekompetenz auf Kompetenzstufe Ia oder darunter liegt, differenziert nach EGP-Klassen (Bezugsperson) zwischen PISA 2000 und 2012 in Deutschland.....	268
Tabelle 8.10:	Prozentuale Anteile der Schülerinnen und Schüler in den Schularten, differenziert nach EGP-Klassen im Vergleich zwischen PISA 2000 und 2012 in Deutschland.....	269
Tabelle 9.1:	Prozentuale Anteile von fünfzehnjährigen Schülerinnen und Schülern mit Zuwanderungshintergrund in ausgewählten OECD-Staaten.....	282
Tabelle 9.2:	Prozentuale Anteile von fünfzehnjährigen Schülerinnen und Schülern mit Zuwanderungshintergrund, die zu Hause die Sprache des Einwanderungslandes sprechen.....	284
Tabelle 9.3:	Disparitäten im sozioökonomischen Status zwischen Familien ohne und Familien mit Zuwanderungshintergrund in ausgewählten OECD-Staaten.....	285
Tabelle 9.4:	Mittlere mathematische Kompetenz von Jugendlichen mit Zuwanderungshintergrund in ausgewählten OECD-Staaten.....	287
Tabelle 9.5:	Prozentuale Anteile fünfzehnjähriger Schülerinnen und Schüler mit Zuwanderungshintergrund in Deutschland.....	290
Tabelle 9.6:	Prozentuale Anteile fünfzehnjähriger Schülerinnen und Schüler mit Zuwanderungshintergrund in Deutschland nach Herkunft und Geburtsland der Eltern	291
Tabelle 9.7:	Prozentuale Anteile fünfzehnjähriger Schülerinnen und Schüler, die zu Hause Deutsch sprechen.....	293

Tabelle 9.8:	Prozentuale Anteile fünfzehnjähriger Schülerinnen und Schüler mit Zuwanderungshintergrund, die zu Hause Deutsch sprechen, nach Herkunft und Geburtsland der Eltern.....	293
Tabelle 9.9:	Unterschiede in der sozialen Herkunft zwischen Jugendlichen mit und Jugendlichen ohne Zuwanderungshintergrund.....	294
Tabelle 9.10:	Unterschiede in der sozialen Herkunft zwischen Jugendlichen ohne und Jugendlichen mit Zuwanderungshintergrund nach Herkunftsland und Geburtsland der Eltern.....	296
Tabelle 9.11:	Prozentuale Anteile fünfzehnjähriger Schülerinnen und Schüler mit Zuwanderungshintergrund aufgeteilt auf die Schularten.....	298
Tabelle 9.12:	Regressionsmodelle zur Schätzung von Disparitäten in der Mathematikkompetenz zwischen Jugendlichen ohne Zuwanderungshintergrund und Jugendlichen mit Zuwanderungshintergrund.....	300
Tabelle 9.13:	Regressionsmodelle zur Schätzung von Disparitäten in der Mathematikkompetenz zwischen Jugendlichen ohne Zuwanderungshintergrund und Jugendlichen mit Zuwanderungshintergrund unterschiedlicher Herkunftsgruppen.....	302
Tabelle 10.1:	Brutto-Stichprobe der Schulen nach Bundesland und Schulart.....	316
Tabelle 10.2:	Anzahl der in PISA 2012 eingesetzten Units und Items nach Domäne.....	324
Tabelle 10.3:	Testheftdesign von PISA 2012.....	325
Tabelle 10.4:	Beispiel einer Q-Matrix für 9 Items und 3 Kompetenzbereiche.....	335
Tabelle 10.5:	<i>Cut-Off</i> -Werte der Kompetenzstufen in PISA 2012.....	341

Anhang

Tabelle 1:	Mittelwerte, Streuungen und Perzentile mathematischer Kompetenz.....	348
Tabelle 2:	Mittelwerte, Streuungen und Perzentile für den mathematischen Inhaltsbereich <i>Veränderung und Beziehungen</i>	349
Tabelle 3:	Mittelwerte, Streuungen und Perzentile für den mathematischen Inhaltsbereich <i>Quantität</i>	350
Tabelle 4:	Mittelwerte, Streuungen und Perzentile für den mathematischen Inhaltsbereich <i>Raum und Form</i>	351
Tabelle 5:	Mittelwerte, Streuungen und Perzentile für den mathematischen Inhaltsbereich <i>Unsicherheit und Daten</i>	352
Tabelle 6:	Mittelwerte, Streuungen und Perzentile naturwissenschaftlicher Kompetenz.....	353
Tabelle 7:	Mittelwerte, Streuungen und Perzentile der Lesekompetenz.....	354
Tabelle 8:	Mittlere Kompetenz von Jugendlichen mit Zuwanderungshintergrund für Mathematik, Lesen und Naturwissenschaften.....	355

Die Autorinnen und Autoren dieses Berichtsbandes

Prof. Dr. Manfred Prenzel (ZIB/TU München)
Prof. Dr. Olaf Köller (ZIB/IPN Kiel)
Prof. Dr. Eckhard Klieme (ZIB/DIPF Frankfurt)
Prof. Dr. Cordula Artelt (Universität Bamberg)
Prof. Dr. Timo Ehmke (Leuphana Universität Lüneburg)
Prof. Dr. Aiso Heinze (IPN Kiel)
Prof. Dr. Kristina Reiss (TU München)
Prof. Dr. Tina Seidel (TU München)
Prof. Dr. Petra Stanat (IQB Berlin)
Dr. Lars Borchert (IEA DPC Hamburg)
Dr. Markus Gebhardt (TU München)
Dr. Katharina Hohn (ZIB/TU München)
Dr. Katharina Müller (TU München)
Dr. Andreas Obersteiner (TU München)
Dr. Dominique Rauch (DIPF Frankfurt)
Dr. Silke Rönnebeck (IPN Kiel)
Dr. Christine Sälzer (ZIB/TU München)
Dr. Anja Schiepe-Tiska (ZIB/TU München)
Dr. Katrin Schöps (IPN Kiel)
Jörg-Henrik Heine (ZIB/TU München)
Julia Mang (ZIB/TU München)
Stefanie Schmidtner (ZIB/TU München)
Heiko Sibberns (IEA DPC Hamburg)

Die *TUM School of Education* an der Technischen Universität München ist Deutschlands erste Fakultät für Lehrerbildung und Bildungsforschung. Sie widmet sich der Aus- und Weiterbildung von Lehrkräften an Gymnasien in den MINT-Fächern (Mathematik, Informatik, Naturwissenschaften, Technik) sowie der Aus- und Weiterbildung von Lehrkräften an beruflichen Schulen im gewerblich-technischen Gebiet. Das nationale Projektmanagement für PISA 2012 und 2015 ist an der *TUM School of Education* angesiedelt.

Das *DIPF (Deutsches Institut für Internationale Pädagogische Forschung)* in Frankfurt am Main unterstützt Forschung, Politik und Praxis im Bildungsbereich durch wissenschaftliche Infrastruktur und vielfältige Forschungstätigkeit. Als Mitglied der Leibniz-Gemeinschaft verbindet das Institut erkenntnisorientierte Grundlagenforschung mit innovativen Entwicklungsarbeiten und Anwendungen zum Nutzen der Gesellschaft und ihrer Mitglieder. Das nationale Projektmanagement für PISA 2009 oblag dem DIPF.

Das IPN (*Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik*) an der Universität Kiel hat zum Auftrag, durch seine Forschungen die Pädagogik der Naturwissenschaften und der Mathematik weiterzuentwickeln und zu fördern. Die interdisziplinären Arbeiten des IPN umfassen Grundlagenforschung in Fragen des Lehrens und Lernens. Das IPN hatte das nationale Projektmanagement für PISA 2003 und 2006 inne.

Das ZIB (*Zentrum für Internationale Vergleichsstudien*, www.zib-cisa.de) ist ein An-Institut der Technischen Universität München und umfasst als gemeinnütziger Verein die drei Institutionen TUM School of Education, IPN Kiel und DIPF Frankfurt. Der Zusammenschluss der drei Institutionen ermöglicht eine effiziente Bündelung von Kompetenz und Expertise im Bereich der Large-Scale-Assessments.

Das IQB (*Institut zur Qualitätsentwicklung im Bildungswesen*) in Berlin unterstützt die Arbeiten der Länder in der Bundesrepublik Deutschland bei der kontinuierlichen Weiterentwicklung und Sicherung von Bildungserträgen im Schulsystem. Eine zentrale Grundlage dieser Arbeiten bilden die länderübergreifenden Bildungsstandards der Kultusministerkonferenz, die definieren, welche Kompetenzen Schülerinnen und Schüler bis zu bestimmten Zeitpunkten in ihrer schulischen Laufbahn erwerben sollen. Das IQB überprüft regelmäßig, inwieweit diese Kompetenzziele in deutschen Schulen erreicht werden und unterstützt die Länder bei der Umsetzung der Bildungsstandards.

Das IEA DPC Hamburg (*Data Processing and Research Center der International Association for the Evaluation of Educational Achievement*) in Hamburg führt seit 1958 international vergleichende Schulleistungsstudien durch und ist in Deutschland seit der ersten PISA-Studie Kooperationspartner der wissenschaftlichen Studienleitung.