

PISA 2003  
Der zweite Vergleich der Länder in Deutschland –  
Was wissen und können Jugendliche?



Manfred Prenzel, Jürgen Baumert, Werner Blum,  
Rainer Lehmann, Detlev Leutner, Michael Neubrand,  
Reinhard Pekrun, Jürgen Rost und Ulrich Schiefele (Hrsg.)

PISA-Konsortium Deutschland

# PISA 2003

Der zweite Vergleich der Länder  
in Deutschland – Was wissen und  
können Jugendliche?



Waxmann  
Münster/New York  
München/Berlin

**Bibliografische Informationen Der Deutschen Bibliothek**

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

ISBN 3-8309-1560-8

© 2005 Waxmann Verlag GmbH

Postfach 8603, D-48046 Münster

Waxmann Publishing Co.

P.O. Box 1318, New York, NY 10028, USA

[www.waxmann.com](http://www.waxmann.com)

E-Mail: [info@waxmann.com](mailto:info@waxmann.com)

Buchumschlag: Christian Averbeck, Münster

Satz: Stoddart Satz- und Layoutservice, Münster

Druck: Runge GmbH, Cloppenburg

Gedruckt auf alterungsbeständigem Papier DIN 6738

Alle Rechte vorbehalten. Printed in Germany.

# Inhalt

Vorwort .....	11
<b>1 Einführung in den Ländervergleich PISA 2003 .....</b>	<b>13</b>
<i>Manfred Prenzel, Barbara Drechsel und Claus H. Carstensen</i>	
1.1 Das Anliegen der PISA-Erweiterung 2003 .....	13
1.2 Theoretischer Rahmen: Literacy, Kompetenzen, Hintergrundvariablen .....	16
1.3 Die nationalen Erweiterungen von PISA .....	20
1.4 Die Anlage der Untersuchung für den Ländervergleich .....	22
1.4.1 Unterschiedliche Bildungsbeteiligung in den Schularten .....	22
1.4.2 Die Stichprobenziehung für den Ländervergleich .....	24
1.4.3 Repräsentativität der Stichproben .....	27
1.4.4 Effekte von Üben und Testmotivation .....	29
1.4.5 Durchführung der Erhebung .....	32
1.4.6 Auswertung und Skalierung .....	33
1.4.7 Berichterstattung und Darstellung .....	35
1.5 Von PISA 2000 nach PISA 2003: Belastbare Aussagen über Veränderungen .....	39
1.6 PISA – Ein kooperatives Unternehmen .....	41
1.7 Was ist neu beim Ländervergleich PISA 2003? .....	44
1.8 Ergebnisse des Ländervergleichs im Überblick .....	46
Literatur .....	49
<b>2 Mathematische Kompetenz im Ländervergleich .....</b>	<b>51</b>
<i>Michael Neubrand, Werner Blum, Timo Ehmke, Alexander Jordan, Martin Senkbeil, Frauke Ulfing und Claus H. Carstensen</i>	
2.1 Mathematical Literacy in PISA 2003 .....	51
2.1.1 Das Konzept Mathematical Literacy .....	51
2.1.2 Konzeption des internationalen PISA-Mathematiktests .....	52
2.1.3 Kompetenzstufen .....	53
2.1.4 Aufgabenbeispiele .....	55
2.2 Befunde zur mathematischen Kompetenz .....	58
2.2.1 Mathematische Kompetenz in den Ländern der Bundesrepublik und den OECD-Staaten .....	59
2.2.2 Kompetenzstufen in Mathematik .....	62
2.2.3 Ergebnisse in den Inhaltsbereichen im nationalen und internationalen Vergleich .....	63
2.2.4 Mathematische Kompetenz von Schülerinnen und Schülern mit Migrationshintergrund .....	71

2.3	Veränderungen in Teilbereichen der mathematischen Kompetenz .....	73
2.3.1	Inhaltsbereich „Veränderung und Beziehungen“ .....	74
2.3.2	Inhaltsbereich „Raum und Form“ .....	75
2.4	Vergleich der mathematischen Kompetenzen in den Gymnasien .....	76
2.5	Profile der Inhaltsbereiche in den Ländern .....	80
2.6	Zusammenfassung und Diskussion .....	82
Literatur .....		83
3	Die Lesekompetenz im Ländervergleich .....	85
	<i>Barbara Drechsel und Ulrich Schiefele</i>	
3.1	Wie wird die Lesekompetenz in PISA erfasst? .....	86
3.2	Wie schneiden die Länder im internationalen Vergleich ab? .....	86
3.3	Wie ist die Lesekompetenz in den Ländern auf die Kompetenzstufen verteilt? .....	90
3.4	Hat sich die Lesekompetenz in den Ländern zwischen PISA 2000 und PISA 2003 verändert? .....	93
3.5	Ergebnisse für Teilpopulationen .....	94
3.5.1	Wie unterscheidet sich die Lesekompetenz von Schülerinnen und Schülern mit und ohne Migrationshintergrund in den Ländern? .....	94
3.5.2	Wie unterscheiden sich Leseleistungen an den Gymnasien der Länder? .....	97
3.6	Zusammenfassung und Diskussion .....	99
Literatur .....		101
4	Naturwissenschaftliche Grundbildung im Ländervergleich .....	103
	<i>Jürgen Rost, Martin Senkbeil, Oliver Walter, Claus H. Carstensen und Manfred Prenzel</i>	
4.1	Konzeption des internationalen Naturwissenschaftstests .....	103
4.2	Die Länder Deutschlands im internationalen Vergleich .....	105
4.3	Veränderungen in der naturwissenschaftlichen Grundbildung zwischen PISA 2000 und PISA 2003 .....	110
4.4	Vergleich der naturwissenschaftlichen Kompetenz für ausgewählte Subpopulationen .....	113
4.4.1	Naturwissenschaftliche Grundbildung von Schülerinnen und Schülern mit Migrationshintergrund .....	113
4.4.2	Vergleich der Länder hinsichtlich der naturwissenschaftlichen Kompetenz im Gymnasium .....	117
4.4.3	Geschlechterdifferenzen in der naturwissenschaftlichen Kompetenz .....	120
4.5	Zusammenfassung und Diskussion .....	121
Literatur .....		123

5	Die Problemlösekompetenz in den Ländern der Bundesrepublik Deutschland .....	125
	<i>Detlev Leutner, Eckhard Klieme, Katja Meyer und Joachim Wirth</i>	
5.1	Problemlösen als fächerübergreifende Kompetenz .....	125
5.2	Die Länder der Bundesrepublik im inter- und intranationalen Vergleich .....	126
5.3	Vergleich der Problemlösekompetenz für ausgewählte Subpopulationen .....	131
5.3.1	Problemlösekompetenz von Schülerinnen und Schülern mit und ohne Migrationshintergrund .....	131
5.3.2	Problemlösekompetenz von Schülerinnen und Schülern an Gymnasien .....	135
5.3.3	Geschlechterdifferenzen in der Problemlösekompetenz .....	138
5.4	Vergleich von Problemlösekompetenz und mathematischer Kompetenz .....	139
5.5	Zusammenfassung und Diskussion .....	143
	Literatur .....	145
6	Schülermerkmale im Ländervergleich .....	147
	<i>Reinhard Pekrun, Anne C. Frenzel, Karin Zimmer und Stephanie Lichtenfeld</i>	
6.1	Selbstvertrauen, Engagement und Lernverhalten im Fach Mathematik .....	147
6.1.1	Variablen und ihre Erfassung .....	148
6.1.2	Selbstvertrauen .....	149
6.1.3	Emotionales und motivationales Engagement .....	151
6.1.4	Lernverhalten: Kognitive und metakognitive Strategien .....	153
6.1.5	Länderunterschiede in PISA 2000 und PISA 2003: Gibt es Veränderungen? .....	154
6.1.6	Schlussfolgerungen .....	154
	Literatur .....	155
7	Die schulische Computernutzung in den Ländern und ihre Wirkungen .....	157
	<i>Martin Senkbeil</i>	
7.1	Computerbildung von Jugendlichen und Aufgaben der Schule .....	157
7.2	Der Einfluss der schulischen Computernutzung auf die computerbezogene Selbstwirksamkeit von Jugendlichen .....	158
7.3	Die Wirksamkeit der schulischen Computernutzung im internationalen Vergleich .....	163
7.4	Zusammenfassung und Diskussion .....	165
	Literatur .....	166

8	Der Blick in die Länder .....	169
	<i>Manfred Prenzel, Karin Zimmer, Barbara Drechsel, Heike Heidemeier und Clemens Draxler</i>	
8.1	Vergleiche innerhalb und zwischen Ländern .....	169
8.2	Baden-Württemberg .....	172
8.3	Bayern .....	175
8.4	Berlin .....	179
8.5	Brandenburg .....	183
8.6	Bremen .....	187
8.7	Hamburg .....	191
8.8	Hessen .....	195
8.9	Mecklenburg-Vorpommern .....	199
8.10	Niedersachsen .....	202
8.11	Nordrhein-Westfalen .....	206
8.12	Rheinland-Pfalz .....	210
8.13	Saarland .....	214
8.14	Sachsen .....	218
8.15	Sachsen-Anhalt .....	221
8.16	Schleswig-Holstein .....	225
8.17	Thüringen .....	229
8.18	Ein abschließender Blick über die Länder .....	232
	Literatur .....	233
9	Soziale Herkunft im Ländervergleich .....	235
	<i>Timo Ehmke, Thilo Siegle und Fanny Hohensee</i>	
9.1	Soziale Herkunft, Kompetenzniveau und Bildungsbeteiligung .....	235
9.2	Familiäre Lebensverhältnisse in den Ländern .....	236
9.2.1	Allgemeine Strukturmerkmale von Familien .....	237
9.2.2	Sozioökonomischer Status .....	239
9.2.3	Erwerbstätigkeit .....	241
9.2.4	Bildungsabschluss und kulturelle Ressourcen .....	244
9.2.5	Zusammenfassung der Befunde .....	249
9.3	Der ökonomische, soziale und kulturelle Status in den Ländern .....	250
9.4	Zusammenhang von sozialer Herkunft und Kompetenz .....	252
9.5	Soziale Herkunft und Bildungsbeteiligung .....	258
9.5.1	Gymnasiale Bildungsbeteiligung in den Ländern .....	260
9.5.2	Relative Wahrscheinlichkeit des Gymnasialbesuchs .....	261
9.6	Zusammenfassung und Diskussion .....	263
	Literatur .....	265



10	Soziokulturelle Herkunft und Migration im Ländervergleich .....	269
	<i>Gesa Ramm, Oliver Walter, Heike Heidemeier und Manfred Prenzel</i>	
10.1	Herausforderung Migration .....	269
10.2	Soziokulturelle Herkunft und Sprachgebrauch .....	271
10.2.1	Zur Herkunft der Jugendlichen und ihrer Eltern .....	272
10.2.2	Zur Verwendung der deutschen Sprache .....	276
10.3	Kompetenzen von Jugendlichen mit Migrationshintergrund in den Ländern .....	280
10.3.1	Mathematische Kompetenz und Herkunft .....	281
10.3.2	Mathematische Kompetenz, Lesekompetenz und Sprachgebrauch .....	283
10.4	Der sozioökonomische und soziokulturelle Status von Migrantenfamilien .....	287
10.5	Befunde zu den beiden größten Herkunftsgruppen .....	291
10.5.1	Aus der ehemaligen Sowjetunion zugewanderte Jugendliche .....	292
10.5.2	In Deutschland geborene Jugendliche mit Eltern aus der Türkei .....	293
10.6	Zusammenfassung und Diskussion .....	294
Literatur	.....	297
11	Schulmerkmale und Schultypen im Vergleich der Länder .....	299
	<i>Martin Senkbeil</i>	
11.1	Wie werden Schulmerkmale in PISA erfasst? .....	299
11.2	Welche Schultypen lassen sich differenzieren? .....	301
11.3	Wie verteilen sich die Schultypen auf die Schularten und Länder der Bundesrepublik? .....	309
11.3.1	Welche Zusammenhänge bestehen zwischen Schultypus und Schulart? .....	309
11.3.2	Die Verteilung der Schultypen auf die Länder der Bundesrepublik .....	311
11.3.3	Die Verteilung der Schultypen unter den Gymnasien der Länder der Bundesrepublik ....	313
11.4	Inwieweit stimmen die Schultypen mit den Einschätzungen der Lehrkräfte und der Schülerschaft überein? .....	315
11.5	Zusammenfassung und Diskussion .....	318
Literatur	.....	320
12	Wirtschaftliche, soziale und kulturelle Lebensverhältnisse und regionale Disparitäten des Kompetenzerwerbs .....	323
	<i>Jürgen Baumert, Claus H. Carstensen und Thilo Siegle</i>	
12.1	Alltagserfahrungen und regionalstatistische Evidenz .....	323
12.2	Regionale Kontexte als Sozialisationsmilieus: Forschungsstand .....	326
12.3	Merkmale familialer Herkunft, wirtschaftliche, soziale und kulturelle Umgebungen von Schulen und regionale Leistungsunterschiede in PISA 2003: Theoretischer Rahmen der Untersuchung .....	330
12.4	Stichprobe und technisches Vorgehen .....	336
12.5	Deskriptive Ergebnisse .....	338
12.6	Individuelle und familiale Voraussetzungen sowie institutionelle und regionale Rahmenbedingungen und der Erwerb von mathematischen Kompetenzen .....	341
12.6.1	Individuelle und familiale Voraussetzungen und der Erwerb von mathematischen Kompetenzen .....	341

12.6.2	Demographische, wirtschaftliche, soziale und ethnisch-kulturelle Rahmenbedingungen der Arbeit von Schulen und der Erwerb von mathematischen Kompetenzen .....	344
12.7	Wechselwirkungen zwischen sozialen Kontexten und familialen Merkmalen .....	350
12.8	Regionale Disparitäten des Kompetenzerwerbs – ein Ergebnis individueller Verteilungsunterschiede und regionaler Strukturdifferenzen? .....	351
12.9	Zusammenfassung .....	358
	Literatur .....	362
13	<b>Ein Fazit des Ländervergleichs: Ausgangslagen, Stärken und Herausforderungen</b> .....	367
	<i>Manfred Prenzel und Claus H. Carstensen</i>	
13.1	Kompetenzvergleiche bei unterschiedlichen Ausgangslagen: Adjustierung nach sozialer Herkunft und Migration .....	368
13.1.1	Adjustierte Vergleiche .....	369
13.1.2	Adjustierter Vergleich der Mathematikkompetenz .....	370
13.1.3	Die Adjustierung im Detail .....	371
13.1.4	Adjustierung für vier Kompetenzen .....	373
13.1.5	Fazit .....	374
13.2	Kompetenzergebnisse im Überblick: Stärken und Entwicklungen .....	375
13.2.1	Kompetenzprofile .....	375
13.2.2	Veränderungen und Entwicklungen .....	377
13.3	Wichtige Erkenntnisse aus dem Ländervergleich: Fortschritte und Herausforderungen .....	379
	Literatur .....	383
14	<b>Technische Grundlagen des Ländervergleichs</b> .....	385
	<i>Claus H. Carstensen, Steffen Knoll, Thilo Siegle, Jürgen Rost und Manfred Prenzel</i>	
14.1	Die Repräsentativität von PISA-E 2003 .....	385
14.1.1	Stichprobenziehung PISA-E .....	386
14.1.2	Die realisierte Stichprobe .....	390
14.1.3	Gewichtung .....	391
14.2	Messwerte für den Ländervergleich 2003 .....	395
14.2.1	Hintergrundmodell PISA-E .....	395
14.2.2	Skalierung für PISA-E .....	396
14.2.3	Das Verbinden der Ergebnisse aus PISA 2003 und PISA 2000 .....	397
14.3	In PISA verwendete Verfahren zur Datenanalyse – Glossar .....	398
	Literatur .....	401
	Anhang .....	403
	Abbildungs- und Tabellenverzeichnis .....	407
	Abkürzungsverzeichnis .....	415

## Vorwort der Präsidentin der Kultusministerkonferenz

Die vorab im Juli 2005 veröffentlichten zentralen Ergebnisse des zweiten Vergleichs der Länder in Deutschland haben deutlich gemacht: Das deutsche Bildungssystem ist in Bewegung geraten. Die Kompetenzen der Schülerinnen und Schüler in den Ländern haben sich in vielen Fällen signifikant gegenüber PISA 2000 verbessert, in manchen Ländern deutlicher und früher als man erwarten konnte. Offensichtlich hat sich ein Mentalitätswandel vollzogen, der Früchte trägt und zu Verbesserungen geführt hat. Die Kultusministerkonferenz betrachtet diese erfreulichen Ergebnisse des Ländervergleichs als Herausforderung, die begonnenen Anstrengungen fortzusetzen und – wo immer möglich – zu intensivieren. Neben der Teilnahme an internationalen Vergleichsuntersuchungen wie PISA und IGLU oder an nationalen Vergleichsuntersuchungen wie DESI bilden die länderübergreifend geltenden Bildungsstandards und die Zusammenarbeit von zahlreichen Ländern bei Untersuchungen wie VERA, aber auch in weiteren Bereichen der Qualitätssicherung, Qualitätsentwicklung und Evaluation wesentliche Säulen eines übergreifenden Bildungsmonitorings, das sich gegenwärtig in Deutschland entwickelt. Die Kultusministerkonferenz hat mit der Gründung des Instituts zur Qualitätsentwicklung im Bildungswesen deutlich gemacht, dass sie diesen Prozess mit Nachdruck verfolgt. Eine weitere Komponente des Bildungsmonitorings ist der künftig in zweijährigem Rhythmus erscheinende Bildungsbericht, der gemeinsam von Bund und Ländern in Auftrag gegeben wird und eine Gesamtschau von Bildung in Deutschland über den Schulbereich hinaus eröffnen wird.

Dass Vergleiche neben Übereinstimmungen auch Unterschiede aufdecken, kann niemanden überraschen. Der jetzt vorliegende analytische Band zum PISA-Ländervergleich wird dazu beitragen, die in den Ländern beobachteten Befunde besser zu verstehen. Der länderübergreifende Bericht von Gymnasien gewinnt durch den Vergleich von PISA 2000 mit PISA 2003 erheblich an Aussagekraft, darüber hinaus werden zentrale Ergebnisse für die Schularten jedes einzelnen Landes nach der internationalen Metrik dargestellt. Auch dies eröffnet neue Möglichkeiten, das durchschnittliche Leistungsniveau einer Schulart in jedem Land exakt zu beschreiben und Veränderungen im durchschnittlichen Kompetenzniveau zwischen den Erhebungszyklen für die einzelnen Schularten zu erfassen. Der Bericht geht der Frage nach, inwieweit unterschiedliche sozioökonomische Ausgangslagen in den Ländern oder unterschiedliche Migrationssituationen die durchschnittlichen Leistungsergebnisse beeinflussen können. Analysen von Schulmerkmalen, von strukturellen Merkmalen auf der Ebene von Regionen werden ebenfalls berichtet.

Diese Analysen tragen dazu bei, dass unser Wissen darüber wächst, wie das Bildungswesen die ihm gestellten Aufgaben lösen kann, welche Veränderungen ins Werk gesetzt

werden sollten und wie wir unsere Ziele erreichen können. Verbesserungen werden wir nur erreichen, wenn wir uns offen den Stärken und Schwächen unserer Systeme zuwenden. Nur wenn wir bereit sind, offen Mängel einzugestehen statt sie ängstlich zu verbergen, wird es uns gelingen, nachhaltige Verbesserungen in unserem Bildungssystem zu erreichen.

In diesem Zusammenhang möchte ich nochmals deutlich unterstreichen, dass es allen Ländern darum geht, mit Hilfe von PISA optimales Steuerungswissen über ihr System zu erhalten. Nur so können wir die gemeinsamen Ziele erreichen:

- Steigerung des allgemeinen Leistungsniveaus
- Erhöhung der Bildungsgerechtigkeit
- Verringerung der Leistungsabstände zwischen den Ländern.

Dem Konsortium gilt mein Dank für die umfangreichen Informationen und Analysen in diesem PISA-Ländervergleich sowie für die umsichtigen Interpretationen und Hinweise zu Möglichkeiten, Bedingungen und Grenzen des Ländervergleichs.

Bonn, im November 2005

Prof. Dr. Johanna Wanka  
Präsidentin der Ständigen Konferenz der Kultusminister der Länder

# 1 Einführung in den Ländervergleich PISA 2003

Manfred Prenzel, Barbara Drechsel  
und Claus H. Carstensen

Im vorliegenden Band werden die Ergebnisse des *OECD Programme for International Student Assessment 2003* (PISA) des zweiten Vergleichs der Länder der Bundesrepublik Deutschland im Rahmen von PISA 2003 vorgestellt. Nach der Darstellung der internationalen Ergebnisse im Dezember 2004 (Prenzel et al., 2004a) wird nun der Frage nachgegangen, wie die Länder der Bundesrepublik im nationalen und internationalen Vergleich abschneiden. Der Vergleich der Länder in Deutschland erfolgt im Auftrag der Ständigen Konferenz der Kultusminister (KMK). Die Studie nutzt eine Option des OECD-Programms: Sie untersucht den Kompetenzstand in regionalen Gliederungen des teilnehmenden Staates, in Deutschland also in den Ländern. Von dieser Möglichkeit machen bei PISA eine Reihe von Staaten Gebrauch, zum Beispiel Australien, Kanada oder die Schweiz (z.B. BFS/EDK, 2005). Voraussetzung für solche ergänzenden Untersuchungen ist eine Vergrößerung der Stichproben. In Deutschland wurden die repräsentativen Stichproben so erweitert, dass die Ergebnisse der Länder untereinander verglichen und international eingeordnet werden können. Letztlich wird damit jedes Land wie ein Staat im OECD-Vergleich behandelt.

Das erste Kapitel führt in den Ländervergleich in PISA 2003 ein und beschreibt die Ziele und die Vorgehensweise bei der Vorbereitung, Durchführung und Auswertung der Studie. Die Abschnitte am Ende dieses Kapitels präsentieren eine kurze Zusammenfassung der Besonderheiten des Ländervergleichs 2003 und seiner wichtigsten Ergebnisse.

## 1.1 Das Anliegen der PISA-Erweiterung 2003

PISA untersucht, wie gut junge Menschen auf Herausforderungen der Wissensgesellschaft vorbereitet sind. Getestet werden fünfzehnjährige Jugendliche am Ende der Pflichtschulzeit. Das Ziel der Studie ist es, zentrale und grundlegende Kompetenzen zu erfassen, die für die individuellen Lern- und Lebenschancen ebenso bedeutsam sind wie für die gesellschaftliche, politische und wirtschaftliche Weiterentwicklung. Von besonderer Bedeutung ist dabei, ob die Jugendlichen ihre Kompetenzen flexibel und situations-

gerecht bei der Lösung von Aufgaben nutzen können, welche vielfältige alltags-, ausbildungs- wie berufsbezogene Anforderungen repräsentieren. Die so gemessenen Kompetenzen sind das Kriterium, mit dessen Hilfe die Leistungsfähigkeit von Bildungssystemen in PISA verglichen und beurteilt wird.

PISA ist der zentrale Teil eines umfassenden Indikatorensystems der OECD. Es dient dazu, die Mitgliedstaaten über Stärken und Schwächen ihrer Bildungssysteme zu informieren. PISA gestattet es auch, Bildungsergebnisse innerhalb der Länder nach internationalen Maßstäben zu beurteilen. Die Studie liefert Bezugspunkte, an denen Möglichkeiten für eine Weiterentwicklung der Bildungssysteme geklärt werden können. In diesem Sinne dient PISA einem *Benchmarking* im Bildungsbereich. Da die Erhebungen regelmäßig in einem Abstand von drei Jahren durchgeführt werden, informieren sie auf längere Sicht darüber, inwieweit ergriffene Maßnahmen die angestrebten Wirkungen erreichen konnten. Unter diesem Aspekt kann PISA als Verfahren zur Dauerbeobachtung von Bildungssystemen im Sinne eines *Bildungsmonitorings* beitragen. PISA stellt den Auftraggebern der Studie empirisch fundierte Erkenntnisse in Aussicht, die ihnen helfen sollen, ihre Bildungssysteme auf der Basis umfassender und zuverlässiger Daten zu steuern.

Befunde über die Stärken und Schwächen von Bildungssystemen sind nicht nur für Bildungsadministration und Bildungspolitik bedeutsam. Auch für die Öffentlichkeit, die Eltern und vor allem die Personen, die sich professionell mit Schule und Bildungsfragen befassen, ist es wichtig zu erfahren, inwieweit junge Menschen im Verlauf der Schulzeit auf zukünftige Anforderungen vorbereitet werden (vgl. Abbildung 1.1).

Das Anliegen von PISA richtet sich deshalb *nicht* darauf, Staaten und Länder in der Art eines sportlichen Leistungswettbewerbs zu vergleichen und mit Rangplätzen auszuzeichnen. Die bei PISA erfassten Kompetenzen beschreiben keine Endzustände eines Trainings für einen Leistungsvergleich, sondern entscheidende Voraussetzungen für weiterführende Lernprozesse mit herausragender Bedeutung für das Individuum und die Gesellschaft. Mit Blick auf die langfristigen Konsequenzen gilt es deshalb, Erkenntnisse über Probleme sowie über notwendige und mögliche Verbesserungen im Bildungsbereich der teilnehmenden Länder zu erhalten. Der internationale und der nationale Vergleich in einem föderalen Bildungssystem helfen, relative Stärken und Schwächen zu erkennen.

Das Design der PISA-Studie ist offen für Erweiterungen und Ergänzungen, sofern der nach internationalen Standards vorgeschriebene Inhalt und Ablauf des Tests nicht beeinflusst wird. In Deutschland wurde von der Möglichkeit einer nationalen Stichprobenerweiterung bereits in PISA 2000 (Baumert et al., 2001; Baumert et al., 2002a; Baumert et al., 2003) Gebrauch gemacht. Auch in PISA 2003 wurde in allen Ländern eine repräsentative Stichprobe untersucht, die für jedes Land und pro Schulart statistisch abgesicherte Aussagen über die Fünfzehnjährigen erlaubt. Dieses Design erlaubt einen systematischen Ländervergleich, über den der vorliegende Band berichtet. Dabei wird die internationale Vergleichsperspektive mit einer nationalen kombiniert.

Abbildung 1.1: PISA 2003 – der Ländervergleich im Überblick

*Inhaltsbereiche*

- PISA untersucht die Kompetenzen von fünfzehnjährigen Schülerinnen und Schülern in den Bereichen Mathematik (Schwerpunktgebiet 2003), Lesen und Naturwissenschaften. Die Erhebungen beruhen auf einem Testansatz, der in allen Inhaltsbereichen zwischen Konzepten, Prozessen und Situationen beziehungsweise Kontexten unterscheidet. Die Testkonzeption ist an einer Vorstellung von lebenslangem Lernen orientiert und betont das Verstehen und die flexible, situationsgerechte Anwendung des Wissens.
- Neben bereichsspezifischen werden bereichsübergreifende Kompetenzen untersucht. Im Zentrum dieser so genannten Cross-Curricular-Competencies steht 2003 das Problemlösen. Neben den Tests zur Problemlösekompetenz ergänzen Erhebungen zu Lernstrategien, Lernmotivation und zur Vertrautheit mit Informationstechnologien den fächerübergreifenden Untersuchungsbereich.

*Erhebungsverfahren*

- Die Erhebung erfolgt an einem Testtag an der Schule, in Gruppen und unter Aufsicht.
- Die Tests bestehen aus Mehrfachauswahlfragen (Multiple-Choice). Sie werden mit Fragen kombiniert, die von den Schülerinnen und Schülern mit eigenen Worten oder Darstellungen beantwortet werden müssen (offene Fragen). Die Einzelfragen (Items) sind thematisch zu Aufgaben gruppiert, die sich auf eine durch einen kurzen Text beschriebene Situation beziehen.
- Die reine Testzeit für jeden Schüler beziehungsweise jede Schülerin beträgt zwei Stunden.
- Die Schülerinnen und Schüler erhielten Testhefte mit unterschiedlichen Aufgabenpaketen. Auf diese Weise kann insgesamt Itemmaterial für mehr als sechseinhalb Stunden Testzeit eingesetzt werden. Der größte Teil dieser Testzeit (3,5 Stunden) entfiel 2003 auf das Schwerpunktgebiet Mathematik.
- Die Schülerinnen und Schüler bearbeiteten weiterhin einen Fragebogen (30 Minuten), der sich auf ihre Herkunft und Umgebung, ihre Lerngewohnheiten und Motivation bezieht. Ein weiterer Fragebogen befasst sich mit der Vertrautheit mit Computern und mit den Vorstellungen über die eigene Bildungskarriere (15 Minuten).
- Die Schulleitungen wurden gebeten, einen Fragebogen zu Merkmalen ihrer Schule (z.B. Ressourcen, Qualifikation der Lehrkräfte, Schulklima) auszufüllen.

*Stichprobe*

- In den 16 Ländern der Bundesrepublik wurden ca. 45000 Schülerinnen und Schüler getestet.
- Die teilnehmenden Schulen und die teilnehmenden Schülerinnen und Schüler wurden mit einem Zufallsverfahren gezogen.

*Ergebnisse*

- Ein Profil von wichtigen Kompetenzen, über die Fünfzehnjährige verfügen.
- Informationen über Zusammenhänge zwischen Kompetenzen und Merkmalen der sozialen Herkunft.
- Erste Einschätzungen der Veränderungen in den Kompetenzen zwischen den Erhebungen in den Jahren 2000 und 2003.

## 1.2 Theoretischer Rahmen: Literacy, Kompetenzen, Hintergrundvariablen

Die theoretischen Grundlagen für PISA 2003 wurden im Anschluss an Konsultationen von internationalen Fachleuten aus den Teilnehmerländern im Einvernehmen mit deren Regierungen entwickelt und in einer Rahmenkonzeption dargelegt (OECD, 2003). Ausgangspunkt der Rahmenkonzeption ist eine Vorstellung von Grundbildung (*Literacy*), bei der es um die Fähigkeit der Schülerinnen und Schüler geht, Kenntnisse und Fertigkeiten in einer Vielzahl von Situationen anzuwenden und zu analysieren, logisch zu denken und in effektiver Weise zu kommunizieren.

*Literacy* im (engeren) Sinne einer Lesekompetenz befähigt, an einer Kultur teilzuhaben, deren Wissen in Texten vorliegt. Das Beispiel Lesekompetenz zeigt die Tragweite einer „kulturellen Teilhabe“, die umfassend die persönlichen Handlungsmöglichkeiten im Alltag, Beruf und gesellschaftlichen Leben betrifft. Auf ähnliche Weise können grundlegende mathematische und naturwissenschaftliche Kompetenzen im (übertragenen) Sinne einer *Literacy* bestimmt und beschrieben werden.

Die PISA zugrunde liegende Bildungskonzeption orientiert sich damit zunächst an der *Funktion* von Kompetenzen im kulturellen und gesellschaftlichen Zusammenhang. So wird die Ausarbeitung der Rahmenkonzeption bestimmt durch die Frage, welches Konzeptwissen („Wissen, dass“) und welches Prozesswissen („Wissen, wie“) für die kulturelle Teilhabe bedeutsam sind und in welchen Situationen entsprechendes Wissen genutzt werden sollte. Ein weiterer Gesichtspunkt kommt hinzu: Um einzugrenzen, welche Kompetenzen für Fünfzehnjährige im Hinblick auf zukünftige Herausforderungen wichtig werden, rückt PISA die *Voraussetzungen für ein weiterführendes Lernen* beziehungsweise für eine weiterführende Auseinandersetzung mit diesen Kulturbereichen in den Blickpunkt. Hier gilt es zu klären, welches begriffliche Verständnis und welches



Prozesswissen gute Chancen bieten, neue Informationen zu verstehen, einzuordnen und damit weiter zu lernen. PISA bemüht sich besonders darum, das für die Kompetenzbereiche grundlegende und anschlussfähige Wissen zu identifizieren und in Aufgaben umzusetzen.

Diese Gesichtspunkte begrenzen die Rahmenkonzeption und eröffnen zugleich einen pragmatischen Weg, theoretisch anspruchsvolle Bildungsergebnisse empirisch zu erfassen und zwischen den teilnehmenden Staaten und Ländern zu vergleichen. In Anbetracht einer großen internationalen curricularen Vielfalt verzichtet PISA auf eine enge Orientierung an den Lehrplänen. Die Studie konzentriert sich vielmehr auf eine Untersuchung von Kompetenzen in Schlüsselbereichen, die als *notwendige Voraussetzungen für weiterführende Lern- und Bildungsprozesse* angesehen werden.

PISA hat in das Spektrum der zu untersuchenden Kompetenzen von Anfang an *bereichsübergreifende Kompetenzen* mit einbezogen. Die Erhebungsrunde 2003 lenkt hier die Aufmerksamkeit auf das Lösen von Problemen, die typisch für viele Situationen im alltäglichen Leben sind. Sie können nicht einem Schulfach oder einer Disziplin zugeordnet werden und verlangen eine gründliche Analyse der Anforderungen und die intelligente Verknüpfung von Wissensbeständen. Als Aspekte fächerübergreifender Kompetenzen werden außerdem Strategien des Lernens und der motivationalen Steuerung betrachtet. Auch Einstellungen zum Lernen und zu bestimmten Gegenstands- beziehungsweise Kulturbereichen werden bei PISA als wichtige Ziele schulischer Bildungsprozesse betrachtet. Der Testansatz von PISA beschränkt sich damit keineswegs auf drei Kernbereiche, sondern erfasst ein relativ breit gefasstes Spektrum von Kompetenzen.

Die Rahmenkonzeption für PISA (OECD, 2003), die diese Ansprüche für die verschiedenen Bereiche umzusetzen versucht, beschreibt damit einen Testansatz, der wichtige Aspekte von *Grundbildung* (im Sinne etwa der deutschsprachigen Diskussion) aufgreift (vgl. Tenorth, 2004). Grundbildungskonzeptionen zeichnen sich dadurch aus, dass sie Ansprüche formulieren, die möglichst von allen Schülerinnen und Schülern erreicht werden sollten. Auch diesem Anspruch wird die internationale Vorstellung von Literacy gerecht. Freilich muss die Testkonzeption auch in der Lage sein, unterschiedliche Niveaus in der Kompetenzentwicklung abzubilden, die von fünfzehnjährigen Schülerinnen und Schülern erreicht werden können. Damit können auch Aussagen über Gruppen getroffen werden, die sich durch besonders niedrige oder hohe Kompetenz auszeichnen. Da sich die Vorstellungen von Literacy und Grundbildung in vielfacher Hinsicht überschneiden, werden wir in diesem Band die Begriffe synonym verwenden.

*Kompetenz als Potential:* PISA vergleicht das Grundbildungsniveau, das in den verschiedenen nationalen Bildungssystemen über den Verlauf der Schulzeit erreicht wird. Damit das angeeignete Wissen beim Weiterlernen oder unter Anforderungen außerhalb der Schule genutzt werden kann, muss es in unterschiedlichen Situationen aktiviert und auf die jeweiligen Anforderungen bezogen werden. Deshalb unterscheidet sich PISA von herkömmlichen Verfahren, Schul- und Buchwissen abzufragen. Die Jugendlichen werden vielmehr mit vielfältigen, realitätsnahen Aufgaben konfrontiert, die eine flexible, an die besondere Situation angepasste Anwendung ihres Wissens verlangen. Die Tests

zielen darauf ab, die Möglichkeiten der Schülerinnen und Schüler auszuloten, in einem bestimmten Gebiet Anforderungen zu bewältigen, Probleme zu lösen und weiter zu lernen. Letztlich soll also das entsprechende *Potential* der Schülerinnen und Schüler erfasst werden. Dieses Potential wollen wir im vorliegenden Band als *Kompetenz* bezeichnen, auch wenn dieser Begriff in der wissenschaftlichen Literatur nicht immer einheitlich gebraucht wird (Klieme et al., 2003; Weinert, 1999). Aus den Antworten auf die Testfragen (aus der Testleistung oder „Performanz“) zu einem bestimmten Teilgebiet wird also die mathematische oder naturwissenschaftliche Kompetenz beziehungsweise die Lese- oder Problemlösekompetenz erschlossen. Der Begriff „Kompetenz“ wird dem Anliegen von PISA, die Vorbereitung auf Anforderungen zu erfassen, sehr viel besser gerecht als die Bezeichnung „Schulleistung“. Freilich waren den Expertengruppen die üblichen curricularen Schwerpunktsetzungen durchaus bewusst, als sie Testanforderungen und Kompetenz im Rahmen einer Grundbildungskonzeption spezifizierten, die das anschlussfähige Wissen und die kulturelle Teilhabe betont.

*Struktur der Kompetenzerhebungen:* Die bei PISA 2000 verwendete Grundstruktur bestimmt auch den Aufbau der Testkonzeptionen der aktuellen Erhebung. Der Testansatz für Problemlösen musste 2003 neu erarbeitet, der für den Bereich Mathematik als Schwerpunktgebiet ausgebaut werden. Im Bereich Naturwissenschaften sind einige kleinere Änderungen zu verzeichnen. Die Testkonzeptionen unterscheiden jeweils drei Aspekte oder Dimensionen:

- die *Inhalte* oder die Konzepte („Wissen, dass“), über die Schülerinnen und Schüler verfügen sollten;
- die *Prozesse* und Prozeduren („Wissen, wie“), die von den Schülerinnen und Schülern verstanden und beherrscht werden sollten;
- die *Situationsklassen* und Kontexte, in denen die Schülerinnen und Schüler ihr Wissen anwenden können sollten.

*Überblick über die Kompetenzbereiche:* Abbildung 1.2 greift diese Unterscheidung auf, um die wichtigsten Merkmale für die bei PISA 2003 untersuchten Kompetenzbereiche zu skizzieren.

Die in Abbildung 1.2 vorgestellten Definitionen lassen die Betonung von Kompetenz im Sinne eines Fähigkeitspotentials ebenso erkennen wie die Orientierung auf die kulturelle Teilhabe im Sinne von Literacy.

*Hintergrundmerkmale:* Die Erhebungen bei PISA beschränken sich nicht nur auf Bildungsergebnisse. Unter dem Begriff „Hintergrundmerkmale“ verbergen sich Merkmale, die mit den Bildungsergebnissen der Schüler und Schülerinnen assoziiert sind und deren Lern- und Lebensumgebungen zugeordnet werden können. Es werden Lern- und Entwicklungsbedingungen erhoben, die vor allem das Elternhaus, die Schule und den Unterricht charakterisieren. Mit der Erhebung von Hintergrundmerkmalen auf diesen unterschiedlichen Ebenen kann PISA nicht nur aufschlussreiche Informationen über Bedingungen des Aufwachsens und diesbezügliche Unterschiede liefern. Die Erhebungen geben auch die Möglichkeit, die unter verschiedenen (und aus einer theoretischen

Abbildung 1.2: Überblick über die Testkonzeptionen (vgl. OECD, 2003)

	<b>Mathematik</b>	<b>Naturwissenschaften</b>	<b>Lesen</b>
Definition	Die Fähigkeit einer Person, die Rolle zu erkennen und zu verstehen, die Mathematik in der Welt spielt, fundierte mathematische Urteile abzugeben und sich auf eine Weise mit der Mathematik zu befassen, die den Anforderungen des gegenwärtigen und künftigen Lebens als konstruktiver, engagierter und reflektierender Bürger entspricht.	Die Fähigkeit, naturwissenschaftliches Wissen anzuwenden, naturwissenschaftliche Fragen zu erkennen und aus Belegen Schlussfolgerungen zu ziehen, um Entscheidungen zu verstehen und zu treffen, welche die natürliche Welt und die durch menschliches Handeln an ihr vorgenommenen Veränderungen betreffen.	Die Fähigkeit, geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, das eigene Wissen und Potential weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen.
„Inhalte“	Übergreifende Ideen: – Quantität – Raum und Form – Veränderung und Beziehungen – Unsicherheit	Bereiche naturwissenschaftlichen Wissens und naturwissenschaftlicher Konzepte wie zum Beispiel – Kraft und Bewegung – Artenvielfalt – physiologische Veränderungen	Textarten: – kontinuierliche Texte, z.B. Erzählungen, Beschreibungen, Argumentationen – nichtkontinuierliche Texte, z.B. Diagramme, Formulare und Listen
„Prozesse“	Die Kompetenzcluster definieren mathematische Fertigkeiten, unterteilt in drei Niveaubereiche: – Reproduktion (Ausführen einfacher Standardtätigkeiten, die direkt aus der Situation entnommen werden können) – Verbindungen (überschaubare Tätigkeiten, welche bereits mehrere Schritte oder die Verknüpfung mehrerer Aufgabenelemente erfordern) – Reflexion (komplexe Tätigkeiten, Verallgemeinerungen oder Reflexionen gefordert)	Die Fähigkeit, naturwissenschaftliches Wissen und Verständnis zu nutzen, naturwissenschaftliche Befunde zu erheben, zu interpretieren und nach ihnen zu handeln. Sie umfasst: – das Beschreiben, Erklären und Vorhersagen naturwissenschaftlicher Phänomene – das Verstehen wissenschaftlicher Forschung – das Interpretieren wissenschaftlicher Befunde und Schlussfolgerungen	Leseaufgaben: – Informationen herausuchen – eine Interpretation entwickeln und/oder – über Inhalt und Form eines Textes reflektieren PISA untersucht eher das „Lesen, um zu lernen“ als das „Lesen lernen“ selbst. Daher werden die grundlegendsten Lesefertigkeiten nicht getestet.
„Situationen“	Das Rahmenkonzept unterscheidet folgende Situationen: – persönliche – ausbildungs- und berufsbezogene – gesellschaftsbezogene – wissenschaftliche	Naturwissenschaftliche Anwendungen in den Bereichen – Leben und Gesundheit – Erde und Umwelt – Naturwissenschaft in Technologien	Der Text dient – privaten Zwecken (z.B. Brief) – öffentlichen Zwecken (z.B. offizielles Dokument) – der beruflichen Qualifikation (z.B. Berichte)

Sicht unterschiedlich unterstützenden) Bedingungen entwickelte Kompetenz von Schülerinnen und Schülern zu vergleichen, auf der nationalen wie internationalen Ebene. Damit können auch Aussagen darüber getroffen werden, inwieweit bestimmte Lebensbedingungen (z.B. Merkmale der Herkunft) in den einzelnen Ländern systematisch mit Kompetenzunterschieden – also unterschiedlichen Chancen auf eine erfolgreiche Kompetenzentwicklung – verknüpft sind. Da der Schwerpunkt im Jahr 2003 auf Mathematik liegt, werden verstärkt Hintergrundmerkmale berücksichtigt, die für die Entwicklung der mathematischen Kompetenz als bedeutsam betrachtet werden.

Für die Erfassung dieser Merkmale nutzt PISA zwei *Informationsquellen*: die Schülerinnen und Schüler selbst und die Schulleitungen. Im *Schülerfragebogen* geben die Jugendlichen Auskunft über Merkmale ihres Elternhauses (ihrer sozialen Herkunft), über Einstellungen, Aktivitäten und über die Wahrnehmung ihres (vor allem Mathematik-)Unterrichts und ihrer Schule (z.B. Schulklima). Der an die Schulleitung gerichtete *Schulfragebogen* erhebt unter anderem die Größe und Ausstattung der Schule sowie Ressourcen, thematisiert aber auch das Schulmanagement, Kooperationen und Verfahren der Qualitätssicherung und fragt nach dem Schulklima und besonderen Unterstützungsangeboten.

### 1.3 Die nationalen Erweiterungen von PISA

In Abstimmung mit der Auftraggeberin (der KMK) wurde das Design der OECD-Studie PISA in Deutschland erweitert und ergänzt. PISA 2003 besteht in Deutschland aus drei Teilstudien (vgl. Prenzel, Drechsel, Carstensen & Ramm, 2004b, S. 23).

Die erste Teilstudie setzt den internationalen Vergleich nach den Standardregeln um und wird als *PISA-I* („PISA-International“) bezeichnet. Der Bericht zu dieser Teilstudie ist im Dezember 2004 erschienen (Prenzel et al., 2004a).

Die vorliegende *zweite* Teilstudie erweitert den internationalen Vergleich um den Vergleich zwischen den Ländern der Bundesrepublik Deutschland. Hier wird die Stichprobe um ein Mehrfaches erweitert. Außerdem wurde eine zusätzliche Zahl von Schülerinnen und Schülern mit Migrationshintergrund in die Stichprobe einbezogen. Im Rahmen des Ländervergleichs bearbeiten nur die fünfzehnjährigen Schülerinnen und Schüler die internationalen Tests. Diese Teilstudie wird mit *PISA-E* (Erweiterungsstudie) etikettiert.

Die *dritte* Teilstudie konzentriert sich auf die Schulen der internationalen Stichprobe. In diesen Schulen wurden zusätzlich zwei komplette neunte Klassen per Zufall ausgewählt. Diese Klassen wurden ein Jahr nach der PISA-Erhebung 2003 (also 2004) ein zweites Mal getestet, um den Lernzuwachs über ein Schuljahr zu bestimmen. Da in dieser Studie – *PISA-I-Plus* – zahlreiche weitere Erhebungsverfahren zum Einsatz kamen, sind Rückschlüsse auf lernförderliche Unterrichts- und Schulbedingungen möglich. Der Bericht zu dieser Teilstudie wird im Herbst 2006 erscheinen.

Die vorliegende Teilstudie *PISA-E* übernimmt die Struktur und die Verfahren des internationalen Tests: An einem Testtag absolvieren die Schülerinnen und Schüler unter Aufsicht externer Testleiter ein Testprogramm in den drei Kompetenzbereichen Mathematik, Lesekompetenz und Naturwissenschaften sowie im Problemlösen als fächerübergreifender Kompetenz. Um Hintergrundmerkmale zu erfassen wird dieser zweistündige Test ergänzt durch einen Fragebogen, in dem die Jugendlichen in ca. 45 Minuten um Angaben zu ihrer Herkunft und Umgebung, ihren Lerngewohnheiten in Mathematik, ihrer Motivation sowie der Vertrautheit mit Computern und den Vorstellungen über ihre Bildungskarriere gebeten werden. Schülerinnen und Schüler aus Sonder- und Förderschulen erhalten ein verkürztes Testheft, in dem sie in einer Stunde die leichtesten PISA-Aufgaben bearbeiten. Auf die Bearbeitung des Fragebogens wird in dieser Schulart verzichtet.

Weiterhin werden – analog zum internationalen Verfahren – von den Schulleitungen in einem Schulfragebogen Auskünfte zu Größe und Ausstattung der Schule, Ressourcen, Management, Kooperation, Verfahren der Qualitätssicherung, Schulklima und besonderen Unterstützungsangeboten eingeholt.

#### *Erweiterung durch zusätzliche Instrumente und Items*

Eine erste Erweiterung des internationalen Testdesigns ist durch einige zusätzliche Fragen im Schülerfragebogen gegeben: Um die Situation von Jugendlichen mit Migrationshintergrund in Deutschland genauer beleuchten zu können, wurden Fragen hinzugefügt, die sich neben Angaben zur Muttersprache insbesondere auf die Bedingungen und den Verlauf des Erwerbs der deutschen Sprache beziehen.

Auch im Schulfragebogen wurden zusätzlich Fragen zur Situation von Schülerinnen und Schülern mit Migrationshintergrund gestellt. Es wurde nach Stütz- und Fördermaßnahmen, insbesondere nach der Realisierung von Maßnahmen zur speziellen Förderung von Schülerinnen und Schülern nichtdeutscher Muttersprache gefragt.

In einem ergänzenden nationalen Schulfragebogen wurden Schulleiterinnen und Schulleiter außerdem um Auskünfte über die Schule und Ansätze zur Schul- und Qualitätsentwicklung gebeten.

#### *Erweiterung durch zusätzliches Einbeziehen von Fünfzehnjährigen mit Migrationshintergrund*

Im Rahmen von PISA 2003 wird besonderes Augenmerk auf die Situation und auf die Kompetenzprofile von Schülerinnen und Schülern nichtdeutscher Muttersprache gerichtet. Neben der für den Ländervergleich erforderlichen Stichprobenerweiterung wurde ein *Oversampling* von Schülerinnen und Schülern mit Migrationshintergrund durchgeführt. Dafür wurden in den Schulen zusätzlich bis zu 15 Fünfzehnjährige aus Familien mit Migrationshintergrund zufällig ausgewählt. Der Stichprobenanteil ist damit ausreichend groß, um systematische und aussagekräftige Vergleiche zwischen Jugendlichen deutscher und Jugendlichen nichtdeutscher Muttersprache durchzuführen.

## 1.4 Die Anlage der Untersuchung für den Ländervergleich

### 1.4.1 Unterschiedliche Bildungsbeteiligung in den Schularten

Betrachtet man die Schulsysteme der Länder in der Bundesrepublik Deutschland aus einer internationalen Perspektive, dann weisen sie viele strukturelle Gemeinsamkeiten und Ähnlichkeiten auf. In allen Ländern finden wir zum Beispiel die charakteristische Schulart Gymnasium, aber auch große Übereinstimmungen im Bereich der Berufsschulen. Bei einer nationalen Betrachtung wiederum fallen eher die Besonderheiten auf, die sich aufgrund unterschiedlicher Traditionen und Herausforderungen entwickelt haben. So finden wir zum Beispiel die Haupt- und die Realschule nicht mehr in allen Ländern vor. Unter dem Etikett der Schulen mit mehreren Bildungsgängen verbergen sich unterschiedlich benannte und konzipierte Schularten (z.B. Regionale Schule, Regel- oder Mittelschule), die zu unterschiedlichen Abschlüssen führen können. Nicht zuletzt tragen Schulen in privater Trägerschaft, aber auch auslaufende oder neu eingeführte Schularten zu einer erheblichen Vielfalt in den Bildungssystemen der Länder bei.

In Anbetracht der Vielfalt von Beschulungsformen stellt sich die Frage nach der Gewichtung und der Vergleichbarkeit. Letztlich wird die Geburtenkohorte der Fünfzehnjährigen in allen Ländern zu bestimmten Anteilen auf die vorhandenen Schularten verteilt. Die Bildungsbeteiligung in den verschiedenen Schularten der Länder stellt die Tabelle 1.1 vor. Sie beschreibt die Anteile der Fünfzehnjährigen in den Schularten und damit die Verteilung gemäß der PISA-Stichprobe. Diese Anteile entsprechen nicht den Zahlen für bestimmte Klassenstufen, die in anderen Statistiken berichtet werden (z.B. Ständige Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland, 2005). Die bei PISA getesteten fünfzehnjährigen Jugendlichen verteilen sich auf mehrere Klassenstufen (vgl. Kapitel 8). Da PISA untersucht, welche Kompetenzen zu einem bestimmten Lebenszeitpunkt entwickelt wurden, sind die Fünfzehnjährigen die adäquate Vergleichsgruppe und nicht die Schülerinnen und Schüler einer bestimmten Klassenstufe.

Die Tabelle 1.1 lässt die Gemeinsamkeiten und die Vielfalt der Schulsysteme in den Ländern auf den ersten Blick erkennen. In Bayern, Baden-Württemberg und Niedersachsen verteilen sich die Schülerinnen und Schüler in etwa zu je einem Drittel auf die drei Schularten *Hauptschule*, *Realschule* und *Gymnasium*. In den meisten östlichen Ländern (mit Ausnahme Brandenburgs) hat sich die *Schule mit mehreren Bildungsgängen* als die allgemeinbildende Schulart für die Mehrzahl der Fünfzehnjährigen etabliert. In diesen Schulen, die in Sachsen als *Mittelschule*, in Thüringen als *Regelschule*, in Sachsen-Anhalt als *Sekundarschule* und in Mecklenburg-Vorpommern als *Verbundene Haupt- und Realschule* beziehungsweise als *Regionale Schule* bezeichnet werden, werden Schülerinnen und Schüler von der fünften bis zur zehnten Jahrgangsstufe unterrichtet. Sie streben teilweise einen Hauptschulabschluss, in der Mehrzahl aber einen mittleren Bildungsabschluss an. Im Saarland findet sich eine ähnliche Struktur. Dort werden über 45 Prozent

Tabelle 1.1: Bildungsbeteiligung der Fünfzehnjährigen in Prozent nach Schularten in den Ländern

	Hauptschule	Schule mit mehreren Bildungsgängen	Integrierte Gesamtschule	Realschule	Gymnasium	Berufsschule	Sonder-/Förderschule
Baden-Württemberg	27.9	–	2.0	30.3	27.8	9.0	3.0
Bayern	32.2	–	0.6	27.2	26.3	11.0	2.6
Berlin	11.2	–	28.5	21.6	34.5	1.0	3.2
Brandenburg	–	–	50.1	15.7	30.8	–	3.4
Bremen	21.7	–	16.5	26.7	30.6	1.1	3.4
Hamburg	10.6	5.1	26.5	14.9	33.4	5.9	3.6
Hessen	15.6	–	17.1	27.0	31.7	5.2	3.3
Mecklenburg-Vorpommern	0.4	53.2	4.0	9.6	25.8	1.7	5.4
Niedersachsen	28.3	0.4	4.1	33.5	26.6	3.6	3.5
Nordrhein-Westfalen	26.6	–	16.2	24.7	28.8	–	3.7
Rheinland-Pfalz	22.9	12.8	4.7	22.2	25.8	8.5	2.9
Saarland	–	45.7	13.5	1.6	25.7	10.8	2.7
Sachsen	–	61.3	–	–	32.0	2.6	4.0
Sachsen-Anhalt	–	60.9	1.1	–	30.6	2.3	5.2
Schleswig-Holstein	29.3	–	7.3	31.4	25.2	3.4	3.5
Thüringen	–	57.6	1.4	–	30.5	5.6	4.9

der Schülerinnen und Schüler in Schulen mit mehreren Bildungsgängen unterrichtet. Von diesen Schülerinnen und Schülern streben über 30 Prozent den mittleren Bildungsabschluss an.

Für einen beträchtlichen Anteil der Jugendlichen aus *Schulen mit mehreren Bildungsgängen* ist jedoch zum Zeitpunkt des PISA-Tests noch nicht entschieden, welchen Abschluss sie anstreben. Der Grund dafür ist, dass sich die Fünfzehnjährigen auf unterschiedliche Jahrgangsstufen verteilen. So ist zum Teil in der siebten oder achten Jahrgangsstufe diese Entscheidung noch nicht getroffen worden. Ein weiterer Anteil der Jugendlichen in diesen Schularten besucht auch sogenannte integrierte Bildungsgänge, die weder dem Realschul- noch dem Hauptschulbildungsgang zuzuordnen sind. Deshalb wurde für den vorliegenden Bericht die Entscheidung getroffen, die Ergebnisse für die Länder nicht nach Bildungsgängen, sondern nach Schularten zu berichten.

Die Anteile der Jugendlichen in *Berufsschulen* sind in den Ländern sehr unterschiedlich. Während in Brandenburg und Nordrhein-Westfalen keine Fünfzehnjährigen in Berufsschulen anzutreffen sind, liegen die Anteile in den anderen Ländern zwischen einem Prozent in Bremen und 11 Prozent im Saarland und in Bayern. Die relativen Anteile der Fünfzehnjährigen in den *Sonder- und Förderschulen* betragen zwischen 2,6 Prozent in Bayern und 5,4 Prozent in Mecklenburg-Vorpommern.

### 1.4.2 Die Stichprobenziehung für den Ländervergleich

Die Stichprobe der nationalen PISA-Erweiterung wurde nach den gleichen Gesichtspunkten gezogen wie die für den internationalen Vergleich: PISA untersucht Fünfzehnjährige, die sich zum Testzeitpunkt noch im Schulsystem befinden. Das sind in Deutschland 96,6 Prozent der Fünfzehnjährigen. Für den internationalen Vergleich waren 220 Schulen per Zufall ausgewählt worden. Für den Ländervergleich wurde diese Stichprobe um 1307 Schulen ergänzt. Innerhalb dieser Schulen wurde wiederum eine Stichprobe von Fünfzehnjährigen für die Untersuchung durch ein Zufallsverfahren ausgewählt. Nicht in die Tests und in die Stichprobenziehung einbezogen wurde eine sehr kleine Gruppe von Jugendlichen, die einige internationale exakt definierte Ausschlusskriterien erfüllte (körperliche, geistige oder emotionale Probleme oder weniger als ein Jahr Deutschunterricht in Deutschland). Entsprechend wurden auf Schulebene in Deutschland die Schulen für geistig, körperlich und mehrfach Behinderte und Kranke ausgeschlossen.

Die für den Ländervergleich erforderliche Stichprobengröße beträgt auf der Schulebene 30 Fünfzehnjährige. Zusätzlich wurden in jeder Schule bis zu fünfzehn Jugendliche (Fünfzehnjährige) mit einem Migrationshintergrund zusätzlich in die Stichprobe aufgenommen (*Oversampling*). In die Stichprobe der Sonder- und Förderschulen und der Berufsschulen wurden jeweils alle Fünfzehnjährigen einer Schule einbezogen.

Die Zahl der gezogenen Schulen je Land und Schulart wurde so festgelegt, dass möglichst gleiche Messgenauigkeiten in jedem Land erreicht werden. Für diese Optimierung wurde auf die erreichten Messgenauigkeiten aus dem Ländervergleich von PISA 2000 zurückgegriffen. Daher unterscheiden sich die Anzahlen von Schulen in den Ländern. Über eine geeignete Gewichtung der Stichprobendaten wird sichergestellt, dass die berichteten Ergebnisse auch bei der Optimierung und bei zusätzlichem *Oversampling* von Jugendlichen mit Migrationshintergrund repräsentativ für die Fünfzehnjährigen in den Bildungssystemen der Länder sind (vgl. Kapitel 14).

Die Verteilung der Schulen in der PISA-Erweiterung nach Ländern und Schularten ist in Tabelle 1.2 zusammengefasst. Für den Ländervergleich wurden keine zusätzlichen Sonder- und Förderschulen untersucht. Über diese Schularten werden im Länderbericht keine Aussagen getroffen. Dies gilt auch für die Berufsschulen. Sie müssen jedoch in die Untersuchung einbezogen werden, da in einigen Ländern nennenswerte Anteile von Fünfzehnjährigen sich bereits in Berufsschulen (z.B. auch in Berufsvorbereitungsjahren) befinden. Die Anzahl der Berufsschulen wurde im Vergleich zu PISA 2000 nahezu verdoppelt, um zuverlässigere Schätzungen für die Leistungen dieser Jugendlichen zu erhalten. Die Zahl der Beruflichen Schulen setzt sich aus 10 Schulen der internationalen PISA-Stichprobe und 42 weiteren Beruflichen Schulen der Ergänzungsstichprobe zusammen.

Die Zahl der 1527 Schulen verringert sich um 36 Schulen, die geschlossen wurden oder an denen keine Fünfzehnjährigen mehr unterrichtet wurden. Vier Schulen der Ergänzungsstichprobe haben die Teilnahme an der Untersuchung verweigert, so dass



Tabelle 1.2: Schulen der PISA-E-Stichprobe nach Ländern und Schulart

Land	Haupt- schule	Schule mit mehreren Bildungs- gängen	Integrierte Gesamt- schule	Real- schule	Gymnasium	Berufs- schule	Sonder-/ Förder- schule	Gesamt
Baden-Württemberg	26	–	2	26	23	10	2	89
Bayern	24	–	2	21	19	10	1	77
Berlin	25	–	29	25	33	2	–	114
Brandenburg	–	–	35	20	20	2	1	78
Bremen	26	–	17	29	34	2	–	108
Hamburg	35	16	37	29	42	2	–	161
Hessen	21	–	27	28	31	4	1	112
Mecklenburg-Vorpommern	2	44	7	23	26	2	–	104
Niedersachsen	24	2	5	26	21	4	1	83
Nordrhein-Westfalen	29	–	24	24	27	–	2	106
Rheinland-Pfalz	30	8	5	22	24	4	–	93
Saarland	–	30	12	11	23	2	–	78
Sachsen	–	59	–	–	26	2	1	88
Sachsen-Anhalt	–	38	2	–	18	2	–	60
Schleswig-Holstein	28	–	15	25	20	2	1	91
Thüringen	–	53	2	–	28	2	–	85
<b>Summe PISA-E</b>	<b>270</b>	<b>250</b>	<b>221</b>	<b>309</b>	<b>415</b>	<b>52</b>	<b>10</b>	<b>1527</b>

Daten aus 1487 Schulen in die Analysen eingehen. Die Teilnahmequote auf Schulebene beträgt 99,7 Prozent.

#### *Untersuchungsbeteiligung nach Land und Schulart*

Die Ausschöpfung der Stichprobe bezogen auf den Test ist in Tabelle 1.3 wiedergegeben. Die linke Spalte weist die Teilnahmequoten für den Test, die rechte die Quoten für den Schülerfragebogen aus.

Wie die letzten beiden Spalten der Tabelle zeigen, wurde die Stichprobe in allen Ländern sehr gut ausgeschöpft. In allen Ländern liegen die Teilnahmequoten für den Test und den Fragebogen über dem international geltenden Kriterium von 80 Prozent. Die Teilnahmequoten für den Test variieren auf Länderebene zwischen 85 Prozent im Saarland und 96 Prozent in Brandenburg, Sachsen und Thüringen. Für den Fragebogen erstrecken sich die Quoten von 95 Prozent in Brandenburg bis zu 80 Prozent in Hamburg. Diese Variationsbreite ist unter anderem auf Unterschiede in den Schulgesetzen der Länder zurückzuführen. In einer Reihe von Ländern ist die Teilnahme am Test verpflichtend. In einem Land (Brandenburg) ist auch die Teilnahme an der Befragung (Schülerfragebogen) verpflichtend (und damit keine Einverständniserklärung der Eltern erforderlich). Mögliche Auswirkungen unterschiedlicher Beteiligungsquoten auf die Ergebnisse werden in Abschnitt 1.4.3 diskutiert.

Betrachtet man die Teilnahmequoten in den Schularten der einzelnen Länder, liegen die Testbeteiligungen an den Hauptschulen, den Schulen mit mehreren Bildungsgängen, den Realschulen, den Integrierten Gesamtschulen und den Gymnasien meist deutlich

Tabelle 1.3: Untersuchungsbeteiligung der Fünfzehnjährigen in Prozent für Test und Fragebogen (FB) nach Ländern und Schulart

	Haupt- schule		Schule mit mehreren Bildungs- gängen		Integrierte Gesamt- schule		Real- schule		Gymnasium		Berufs- schule		Summen	
	Test	FB	Test	FB	Test	FB	Test	FB	Test	FB	Test	FB	Test	FB
Baden-Württemberg	91	91	–	–	88	88	91	91	93	93	59	59	89	89
Bayern	88	88	–	–	84	84	94	94	96	96	68	69	90	89
Berlin	89	72	–	–	94	82	96	87	97	92	65	42	94	84
Brandenburg	–	–	–	–	95	95	96	96	98	98	–	–	96	95
Bremen	89	85	–	–	96	93	95	89	97	92	100	100	95	90
Hamburg	89	65	94	82	95	79	96	78	97	89	23	18	94	80
Hessen	92	90	–	–	94	89	96	93	97	96	89	74	95	91
Mecklenburg-Vorpommern	100	100	96	89	96	88	96	93	96	94	35	10	96	91
Niedersachsen	88	87	89	89	83	83	94	94	96	96	61	61	91	91
Nordrhein-Westfalen	85	76	–	–	88	82	94	91	96	95	–	–	90	85
Rheinland-Pfalz	94	88	95	86	95	83	96	94	98	96	76	74	95	91
Saarland	–	–	80	80	81	81	93	93	93	93	80	80	85	85
Sachsen	–	–	96	91	–	–	–	–	97	94	64	64	96	91
Sachsen-Anhalt	–	–	93	89	94	89	–	–	98	95	67	33	94	90
Schleswig-Holstein	83	83	–	–	84	83	92	92	93	93	38	38	87	87
Thüringen	–	–	97	91	98	74	–	–	97	94	71	44	96	90
<b>Deutschland (E)</b>	<b>89</b>	<b>82</b>	<b>93</b>	<b>88</b>	<b>92</b>	<b>85</b>	<b>95</b>	<b>91</b>	<b>96</b>	<b>94</b>	<b>66</b>	<b>60</b>	<b>93</b>	<b>88</b>

über dem Kriterium von 80 Prozent. Um in Berufsschulen akzeptable Beteiligungsquoten zu erreichen, wurden in PISA 2003 erhebliche Anstrengungen unternommen (z.B. mehrtägige Testleiterbesuche und Testgelegenheiten bei Teilzeitberufsschülerinnen und -schülern). In Berufsschulen sowie Sonder- und Förderschulen beteiligten sich durchschnittlich knapp 70 Prozent, was unter den besonderen Umständen der Testung in diesen Schulen als annehmbar gelten kann. Relativ schlecht ist die Beteiligungsquote an den Berufsschulen in Hamburg, Mecklenburg-Vorpommern und Schleswig-Holstein.

Die Teilnahmequoten für den Fragebogen sind an den Gymnasien, den Realschulen und den Schulen mit mehreren Bildungsgängen in allen Ländern unproblematisch. Geringer fällt die Beteiligung in den Hauptschulen und den Berufsschulen mancher Länder aus. Insgesamt erfüllen die Beteiligungsquoten jedoch in allen Ländern die internationalen Anforderungen. Dementsprechend können alle Länder mit ihren Datensätzen in die Auswertungen einbezogen werden.

### *PISA-I und PISA-E im Vergleich*

Um die Leistungen der Schülerinnen und Schüler in Deutschland international einordnen zu können, ist ein geringerer Stichprobenumfang erforderlich als für den Ländervergleich in Deutschland. Wie bereits ausgeführt wurde, beruht der internationale Vergleich auf einer Stichprobe von 220 Schulen in Deutschland; der Ländervergleich erfordert eine Erweiterung um etwa 1300 Schulen. Beide Stichproben sollen die Leistungen der Fünfzehnjährigen in Deutschland repräsentativ abbilden. Entsprechend dürfen die

Tabelle 1.4: Kompetenzmittelwerte der Fünfzehnjährigen in Deutschland auf der Grundlage der verschiedenen Stichproben

	Gesamtstichprobe des Ländervergleichs (PISA-E)	Deutsche Stichprobe des internationalen Vergleichs (PISA-I)
	MW (S.E.)	MW (S.E.)
Mathematik	502 (1.5)	503 (3.3)
Lesen	491 (1.6)	491 (3.4)
Naturwissenschaften	503 (1.5)	502 (3.6)
Problemlösen	511 (1.3)	513 (3.2)

Kompetenzmittelwerte aus beiden Stichproben sich nicht systematisch, sondern lediglich im Rahmen der Stichprobenfehler (S.E.) unterscheiden.

Wie Tabelle 1.4 zeigt, weichen die Kompetenzmittelwerte für die beiden Stichproben nicht beziehungsweise geringfügig voneinander ab. In allen Fällen liegen die Werte innerhalb des Vertrauensintervalls durch den Stichprobenfehler (S.E.). Anhand der Stichprobe des Ländervergleichs (PISA-E) ergeben sich die gleichen Mittelwerte wie die anhand der PISA-I-Stichprobe bereits berichteten. Damit ist die Vergleichbarkeit der Länderstichproben am internationalen Maßstab gesichert,<sup>1</sup> und es ist gerechtfertigt, die Kompetenzen in den Ländern am jeweiligen OECD-Mittel zu vergleichen, sofern die Ergebnisse für jedes Land als repräsentativ und nicht verzerrt gelten können. Diese Überprüfung ist in den folgenden Abschnitten dargestellt.

### 1.4.3 Repräsentativität der Stichproben

Die Untersuchungsbeteiligung in allen Ländern genügt den Anforderungen, die international bei PISA für die vergleichenden Analysen und die Berichterstattung gestellt werden. Jedoch ließen die in Tabelle 1.2 berichteten Ausschöpfungsquoten in den Ländern und nach Schularten doch eine gewisse Streubreite erkennen. Somit stellt sich etwa die Frage, inwieweit Stichproben tatsächlich vergleichbar sind, die 95 Prozent oder nur 85 Prozent der Grundgesamtheit enthalten. Mit Aspekten der Repräsentativität und Vergleichbarkeit beschäftigen sich die folgenden Abschnitte.

Länder, die eine Teilnahme am Test durch gesetzliche Regelungen verbindlich vorschreiben, erreichen höhere Ausschöpfungsquoten als Länder, in denen eine Einwilligung der Eltern und der Schülerinnen und Schüler erforderlich ist. In beiden Fällen kann freilich eine Erkrankung die Teilnahme am Test verhindern. Der Anteil von kranken Schülerinnen und Schülern wird über die Länder weitgehend vergleichbar sein, und eine Krankheit wird nicht vorzugsweise leistungsstarke oder leistungsschwache Jugend-

<sup>1</sup> Die – aus Gründen der Konsistenz – mit der internationalen Stichprobe berechneten Durchschnittswerte in den Tabellen und Abbildungen in diesem Band werden mit *Deutschland* bezeichnet. Die Werte, die mit der Erweiterungsstichprobe berechnet sind, werden *Deutschland (E)* genannt.

liche treffen. Ist die Teilnahme prinzipiell freigestellt, gibt es sehr viele Gründe, die von der Mitwirkung abgehalten haben können. So könnten unter diesen Umständen sehr viele Schülerinnen und Schüler eine Testbeteiligung als unnötig betrachten oder einfach vergessen, die Einwilligungsschreiben rechtzeitig zum Testzeitpunkt vorzulegen. Manche der Gründe für eine Nichtteilnahme könnten die Vergleichbarkeit der Ergebnisse zwischen den Ländern beeinträchtigen. Denkbar ist, dass Eltern aus politischen Überzeugungen ihren Kindern die Teilnahme verweigerten. Eltern aus bestimmten Milieus könnten es häufiger versäumt haben, das Einwilligungsschreiben zu unterzeichnen und ihren Kindern mitzugeben. Vor allem aber könnten Schülerinnen und Schüler, die generell Probleme mit der Schule und mit dem Lernen haben, auf eine Testteilnahme verzichten wollen. Nicht auszuschließen ist auch, dass eher leistungsschwächeren Schülerinnen und Schülern bedeutet wird, dass ihre Testteilnahme nicht unbedingt erforderlich ist.

Entsprechende Überlegungen betreffen mögliche Verzerrungen der Ergebnisse durch eine *selektive Teilnahme* am Test. Um die Vergleichbarkeit sicherzustellen, muss ausgeschlossen werden, dass Hinweise auf eine selektive Teilnahme am Test zu finden sind. Es gibt verschiedene Möglichkeiten, diese Kontrolle durchzuführen (vgl. Baumert et al., 2002b, S. 27). Am effektivsten ist eine Kontrolle anhand der Schulnoten (vgl. auch Prenzel et al., 2004b). Für alle Schülerinnen und Schüler, die für die Teilnahme an PISA ausgewählt worden waren, wurden von den Schulen die Informationen über ihre Noten in Hauptfächern abgefragt. Damit lagen auch Informationen über die Leistungsstärke der Jugendlichen vor, die nicht am Test teilnahmen. Da die Schulnoten die Leistungsunterschiede in den Klassen gut repräsentieren, kann anhand der Noten zuverlässig abgeschätzt werden, ob die am Test teilnehmenden Schülerinnen und Schüler im Mittel höhere Leistungen zeigen als die kompletten Klassen (vgl. auch Kienzl, 2005).

Tabelle 1.5 zeigt, dass sich für kein Land die Durchschnittsnote der Teilnehmerinnen und Teilnehmer von der Durchschnittsnote aller gezogenen Schülerinnen und Schüler unterscheidet. Auch nach Schularten getrennt ergibt sich das gleiche Bild. Lediglich in den (vier) Integrierten Gesamtschulen in Niedersachsen sind die Noten der Fünfzehnjährigen, die am Test teilnehmen, signifikant besser als die aller gezogenen Jugendlichen.<sup>2</sup> Über die Schulen aller Schularten in Niedersachsen gesehen, zeigt sich jedoch kein Unterschied in den Noten zwischen den teilnehmenden und allen gezogenen Schülerinnen und Schülern. Insgesamt konnten über alle Schulen und Schularten der sechzehn Länder aufgrund der Notenvergleiche keine Hinweise auf eine selektive Testteilnahme gefunden werden. Damit belegen die durchgeführten Analysen, dass die Stichproben der Länder auch bei unterschiedlichen Ausschöpfungsquoten nicht verzerrt und damit vergleichbar sind.

---

2 Hier haben 21 von 114 Jugendlichen aus Integrierten Gesamtschulen nicht am Test teilgenommen.

Tabelle 1.5: Schulnoten (Mittelwerte und Streuungen) der für den Test ausgewählten und am Test teilnehmenden Fünfzehnjährigen nach Ländern

Land	Noten				Unterschied*
	Alle gezogenen Fünfzehnjährigen		Teilnehmerinnen und Teilnehmer		
	Mittelwert	SD	Mittelwert	SD	
Baden-Württemberg	3.06	1.0	3.05	1.0	n.s.
Bayern	3.32	1.1	3.27	1.1	n.s.
Brandenburg	3.22	1.0	3.20	1.0	n.s.
Berlin	3.44	1.1	3.40	1.0	n.s.
Bremen	3.40	1.1	3.37	1.1	n.s.
Hamburg	3.33	1.0	3.30	1.0	n.s.
Hessen	3.30	1.1	3.27	1.0	n.s.
Mecklenburg-Vorpommern	3.17	0.9	3.16	0.9	n.s.
Niedersachsen	3.40	1.0	3.35	1.0	n.s.
Nordrhein-Westfalen	3.35	1.1	3.30	1.0	n.s.
Rheinland-Pfalz	3.31	1.0	3.28	1.0	n.s.
Saarland	3.25	1.1	3.20	1.1	n.s.
Sachsen	3.12	1.0	3.11	0.9	n.s.
Sachsen-Anhalt	3.14	1.0	3.10	1.0	n.s.
Schleswig-Holstein	3.36	1.0	3.33	1.0	n.s.
Thüringen	2.96	1.0	2.95	1.0	n.s.

\*:  $p = 0.05$  n.s. = nicht signifikant

#### 1.4.4 Effekte von Üben und Testmotivation

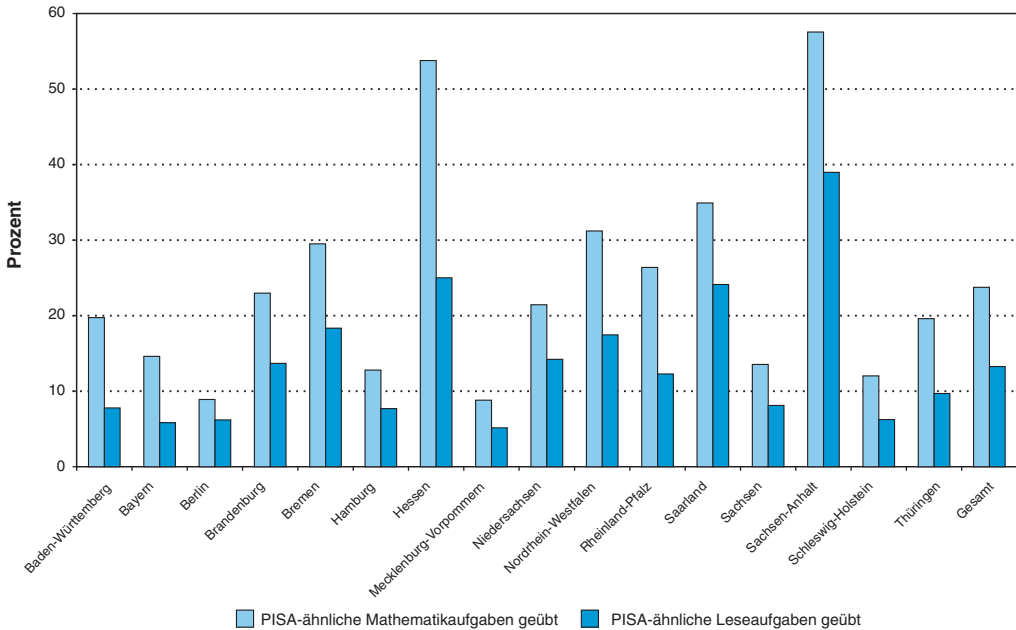
##### *Gibt es Auswirkungen von Üben für den PISA Test?*

Die Frage nach Übungseffekten in *low stakes tests* (also Tests wie PISA, die anders als z.B. Einstufungstests oder benotete Tests für die getestete Person wenig relevant sind) wurde nach PISA 2000 vielerorts diskutiert. Brunner, Artelt, Krauss & Baumert (in Druck) führten daher eine Studie zu Coaching- und Übungseffekten in Mathematik und Lesen durch. Sie fanden geringe Effekte eines gezielten Testtrainings und Übens auf die Leistung, im Gymnasium deutlicher als in Hauptschulen und im Fach Mathematik stärker als im Lesen. Solange nicht eine Kombination aus Vortest und Coaching eingesetzt wurde, sind die Effekte nicht nennenswert. Die Autoren kommen zu dem Schluss, dass Coaching-Effekte für PISA als sehr unwahrscheinlich eingeschätzt werden können, da das Testmaterial unter Verschluss bleibt und es keine Möglichkeit zum Vortesten gibt.

Um der Frage nach der Vergleichbarkeit der Testergebnisse in PISA 2003 weiter nachgehen zu können, wurde im Schülerfragebogen gefragt, ob im Unterricht für den PISA-Test geübt worden war. Die Angaben der Schülerinnen und Schüler geben einen Anhaltspunkt dafür, ob in einzelnen Ländern (und/oder Schularten) systematisch versucht wurde, die Fünfzehnjährigen auf den Test vorzubereiten. Nach Angaben der Schülerinnen und Schüler wurden, wenn Übungen stattfanden, Aufgaben verwendet, die den PISA-Aufgaben ähnlich waren. Für Mathematik wurde mehr geübt als für den Bereich Lesen. Im Vergleich der Länder fällt auf, dass *PISA-ähnliche* Aufgaben häufiger in Hes-

sen und Sachsen-Anhalt geübt wurden. Nach den Angaben der Schülerinnen und Schüler wurde am meisten in den Gymnasien und am wenigsten in den Hauptschulen geübt (vgl. Abbildung 1.3).

Abbildung 1.3: Anteile von Jugendlichen, die PISA-ähnliche Aufgaben in Mathematik und Lesen geübt haben



Um zu prüfen, inwieweit sich die Kompetenzen zwischen den Jugendlichen unterscheiden, die angegeben haben, PISA-ähnliche Aufgaben geübt oder nicht geübt zu haben, werden mittlere Kompetenzunterschiede betrachtet. Da in den Schularten unterschiedlich viele Jugendliche angeben, geübt zu haben, werden die Unterschiede nach Schulart und Land differenziert betrachtet (vgl. die Analyse für PISA-I, Prenzel et al., 2004b, S. 40). In Tabelle 1.6 sind diese Unterschiede wiedergegeben.

Die Unterschiede fallen sowohl positiv als auch negativ aus, das heißt, die Kompetenzen der Jugendlichen, die angeben, geübt zu haben, sind zum Teil höher und zum Teil niedriger als die Kompetenzen derjenigen Jugendlichen, die keine PISA-ähnlichen Mathematik-Aufgaben geübt haben. In keiner Schulart je Land werden die Unterschiede statistisch signifikant, so dass von keinem systematischen Einfluss auf die Kompetenz ausgegangen werden kann.

Tabelle 1.6: Differenzen in der Mathematik-Kompetenz zwischen Jugendlichen, die angeben, PISA-ähnliche Mathematik-Aufgaben geübt beziehungsweise nicht geübt zu haben

Land	Hauptschule	Schule mit mehreren Bildungsgängen	Integrierte Gesamtschule	Realschule	Gymnasium
Baden-Württemberg	5	–	–	12	8
Bayern	5	–	–	15	16
Berlin	-17	–	-33	-21	-12
Brandenburg	–	–	10	-9	-2
Bremen	22	–	-4	15	-16
Hessen	2	–	3	12	-1
Hamburg	-1	-38	16	-8	5
Mecklenburg-Vorpommern	–	0	–	-19	1
Niedersachsen	14	–	–	-13	5
Nordrhein-Westfalen	15	–	17	8	-7
Rheinland-Pfalz	-5	15	–	-8	4
Saarland	–	9	-4	–	-6
Sachsen	–	-8	–	–	-8
Sachsen-Anhalt	–	8	–	–	-15
Schleswig-Holstein	15	–	33	30	16
Thüringen	–	-4	–	–	0

Eine positive Differenz weist auf einen Kompetenzvorteil der Jugendlichen, die angeben, geübt zu haben

### *Wie stark haben sich die Jugendlichen im Test angestrengt?*

Wie in Prenzel et al. (2004b, S. 40) dargestellt, berichteten die Schülerinnen und Schüler in PISA 2003 eine höhere Testmotivation als in PISA 2000. Die Schülerinnen und Schüler waren gebeten worden, anhand einer Skala ihre Anstrengung bei der Testbearbeitung einzuschätzen. Da speziell in Deutschland aufgrund der öffentlichen Aufmerksamkeit für PISA Auswirkungen auf die Testmotivation nicht auszuschließen waren, haben wir die entsprechenden Kennwerte für die beiden PISA-Erhebungen verglichen. Die Befunde weisen darauf hin, dass die Schülerinnen und Schüler in Deutschland 2003 den PISA-Test ernster nahmen als 2000. Tabelle 1.7 zeigt die Testmotivation, die auf einer Skala zwischen 1 (geringe Testmotivation) und 10 (sehr hohe Testmotivation) anzugeben war, für die Länder.

In allen Ländern erreichen die Jugendlichen einen Mittelwert von mindestens 7 auf der Skala von 1 bis 10. Es ist zu erkennen, dass sich die Schülerinnen und Schüler in den Ländern in ihrer Testmotivation unterscheiden. Allerdings sind Unterschiede nicht gravierend. In allen Ländern berichten die Schülerinnen und Schüler 2003 über eine (statistisch signifikant) höhere Bereitschaft, sich im PISA-Test anzustrengen.

Tabelle 1.7: Testmotivation für PISA 2000 und PISA 2003, Einschätzskalen am Ende der Testhefte

Land	PISA 2000		PISA 2003		Unterschied in Effektstärke <i>d</i>
	MW	SD	MW	SD	
Baden-Württemberg	7.4	2.0	7.6	1.8	<b>0.13</b>
Bayern	7.2	1.9	7.5	1.8	<b>0.16</b>
Berlin	7.2	2.0	7.4	2.0	<b>0.07</b>
Brandenburg	7.4	1.9	7.7	1.8	<b>0.14</b>
Bremen	6.8	2.2	7.5	1.9	<b>0.32</b>
Hamburg	6.7	2.1	7.1	2.1	<b>0.17</b>
Hessen	7.0	2.1	7.4	2.0	<b>0.20</b>
Mecklenburg-Vorpommern	7.4	1.9	7.6	1.8	<b>0.12</b>
Niedersachsen	6.9	2.1	7.3	2.0	<b>0.21</b>
Nordrhein-Westfalen	7.0	2.1	7.5	1.9	<b>0.24</b>
Rheinland-Pfalz	7.1	2.1	7.6	1.9	<b>0.24</b>
Saarland	7.1	2.0	7.4	1.9	<b>0.16</b>
Sachsen	7.2	1.9	7.6	1.8	<b>0.22</b>
Sachsen-Anhalt	7.3	2.0	7.8	1.8	<b>0.30</b>
Schleswig-Holstein	6.9	2.1	7.4	1.9	<b>0.27</b>
Thüringen	7.4	1.8	7.7	1.8	<b>0.12</b>

Signifikante Unterschiede sind in Fettschrift gekennzeichnet

### 1.4.5 Durchführung der Erhebung

Die Erhebungen zum Ländervergleich 2003 wurden in enger Zusammenarbeit zwischen dem PISA-Konsortium Deutschland, dem Data Processing Center (DPC) der *International Association for the Evaluation of Educational Achievement* (IEA) in Hamburg sowie den in jedem Land zuständigen PISA-Beauftragten und den Projekt-Koordinatorinnen geplant und durchgeführt. Die Datenschutzbeauftragten der Länder waren an der Genehmigung der Testung in jedem Land beteiligt.

Je nach Gesetzeslage in den Ländern war die Teilnahme am Test und/oder am Fragebogen verpflichtend oder freiwillig und damit abhängig von der Genehmigung der Eltern. Diese Regelungen können die Ausschöpfung der Stichprobe in den Ländern unterschiedlich beeinflussen (vgl. Abschnitt 1.4.3).

Die Verantwortung für die Durchführung der Erhebung in den Schulen der Erweiterungsstichprobe lag bei den Ländern selbst. In jedem Land wurden PISA-Beauftragte sowie PISA-Koordinatorinnen und -Koordinatoren benannt. Die Koordinatorinnen und Koordinatoren in den Ländern wurden über die Ziele der Studie informiert und durch das DPC und das PISA-Konsortium sorgfältig auf ihre Aufgaben vorbereitet. Sie waren dafür verantwortlich, die Testleiterinnen und Testleiter zu rekrutieren, die Schülerlisten vorzubereiten, die Testtermine zu organisieren und die Schulen zu betreuen.

Nach der Ziehung der Schulstichprobe durch das DPC wurden die Ministerien darüber informiert, welche Schulen in jedem Land für die Teilnahme an PISA ausgewählt wurden. Die gezogenen Schulen wurden von ihren Ministerien und dem PISA-Konsortium Deutschland jeweils in einem Brief um ihre Mitarbeit gebeten. Die Schul-



leitungen der gezogenen Schulen hatten die Möglichkeit, an Veranstaltungen zu PISA 2003 teilzunehmen, in denen ausführlich über die Ziele und Inhalte der Studie sowie den Ablauf und die organisatorischen Besonderheiten der Erhebung informiert wurde. In jeder Schule wurde ein PISA-Koordinator oder eine PISA-Koordinatorin für die notwendigen organisatorischen Arbeiten benannt.

Die Auswahl der Testleiterinnen und Testleiter für die PISA-Erweiterung wurde in den Ländern vorgenommen. Das DPC schulte die Testleiterinnen und Testleiter und machte sie mit den international festgelegten Prozeduren vertraut. An einem Testtag im Zeitraum von März bis Mai 2003 wurden in den Schulen die Tests durchgeführt. Zur Sicherstellung der Testsicherheit und der Einhaltung der international festgelegten Regeln hat das PISA-Konsortium Deutschland vielfältige Maßnahmen ergriffen. Dazu gehört unter anderem, alle an der Studie Beteiligten sorgfältig zu informieren. Alle Personen, die Einsicht in die Materialien hatten, wurden schriftlich zur Vertraulichkeit verpflichtet. Die Landesbeauftragten für den Datenschutz wurden in die Festlegung der Prozeduren einbezogen, die die Anonymität der Teilnehmenden sicherstellte.

#### *Qualitätsmonitoring und Standards in der Durchführung*

Die Durchführung der Erhebungen bei PISA folgt einem differenzierten Plan. Ausführliche Manuale regeln die Vorbereitungsarbeiten an den Schulen und das Handeln der Testleiter, die für ihre Aufgaben gründlich geschult werden. Für die internationale Schulstichprobe und für die Berufsschulen wurden zentral Testleiterinnen und Testleiter ausgewählt und eingesetzt. Für die Schulen der Erweiterungsstichproben wurden Testleiterinnen und Testleiter in den Ländern gewonnen und zentral auf ihre Aufgaben vorbereitet. Außerdem wurde ein Qualitätsmonitoring durchgeführt, bei dem zufällig ausgewählte Schulen der nationalen und internationalen PISA-Stichprobe Besuch von geschulten Beobachtern erhielten. Die Beobachter hatten die Aufgabe, die gesamte Testsituation (inklusive deren Vorbereitung und Abschluss) zu überwachen und mit Hilfe eines standardisierten Fragebogens zu protokollieren. Das Ziel des PISA-Qualitätsmonitorings bestand darin, verlässliche Informationen über die Validität der erhobenen Daten zu gewinnen. Aus der Erweiterungsstichprobe wurde eine Stichprobe von 100 Schulen gezogen, die am Testtag von den Beobachtern besucht wurden. Der Bericht über das Qualitätsmonitoring gelangt insgesamt zu der Aussage, dass die Testung in allen entscheidenden Punkten dem vorgesehenen Reglement entsprach (vgl. Köller, 2003). In einzelnen Fällen wurden kleinere Abweichungen vom Erhebungsskript festgestellt, die allerdings für die Qualität der Datenerhebung ohne Belang waren.

### 1.4.6 Auswertung und Skalierung

Die Auswertung der Daten der Erweiterungsstichprobe erfolgte nach Abschluss der Auswertung der internationalen Stichprobe und nach dem gleichen Plan. Die Datenbasis enthält ca. 535 Variablen. Die Testhefte wurden eingesehen und die „offenen“ Testauf-

gaben wurden – z.T. durch mehrere unabhängige Auswerter – kodiert. Diese Arbeiten wurden am DPC in Hamburg durchgeführt. Die Daten der Erweiterungsstichprobe wurden am IPN weiter aufbereitet und skaliert. Dabei wurden exakt die gleichen Verfahren angewendet, die auch bei der internationalen Skalierung und Datenanalyse genutzt wurden (OECD, 2005, vgl. auch Kapitel 14).

Das Ziel der *Skalierung* ist es, die Antworten der Schülerinnen und Schüler auf die Testaufgaben auf einer Skala abzubilden, so dass die Kompetenzen der Jugendlichen miteinander verglichen werden können, auch wenn die Jugendlichen unterschiedliche Testhefte bearbeitet haben. Die Skalierung dient auch dazu, die Schwierigkeit der Testaufgaben zu bestimmen.

Die bei PISA 2003 verwendeten Skalierungsverfahren sind in Carstensen, Knoll, Rost & Prenzel (2004) dargestellt. International wie national wurden bei der Auswertung der Testdaten Analyseverfahren verwendet, die auf Modellen der Item-Response-Theorie beruhen (vgl. Adams & Wu, 2002; OECD, 2005; Rost, 2004). Auswertungen mit diesem Ansatz liefern Informationen über die Dimensionalität eines Tests und seine Messgenauigkeit. Anhand der Itemkennwerte können die gemessenen Kompetenzen der Schülerinnen und Schüler unter inhaltlichen Bezugskriterien interpretiert und – theoretisch begründet – so genannten Kompetenzstufen zugeordnet werden. Dieser Ansatz gestattet es ebenso, die Verteilungen von Kompetenzen zu betrachten und die Leistungen von Schülergruppen (bis hin zu nationalen Stichproben) zu vergleichen. Eine für PISA wesentliche Eigenschaft der Verfahren der Item-Response-Theorie besteht darin, dass auch bei einem Testdesign mit rotierten Aufgabenblöcken die Kompetenzen der Schülerinnen und Schüler auf einer gemeinsamen Skala abgebildet und verglichen werden können. Will man anhand eines Testdesigns mit mehreren Testheften, unterschiedlichen Aufgabenblöcken und vergleichsweise wenigen Aufgaben je Schülerin oder Schüler zu den einzelnen Kompetenzbereichen die Mittelwerte und Streuungen für Populationen korrekt schätzen, sind weiterführende Datenaufbereitungs- und Analyseverfahren erforderlich (vgl. Adams & Wu, 2002; Mislevy, Beaton, Kaplan & Sheehan, 1992). Auch hier wurden bei der Auswertung der Erweiterungsstudie die international gebräuchlichen Verfahren angewendet.

Etwas einfacher stellt sich die Auswertung der *Fragebogendaten* dar. Allerdings waren auch hier zunächst einige Kodierungsschritte bei „offenen“ Fragen (z.B. zur Klassifikation der Berufsangaben) erforderlich. Eine ganze Reihe der mit Fragebögen erhobenen Daten bezieht sich auf Merkmale, die mit einer Angabe beschrieben werden können (z.B. Alter oder Geschlecht). Die Fragebögen zielen aber auch auf Merkmale („Konstrukte“), die erst über mehrere Fragen (z.B. auch über Einschätzfragen, so genannte Ratings) zuverlässig eingeschätzt werden können. Typische Beispiele für solche Konstrukte sind zum Beispiel das Mathematikinteresse oder die Einschätzung des Schulklimas. Die verschiedenen Fragen (bzw. Items) zu diesem Konstrukt müssen ebenfalls skaliert werden. Auch hier informieren Analysen der Dimensionalität oder Zuverlässigkeit über die Qualität und die Eignung der Skala für weitere Auswertungen. Das PISA-

Konsortium Deutschland hat auch die Konstrukte der internationalen Fragebögen auf der Basis der Item-Response-Theorie analysiert.

### 1.4.7 Berichterstattung und Darstellung

Der vorliegende Bericht über den Ländervergleich im Rahmen von PISA 2003 konzentriert sich auf eine Darstellung deskriptiver Befunde. In erster Linie beschreiben wir bedeutsame *Bildungsergebnisse*, die in den Ländern der Bundesrepublik Deutschland erreicht wurden.

Die Bildungsergebnisse kann man aus unterschiedlichen *Perspektiven* betrachten und analysieren. Eine erste Perspektive bietet der internationale Vergleich. Wir wollen die Befunde zweitens unter inhaltlichen Kriterien beleuchten, die aus Ansprüchen an eine zeitgemäße und anschlussfähige Grundbildung im Sinne von Literacy resultieren. Unter einer dritten Perspektive liegt es nahe, die Ergebnisse zu vergleichen, die bei PISA 2000 und 2003 erzielt wurden.

PISA gibt auch die Möglichkeit, über Merkmale von schulischen und außerschulischen Lernumgebungen zu berichten, die aus einer theoretischen Sicht als Ressourcen oder Voraussetzungen für die Entwicklung von Kompetenz gelten können. Auch hier können die Befunde für sich stehend unter Qualitätskriterien diskutiert werden. Allerdings liegt hier die Versuchung nahe, aus dem Vergleich von Bildungsergebnissen und Merkmalen von Lernumgebungen *Erklärungen* (Aussagen über Ursachen) für feststellbare Unterschiede abzuleiten. Statistische Auswertungen, die korrelative Zusammenhänge zwischen Umgebungsmerkmalen und Kompetenzen beschreiben, werden gerne als Bedingungsanalysen verstanden, obwohl sie tatsächlich keine Aussagen über Bedingungen zulassen (allenfalls Vorhersagen von Kompetenzunterschieden auf der Basis von Merkmalsunterschieden, die oft als Varianzerklärung bezeichnet werden). So anregend solche Assoziationen und Schlussfolgerungen sein mögen, sie sind nicht mehr als Vermutungen, die einer kritischen empirischen Überprüfung bedürfen. Der PISA-Datensatz lässt eine angemessene Prüfung von Hypothesen über kausal relevante Bedingungen nicht zu. Das PISA-Konsortium Deutschland wird auch in diesem Band keine Aussagen über mögliche Ursachen (kausal relevante Bedingungen) der Bildungsergebnisse treffen. Wir verweisen auf nachfolgende Analysen mit einem erweiterten nationalen Design, das durch zwei Erhebungszeitpunkte und systematische Erhebungen auf mehreren Ebenen empirisch tragfähige Bedingungsanalysen und weiterführende Erkenntnisse zulässt. Über diese Ergebnisse berichten wir im dritten Band, der im Herbst 2006 erscheinen wird.

Die folgenden Abschnitte erläutern, wie wir in diesem Band die Befunde unter den drei genannten Perspektiven darstellen.

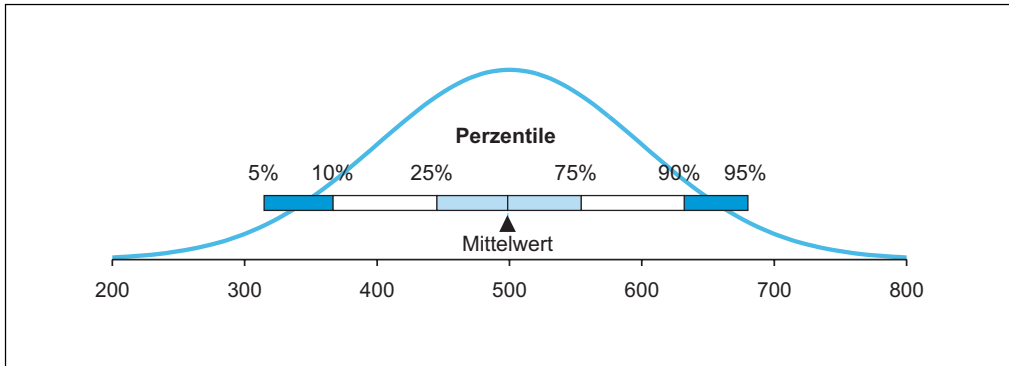
In der PISA-Rezeption wird der *Vergleich von Staaten und Ländern* häufig nur als Rangordnung gelesen. Tatsächlich werden die Rangordnungen oft unzulässig interpretiert und viele andere Informationen aus dem Vergleich übersehen. Bezugspunkt für die

Vergleiche sind die international skalierten Tests. So wie es mehrere Möglichkeiten gibt, die Temperatur zu skalieren, können auch *Kompetenzskalen* unterschiedliche Zahlenwerte und Abstände zugeordnet werden. Die internationalen Kompetenzskalen sind so definiert, dass ihr *Mittelwert bei 500* liegt. Dieser Wert entspricht also dem internationalen Durchschnitt. Die internationale Skala ist weiterhin dadurch bestimmt, dass die *Standardabweichung* bei einem Wert von 100 liegt. Die Standardabweichung ist eine statistische Maßzahl, die die durchschnittliche Streuung der Werte einer Verteilung um den Mittelwert beschreibt. Bei PISA liegen im Bereich zwischen 400 und 600 (also Mittelwert  $\pm$  eine Standardabweichung) ca. zwei Drittel (genau: 68,2 Prozent) der getesteten Schülerinnen und Schüler. Bei einem Abstand von zwei Standardabweichungen (also 700 bzw. 300 Punkten) nähert man sich schon den Enden der Verteilung. Dann schneiden nur mehr jeweils 2,3 Prozent der Schülerinnen und Schüler besser (als 700 Punkte) beziehungsweise schlechter (als 300 Punkte) ab. Anhand der Maßzahlen „Mittelwert“ (MW) und „Standardabweichung“ (SD) können auch die Ergebnisse der Länder national und international eingeordnet verglichen werden. Erzielt ein Land einen Mittelwert von 600, dann liegt es über dem internationalen Durchschnitt, und zwar erheblich, denn wir wissen, dass dieser Wert international nur mehr von 15,9 Prozent der getesteten Schülerinnen und Schüler übertroffen wird. Aber auch der Kennwert für die Standardabweichung in einem bestimmten Land ist aussagekräftig. Liegt die Standardabweichung über 100, dann streuen dort die Leistungen stärker als im internationalen Durchschnitt. Bei einer Standardabweichung unter 100 wären die Leistungen dementsprechend relativ homogen.

Viele der Tabellen und Abbildungen in diesem Band präsentieren zusätzlich Angaben über so genannte *Perzentile* (genauer für bestimmte Perzentilwerte, nämlich 5, 25, 75 und 95). Diese Maßzahlen informieren darüber, wie hoch der Kennwert für die Kompetenz an bestimmten Abschnitten der Verteilung ist. Bei einem Perzentil von 95 wird also der Punktwert berichtet, ab dem die besten 5 Prozent einer Verteilung beginnen. Die Perzentilwerte 75 und 25 betreffen das obere oder untere Viertel der Verteilung. Die Perzentilwerte informieren somit vor allem darüber, wie stark die Leistungsspitze (die oberen 5 bzw. 25 Prozent) und wie schwach die untersten Gruppen (die unteren 5 bzw. 25 Prozent) sind. Diese Angaben stellen wir graphisch auch in so genannten *Perzentilbändern* dar, um die Kompetenzverteilungen zu veranschaulichen. In Abbildung 1.4 ist ein Perzentilband dargestellt, zur Veranschaulichung ist ebenfalls eine Normalverteilung eingezeichnet. Mit der Darstellung von Perzentilbändern wird keine Normalverteilung der Kennwerte vorausgesetzt.

Die Tabellen zu den nationalen und internationalen Vergleichen enthalten außerdem Angaben über den so genannten *Standardfehler* (S.E.) der Schätzung des Populationskennwertes. Mit Hilfe von Zufallsstichproben wird ja versucht, Aussagen über Merkmalsverteilungen in einer Population (Grundgesamtheit) zu treffen. Allerdings sind diese Schätzungen auf der Basis von Stichproben immer fehlerbehaftet. Die Größe des Fehlers wiederum lässt sich anhand der (gemessenen) Streuung in der Stichprobe im Verhältnis zur Stichprobengröße schätzen (die Wurzel aus diesem Wert ist der Standard-

Abbildung 1.4: Ein Beispiel für ein Perzentilband und eine Normalverteilung



fehler). Mit Hilfe des Standardfehlers kann man mit einer bestimmten *Irrtumswahrscheinlichkeit* (z.B. 5 Prozent) abschätzen, in welchem Wertebereich der „wahre“ Wert der Population liegt (nämlich in einem Intervall von  $\pm$  zwei Standardfehlern). Ein statistischer Test auf Unterschiede in den Mittelwerten zweier Länder (Signifikanztest) berücksichtigt die Standardfehler für beide Mittelwerte gleichzeitig.<sup>3</sup> Wird mehr als ein Signifikanztest an einem Datensatz durchgeführt, erhöht sich die Irrtumswahrscheinlichkeit. Um multiple Vergleiche mit einer festgelegten Irrtumswahrscheinlichkeit über alle Vergleiche durchführen zu können, muss die zugelassene Irrtumswahrscheinlichkeit für jeden einzelnen Vergleich herabgesetzt werden; in PISA wird die Bonferroni-Korrektur verwendet.<sup>4</sup>

Wenn die bei PISA teilnehmenden Staaten und Länder nun anhand ihrer Mittelwerte in einer Tabelle angeordnet werden, erzeugt man eine *Rangfolge*. Die Ausführungen zum Standardfehler für die Populationsschätzung zeigen, dass die Unterschiede in den Stichprobenmittelwerten nicht immer substantielle Unterschiede (zwischen den Populationen) abbilden. Deshalb müssen geeignete statistische Verfahren zum *Mittelwertsvergleich* angewendet werden. In den Vergleichstabellen werden entsprechend *drei Blöcke* gebildet (OECD-Durchschnitt, oberhalb und unterhalb des OECD-Durchschnittes). Innerhalb dieser Blöcke sind Unterschiede zwischen den Staaten statistisch nicht mehr zuverlässig abzusichern. Folgt man dieser Betrachtung, dann verbietet es sich, die Tabellenplätze durchzunummerieren. Im Anhang befinden sich Tabellen, die im Detail über signifikante Unterschiede bei Vergleichen einzelner Staaten berichten.

Wenden wir bei PISA die gebräuchlichen statistischen Verfahren zur Prüfung von Mittelwertsunterschieden an, müssen wir aufgrund der großen Stichprobenumfänge damit rechnen, dass sich auch kleinere Mittelwertsunterschiede als *statistisch signifikant* erwei-

3 Die Durchführung eines Signifikanztests gleicht nicht dem Vergleich zweier Konfidenzintervalle; vielmehr wird ein Konfidenzintervall für die Differenz anhand der Standardfehler beider Mittelwerte gebildet.

4 Bei 34 Vergleichen und 5 Prozent zugelassener Irrtumswahrscheinlichkeit ergibt sich ein kritisches Intervall von 3,18 Standardfehlern, im Vergleich dazu ergeben sich 1,96 Standardfehler bei nur einem Vergleich.

sen können (auch bei Irrtumswahrscheinlichkeiten unter 1 Prozent). Allerdings kann es sein, dass diese kleinen Mittelwertsunterschiede aus einer praktischen Perspektive kaum mehr bedeutsam sind, etwa wenn sich die Verteilungen der beiden Vergleichsgruppen für ein Merkmal fast vollständig überlappen. Aus diesem Grund werden wir bei entsprechenden Vergleichen (und signifikanten Unterschieden) auch das Maß der *Effektstärke* nutzen, um das Ausmaß des Unterschiedes statistisch darzustellen. Die Effektstärke (die Differenz der Mittelwertsunterschiede, geteilt durch die Standardabweichung) gibt an, wie stark sich Verteilungen überlappen. Das in diesem Band verwendete Effektstärkemaß ( $d$ ) wird häufig so interpretiert, dass Werte in der Größenordnung von  $d = 0.2$  als „kleine“ Effekte,  $d = 0.5$  als „mittlere“ und  $d = 0.8$  als „große“ Effekte bezeichnet werden.

Bei der Beurteilung von Kompetenzunterschieden zwischen Gruppen in großen Schul Leistungsvergleichen hat es sich als praktisch erwiesen, diese in ungefähre Entwicklungszeiten (Schuljahre) zu übersetzen. Als Anhaltspunkt dienen hier die Informationen über durchschnittliche Kompetenzunterschiede zwischen Schülerinnen und Schülern verschiedener Klassenstufen. Im Sinne einer groben Faustregel können Unterschiede in der Größenordnung von etwa 40 Punkten in einen *Abstand von einem Schuljahr* umgerechnet werden.

Mit dem Kriterium einer Vorbereitung auf die Teilhabe an der Wissensgesellschaft und mit entsprechenden Testkonzeptionen für die Kompetenzbereiche beschränkt sich PISA nicht auf Mittelwertsvergleiche. Die Länder sollen auch informiert werden, über welche *inhaltlichen Kompetenzen* ihre Schülerinnen und Schüler verfügen und wie sich solche Kompetenzen auf Schülergruppen innerhalb des Landes verteilen. Zu diesem Zweck werden bei PISA für die jeweiligen Schwerpunktgebiete Tests und *Subskalen für Teilkompetenzen* entwickelt. Die Information aus den Gesamt- und Subskalen kann im Rahmen der bei PISA durchgeführten Skalierung weiter aufgeschlüsselt werden. Die Skalierung beschreibt die Schwierigkeit der Testitems. Aus der theoretischen Sicht der Testkonzeption können wiederum mögliche Abstufungen der Kompetenz inhaltlich beschrieben und auf die vorfindbaren Items mit ihren Schwierigkeitskennwerten und Anforderungen bezogen werden. Auf diese Weise können unterschiedliche Niveaus einer Kompetenz bestimmt und anhand von Aufgabenanforderungen inhaltlich charakterisiert werden (vgl. Baumert et al., 2001; Klieme, Baumert, Köller & Bos, 2000). Formal sind die *Kompetenzstufen* so definiert, dass Schülerinnen und Schüler auf dieser Stufe zugeordnete Schwellenitems mit einer bestimmten Wahrscheinlichkeit (62 Prozent) lösen. Aufgaben, die höheren Kompetenzstufen entsprechen, werden mit einer sehr viel geringeren Wahrscheinlichkeit gelöst. Die inhaltliche Interpretation der Tests mit Hilfe von Kompetenzstufen wurde bisher bei PISA (Baumert et al., 2001; OECD, 2001; OECD, 2005; Turner, 2002) dazu genutzt, Gruppen (Anteile) von Schülerinnen und Schülern zu identifizieren, die aufgrund ihrer Kompetenz sehr schlechte Chancen für ein nachfolgendes Lernen innerhalb und außerhalb der Schule haben. Für Schülerinnen und Schüler, die auf oder unterhalb der ersten Kompetenzstufe anzusiedeln sind, ist die Prognose für die weitere Bildungskarriere (auch bezogen auf eine berufliche Ausbildung) ungünstig. Umgekehrt kann im Spitzenbereich (der obersten Kompetenzstufe)

gefragt werden, wie groß die Anteile von Schülerinnen und Schülern sind, die über exzellente Kompetenz auf einem Gebiet verfügen. In den entsprechenden Kapiteln werden wir deshalb über die entsprechenden Anteile von Schülerinnen und Schülern berichten.

An dieser Stelle weisen wir noch auf eine Auswertungs- und Darstellungsform hin, die in diesem Band gelegentlich verwendet wird, wenn *außerhalb der Kompetenzbereiche* über skalierte Merkmale (Aussagen) von Schülerinnen und Schülern, Lehrkräften oder Schulleitungen berichtet wird. Skalen zu solchen Konstrukten (z.B. Motivation, Schulklima) werden häufig so transformiert, dass sie einen Mittelwert von 0 und eine Standardabweichung von 1 aufweisen. Entsprechend transformierte Skalen stellen anschaulich die Unterschiede zwischen Teilgruppen dar, zum Beispiel zwischen Staaten oder Ländern, aber auch innerhalb der Länder beispielsweise zwischen Jungen und Mädchen oder Schularten. Ein Glossar in Kapitel 14 klärt einige weitere methodische Besonderheiten.

Der *Vergleich von Ergebnissen über Erhebungszeitpunkte*, der über die PISA-Zyklen möglich wird, dürfte schließlich die auf längere Sicht interessanteste Perspektive eröffnen. Über mehrere PISA-Erhebungen können die Entwicklungen von Bildungssystemen mit definierten Kriterien und geeigneten Erhebungsverfahren empirisch beschrieben werden. Die Staaten beziehungsweise Länder erhalten dabei auch Rückmeldung darüber, inwieweit Maßnahmen – die möglicherweise als Reaktion auf frühere PISA-Befunde eingeleitet wurden – gegriffen haben und die beabsichtigten Wirkungen erzielen. Dabei muss in Rechnung gestellt werden, dass Eingriffe und Veränderungen in Bildungssystemen sich oft erst mit erheblicher Verzögerung in Bildungsergebnissen niederschlagen: Die Bildungsergebnisse, die PISA bei Jugendlichen misst, repräsentieren eine fünfzehnjährige Lerngeschichte. Diese Bedingungen sollte man sich vor Augen führen, wenn man Erkenntnisse aus dem Vergleich zweier PISA-Erhebungen erwartet. Es gibt weitere methodische Einschränkungen für diesen Vergleich, die wir im folgenden Abschnitt behandeln.

## 1.5 Von PISA 2000 nach PISA 2003: Belastbare Aussagen über Veränderungen

Bei PISA 2000 und PISA 2003 wurden in den Ländern unterschiedliche Populationen (Kohorten) mit vergleichbaren Erhebungsverfahren getestet. Zu beiden Zeitpunkten werden die Populationen durch Stichproben repräsentiert. Diese Tatsache müssen wir ebenso berücksichtigen wie die Zuverlässigkeit von Erhebungsverfahren, wenn wir die Ergebnisse aus beiden PISA-Runden nebeneinander stellen. Letztlich gilt es, die substantiellen Unterschiede zwischen den Erhebungen – unter Kontrolle von eventuellen Stichprobenfehlern und Messungenauigkeiten – herauszuarbeiten. Gerade wenn man PISA auch als Feedback über erreichte Veränderungen nutzen möchte, müssen wir die gemessenen Ergebnisse kritisch und mit größter methodischer Sorgfalt prüfen.

Da PISA bisher nur einen Vergleich über zwei Messzeitpunkte gestattet, fehlt die empirische Grundlage, um von einer *Entwicklung* oder von einem *Trend* zu sprechen.

Beide Bezeichnungen unterstellen eine substantielle Veränderung und suggerieren eine Progression. Von einem Trend sollte erst dann gesprochen werden, wenn Veränderungen über drei Messzeitpunkte in die gleiche Richtung weisen. In diesem Band begnügen wir uns damit, von *Unterschieden* zwischen den Erhebungszeitpunkten beziehungsweise von *Veränderungen* zu sprechen, wenn die Unterschiede als substantiell erscheinen.

Um einzuschätzen, ob Veränderungen in Kennwerten substantielle Änderungen in den Bildungsergebnissen anzeigen, müssen wir uns mit Besonderheiten der Stichproben, der Testdurchführung und -beteiligung sowie der Zuverlässigkeit der Erhebungsverfahren befassen.

Unterschiede in den typischen Kennwerten (Mittelwert, Standardabweichung) über zwei Erhebungszeitpunkte können auf Unterschiede in der *Ausschöpfung der Population und der Stichprobe* zurückzuführen sein. Veränderungen der Anteile der Zielpopulation (Fünfzehnjährige in schulischer Ausbildung) an der Population (Fünfzehnjährige) oder die Ausschöpfung der Stichprobe (z.B. Testung von Jugendlichen in beruflichen Teilzeitschulen) können sich deutlich in den Ergebnissen niederschlagen. Ebenso könnten Unterschiede auftreten, wenn die Testbeteiligung zu einem Zeitpunkt systematisch verzerrt wäre. Für den Ländervergleich haben wir ausgeschlossen, dass 2003 systematisch leistungsschwächere Schülerinnen und Schüler von einer Testteilnahme absahen (s.o.). Um zu prüfen, ob die Unterschiede zwischen den Stichproben der Jahre 2000 und 2003 auch Unterschiede zwischen den Populationen repräsentieren, müssen wiederum (s.o.) Standardschätzfehler berücksichtigt beziehungsweise geeignete Signifikanzprüfungen durchgeführt werden.

Erhebliche Auswirkungen auf die Testergebnisse kann auch die Art und Weise der *Testadministration* haben. So bereiteten unter anderem Testheft- und Positionseffekte eine Reihe von Problemen, die Kompetenzkennwerte zu den zwei Zeitpunkten zu vergleichen (OECD, 2005). Entsprechend ist sorgfältig zu prüfen, ob die Testdurchführung oder auch die Qualität der Übersetzungen zu beiden Erhebungszeitpunkten in jeder Hinsicht identisch war.

Die Mittelwerte der in PISA 2000 erhobenen Kompetenz-Skalen wurden für den OECD-Durchschnitt auf einen Mittelwert von 500 und eine Standardabweichung von 100 festgelegt. Um die Ergebnisse von PISA 2003 auf die vorherige Studie beziehen zu können, muss eine gemeinsame Verankerung der Kompetenzdaten beider Studien gewählt werden. Für die *Lesekompetenz* und die *Naturwissenschaften* wurde als Bezugspunkt die Verankerung auf der PISA-Skala 2000 beibehalten.

Für *Mathematik* wurde PISA 2003 als Referenzpunkt gewählt, da hier ein vollständiger Mathematiktest eingesetzt wurde. Bezogen auf zwei Teilskalen wurden die Daten aus PISA 2000 auf diesen neuen Bezugspunkt umgerechnet. Die Ergebnisse aus PISA 2000 und PISA 2003 müssen anhand der neu normierten Kompetenzwerte aufeinander bezogen werden. Die *Problemlösekompetenz* wurde in PISA 2003 erstmals erfasst und entsprechend für PISA 2003 bei einem Mittelwert von 500 und einer Standardabweichung von 100 Punkten verankert.



## 1.6 PISA – Ein kooperatives Unternehmen

Eine aufwändige Studie wie PISA-E erfordert die Zusammenarbeit zahlreicher Organisationen, Institute und Personen. In Deutschland wurde der Auftrag für die Durchführung des PISA-2003-Ländervergleichs (nach Ausschreibung) durch die Ständige Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland (KMK) an ein nationales Konsortium vergeben. Das Konsortium stimmt die Arbeit mit der Amtschefscommission „Qualitätssicherung in Schulen“ ab (Vorsitz: Ministerialdirektor Josef Erhard, Bayern und Staatssekretär Elmar Schulz-Vanheyden, Nordrhein-Westfalen sowie Staatssekretär Wolfgang Meyer-Hesemann, Schleswig-Holstein) ab. Die Amtschefscommission bildet zusammen mit wissenschaftlichen Vertreterinnen und Vertretern (Rainer Bromme, Münster, Helmut Fend, Zürich, Kurt Heller, München, Andreas Helmke, Landau, Klaus Klemm, Essen, Friederike Klippel, München, Kristina Reiss, Augsburg, Kaspar Spinner, Augsburg, Elke Sumfleth, Essen) einen Beirat für Vergleichsstudien in Deutschland.

Die Federführung des PISA-Konsortiums Deutschland liegt am Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN) in Kiel (<http://www.pisa.ipn.uni-kiel.de>). Dem Konsortium gehören folgende Wissenschaftler an:

- Manfred Prenzel, Kiel (Sprecher)
- Jürgen Baumert, Berlin
- Werner Blum, Kassel
- Rainer Lehmann, Berlin
- Detlev Leutner, Essen
- Michael Neubrand, Oldenburg
- Reinhard Pekrun, München
- Jürgen Rost, Kiel
- Ulrich Schiefele, Bielefeld

Für die Projektkoordination für PISA-E 2003 am IPN ist Barbara Drechsel verantwortlich. Der PISA-Arbeitsgruppe am IPN gehören Désirée Burba, Claus H. Carstensen, Clemens Draxler, Barbara Drechsel, Timo Ehmke, Heike Heidemeier, Fanny Hohensee, Manfred Prenzel, Gesa Ramm, Jürgen Rost, Ute Schröder, Martin Senkbeil, Thilo Siegle, Antje Stick, Beate von der Heydt, Oliver Walter und Karin Zimmer sowie zahlreiche studentische Hilfskräfte an. Die Skalierung des Ländervergleichs und die Datenaufbereitung am IPN leitet Claus H. Carstensen.

Die Organisation der Datenerhebung sowie der Datenverarbeitung unternimmt das *IEA Data Processing Center* (DPC) in Hamburg. Zuständig waren Susan Böhmer, Regina Borchert, Falk Brese, Jens Gomolka, Steffen Knoll (Projektkoordination), Cornelia Kutter, Heiko Sibberns (Leitung) und Anja Waschk.

Die Verantwortung für die Organisation der PISA-Erhebung in den Ländern lag bei den Ländern selbst. In jedem Land wurde eine PISA-Beauftragte beziehungsweise ein

PISA-Beauftragter und ein PISA-Koordinator beziehungsweise eine PISA-Koordinatorin benannt. Die Planung und Durchführung der Untersuchung wurde in Zusammenarbeit mit dem IEA-DPC und dem PISA-Konsortium Deutschland durchgeführt. In den Ländern waren für PISA-E 2003 folgende Personen verantwortlich:

Klaus Happold, Franz Rapf (Baden-Württemberg)

Elfriede Ohrnberger, Fritz Bedall (Bayern)

Tom Stryck, Christian Ernst (Berlin)

Hans-Jürgen Kuhn (Brandenburg)

Dietmar Kirchhoff (Bremen)

Jan Poerschke, Peter May (Hamburg)

Gabriele Vogt, Ulrich Steffens (Hessen)

Heidrun Breyer, Heidemarie Thiele, Uwe Feiste (Mecklenburg-Vorpommern)

Vera Reineke, Ulrich Kirschner (Niedersachsen)

Gerhard Orth, Christoph Burkard (Nordrhein-Westfalen)

Winfried Eschmann (Rheinland-Pfalz)

Christian Heib, Wolfgang Pfaff, Herbert Sauder, Sylvia Behet (Saarland)

Barbara Meyer-Wyk, Dorit Stenke, Simone Heinrich (Sachsen)

Horst Koehn, Frank Kirchner (Sachsen-Anhalt)

Werner Klein, Kirk Fänderich, Christian Behrendt (Schleswig-Holstein)

Rupert Deppe, Johann Fackelmann, Roberto Napierski (Thüringen)

Die Arbeit des PISA-Konsortiums Deutschland wird durch Expertengruppen tatkräftig unterstützt. Die nationalen Expertengruppen setzen sich wie folgt zusammen.

### **Mathematik**

Werner Blum (Sprecher), Kassel

Michael Neubrand (Sprecher), Oldenburg

Regina Bruder, Darmstadt

Elmar Cohors-Fresenborg, Osnabrück

Lothar Flade, Magdeburg

Rudolf vom Hofe, Regensburg

Alexander Jordan, Kassel

Norbert Knoche, Essen

Detlef Lind, Wuppertal

Wolfgang Löding, Hamburg

Gerd Möller, Düsseldorf

Johanna Neubrand, Vechta

Alexander Wynands, Bonn

Frauke Ulfig, Oldenburg

### **Leseverständnis**

Ulrich Schiefele (Sprecher), Bielefeld

Cordula Artelt, Berlin

Jens Möller, Kiel

Wolfgang Schneider, Würzburg

Wolfgang Schnotz, Landau

Petra Stanat, Berlin

Lilian Streblov, Bielefeld

**Naturwissenschaften**

Jürgen Rost (Sprecher), Kiel  
 Horst Bayrhuber, Kiel  
 Wolfgang Bündler, Kiel  
 Claus H. Carstensen, Kiel  
 Reinders Duit, Kiel  
 Manfred Euler, Kiel  
 Hans E. Fischer, Essen  
 Alfred Flint, Rostock  
 Peter Häußler, Kiel  
 Marcus Hammann, Kiel

Rainer Klee, Gießen  
 Michael Komorek, Kiel  
 Armin Lude, Kassel  
 Jürgen Mayer, Gießen  
 Peter Nentwig, Kiel  
 Sabine Nick, Kiel  
 Helmut Prechtel, Kiel  
 Manfred Prenzel, Kiel  
 Martin Senkbeil, Kiel  
 Oliver Walter, Kiel

**Problemlösen**

Detlev Leutner (Sprecher), Essen  
 Cordula Artelt, Berlin  
 Joachim Funke, Heidelberg  
 Eckhard Klieme, Frankfurt/Main

Stephan Kröner, Erfurt  
 Petra Stanat, Berlin  
 Joachim Wirth, Essen

**Elternhaus/Schülervoraussetzungen**

Reinhard Pekrun (Sprecher), München  
 Anne Frenzel, München  
 Thomas Götz, München  
 Bettina Hannover, Berlin

Stephanie Lichtenfeld, München  
 Sabine Walper, München  
 Elke Wild, Bielefeld

**Schulkontext**

Rainer Lehmann (Sprecher), Berlin  
 Brigitte Steinert, Frankfurt/Main

Astrid Neumann, Berlin

**Bedingungsbereich Mathematik**

Jürgen Baumert (Sprecher), Berlin  
 Werner Blum, Kassel  
 Stefan Krauss, Berlin

Mareike Kunter, Berlin  
 Michael Neubrand, Oldenburg

Dass die PISA-Erhebung in Deutschland 2003 erfolgreich durchgeführt wurde, verdanken wir dem Engagement der Koordinatorinnen und Koordinatoren in den Ländern und an den einzelnen Schulen. Dieses Engagement drückt die Bereitschaft aus, sich Vergleichen zu stellen und aus Vergleichen zu lernen. Das PISA-Konsortium Deutschland möchte an dieser Stelle ebenfalls allen Schülerinnen und Schülern, Eltern, Lehrkräften und Schulleitungen für ihre Mitarbeit bei PISA 2003 herzlich danken.

## 1.7 Was ist neu beim Ländervergleich PISA 2003?

Eine besondere Stärke von PISA besteht darin, dass in regelmäßigen Abständen Erhebungen zu den gleichen Kompetenzen durchgeführt werden. Auf diese Weise können die Ergebnisse der Erhebungsrounden verglichen werden, um Veränderungen beziehungsweise auf längere Sicht Entwicklungstrends zu berichten. Vergleiche über Erhebungszyklen setzen aber voraus, dass die Tests und Fragebögen zu allen Zeitpunkten eine Menge gleicher Aufgaben und Fragen enthalten. Damit bleiben relativ kleine Spielräume, die Erhebungen zu einem Zeitpunkt neu auszurichten. Allerdings setzt PISA bei den Kompetenzerhebungen jeweils einen inhaltlichen Schwerpunkt und akzentuiert entsprechend die Fragebögen. Auf der nationalen Ebene ergeben sich zusätzliche Freiheitsgrade für vertiefende oder neue Fragestellungen, wenn Optionen der Stichprobenziehung oder der Ergänzung durch zusätzliche Instrumente an einem zweiten Testtag genutzt werden. Da Deutschland von diesen Möglichkeiten gezielt Gebrauch macht, können in diesem Bericht neue und vertiefende Erkenntnisse vorgestellt werden, die über die internationale Studie 2003 hinausgehen und sich deutlich auch von dem Ländervergleich in Deutschland im Jahr 2000 unterscheiden. Auf diese Besonderheiten weist die folgende Vorschau über die Kapitel hin.

Das *zweite Kapitel* in diesem Band behandelt das Schwerpunktgebiet von PISA 2003, die *mathematische Kompetenz*. Der Vergleich der Länder erfolgt hier nicht nur an einem Gesamtwert, sondern an vier Teilskalen, die mathematische Inhaltsgebiete (im Sinne übergreifender Ideen) aufschlüsseln und damit Hinweise auf relative Stärken und Schwächen der Länder geben. Das Kapitel schließt einen Vergleich der Ergebnisse ein, die in den Gymnasien der Länder erzielt wurden. Im Unterschied zu PISA 2000 können nun für zwei Teilgebiete der Mathematik die Veränderungen berichtet werden, die in den Ländern insgesamt und im Besonderen an den Gymnasien im Verlauf der letzten drei Jahre festzustellen sind.

Etwas knapper fällt das *dritte Kapitel* aus. Es widmet sich der *Lesekompetenz*, über die bei PISA 2000 sehr ausführlich berichtet wurde. Dieses Kapitel stellt die aktuellen Länderergebnisse vor und informiert darüber, ob seit PISA 2000 in einzelnen Ländern deutliche Verbesserungen der Lesekompetenz zu verzeichnen sind. Auch hier wird der Gymnasialvergleich über die Länder genutzt, um die Veränderungen der Lesekompetenz im Verlauf der letzten Jahre zurückzumelden.

Auch das *vierte Kapitel*, das sich mit der *naturwissenschaftlichen Kompetenz* befasst, folgt dieser Struktur und berichtet neben den aktuellen Ergebnissen die Veränderungen gegenüber dem Ländervergleich bei PISA 2000, auch für die Gymnasien.

Thematisch neu ist der Bericht über den Ländervergleich zum *Problemlösen* im *fünftten Kapitel*. Hier wird untersucht, inwieweit die sich abzeichnende Stärke der Schülerinnen und Schüler Deutschlands auch in den einzelnen Ländern wiederzufinden ist. Damit verbunden ist die Frage, ob es den Ländern im gleichen Maße gelingt, das in der Problemlösekompetenz erkennbare kognitive Potential in fachspezifische (z.B. mathe-

matische) Kompetenz umzusetzen. Diese Frage betrifft auch den Ländervergleich zur Problemlösekompetenz in den Gymnasien.

Im *sechsten Kapitel*, das sich mit *Lernstrategien und motivationalen Merkmalen* befasst, werden erstmals auch (mathematikbezogene) Emotionen berücksichtigt. Besonders aufschlussreich ist hier die Angst vor der Mathematik. Neben Länderunterschieden in der Selbstwirksamkeitswahrnehmung und im Einsatz von Lernstrategien sind in diesem Bericht die Geschlechterdifferenzen von besonderem Interesse.

Im *siebten Kapitel* zur *Vertrautheit mit dem Computer* gewinnt der Ländervergleich zusätzlich an Reiz durch die Rückbindung der Befunde zu einem kleinen ergänzenden Test der „Computer Literacy“, der an einer Teilstichprobe durchgeführt worden war.

Ein ganz neues Spektrum an Informationen über die *Schulen in den Ländern* liefert das *achte Kapitel*. Für jedes Land werden Kennzahlen über die Verteilung der Fünfzehnjährigen an den Schulen, besonders aber über die Häufigkeiten von verzögerten Schullaufbahnen berichtet. Damit steht die Zeitnutzung im Blickpunkt, die auch durch Angaben von Lernzeiten im Unterricht wie außerhalb ergänzt wird. Für jedes Land wird außerdem dargestellt, welches Niveau die einzelnen Schularten in den vier Kompetenzbereichen erzielen und wie sich die Leistungen innerhalb beziehungsweise über die Schularten verteilen.

Auch das *neunte Kapitel* zur *sozialen Herkunft der Schülerinnen und Schüler* enthält eine Anzahl wichtiger Neuerungen, ebenfalls gegenüber den nationalen Auswertungen von PISA 2000. Insbesondere wird hier die Kopplung zwischen sozialer Herkunft und mathematischer Kompetenz über die Länder anhand eines neu gebildeten Index bestimmt. Dieser von der OECD bei PISA 2003 erstmals verwendete Index erfasst materielle, soziale und kulturelle Ressourcen des Elternhauses und ist so angelegt, dass Kompetenzunterschiede möglichst gut vorhergesagt werden. Außerdem berichtet das Kapitel über einen neuen Kennwert zur Beschreibung der Kopplung (das Maß der aufgeklärten Varianz), der das Maß des aus PISA 2000 bekannten „sozialen Gradienten“ ergänzt und vermutlich ablösen wird.

Das Oversampling von *Jugendlichen mit Migrationshintergrund* gestattet differenzierte Analysen von Teilgruppen, über die das *zehnte Kapitel* informiert. Eine Differenzierung betrifft die Herkunftsländer, eine zweite die unterschiedlichen Typen der Aneignung der Sprache des Einwanderungslandes. Kombinierte Analysen solcher Typen der Akkulturation mit Merkmalen der sozialen Lage oder des Herkunftslandes tragen dazu bei, die Situation von Jugendlichen mit Migrationshintergrund im Bildungssystem besser einordnen zu können.

Das *elfte Kapitel* stellt die Ergebnisse eines neuen Zugangs zu den Informationen aus dem Schulfragebogen dar, in dem es *Typen von Schulen* identifiziert und in den Verteilungen nach Land und Schulart berichtet. Die Typologie greift auf Merkmale zurück, die Schulen als „belastet“ erscheinen lassen, und solche, die „aktive“ Ansätze zur Problemlösung und Weiterentwicklung der Schule erkennen lassen.

Einen weiteren umfassenden Analysezugang zu PISA-Daten stellt das *zwölfte Kapitel* über *regionale Disparitäten* vor. Diese Auswertungen beziehen sich auf kleinere regionale Einheiten (als Länder) und greifen auf entsprechende Regionaldaten (z.B. wirtschaftliche Kennzahlen) zurück, die als einflussreiche Rahmenbedingungen für Bildungsprozesse in Elternhaus und Schule gelten. Dabei wird die Frage untersucht, inwieweit entsprechende Regionaldaten Unterschiede in den Kompetenzen vorhersagen können und welcher Erklärungsbeitrag für Merkmale der Schulen bleibt.

Das *Kapitel 13* rundet den Band mit einem *Fazit des Ländervergleichs* ab. Dieses Kapitel befasst sich noch einmal kritisch mit der Vergleichbarkeit der Länderergebnisse und stellt dann kompetenzübergreifend die Ergebnisse und die bei PISA 2003 beobachteten Veränderungen zusammen, um diese abschließend zu diskutieren.

Für die Leserinnen und Leser, die mehr über die beim Ländervergleich im Rahmen von PISA 2003 verwendeten *Methoden* erfahren wollen, stellt das *Kapitel 14* Basisinformationen und Hinweise auf weiterführende Literatur bereit.

## 1.8 Ergebnisse des Ländervergleichs im Überblick

Im internationalen Bericht hatten sich bereits positive Veränderungen für die Stichprobe aus Deutschland abgezeichnet (vgl. Prenzel et al., 2004a). Der Vergleich der Länder mit einer erweiterten Stichprobe bestätigt diese Entwicklungen und hilft diese einzuordnen.

Insgesamt konnten sich bei PISA 2003 zahlreiche Länder *bezogen auf den OECD-Durchschnitt* sehr viel besser positionieren als bei PISA 2000. In der aktuellen Erhebung liegen drei Länder in allen Kompetenzbereichen signifikant *über* dem OECD-Mittelwert. Mindestens einem dieser Länder ist der Anschluss an die internationale Spitzengruppe in allen Kompetenzbereichen gelungen. Gegenüber 2000 befindet sich nun eine größere Zahl von Ländern *innerhalb des* internationalen Durchschnittsbereichs. Schwächen zeichnen sich jedoch noch bei der Hälfte der Länder in der Lesekompetenz ab. Bei PISA 2003 liegen nur mehr wenige Länder in mehreren Kompetenzbereichen *unter* dem internationalen Durchschnittswert.

Betrachtet man die bei PISA 2003 erreichten Ergebnisse unter einer *inhaltlichen, kriteriumsorientierten Bezugsnorm*, dann kann generell von einer Stabilisierung, in vielen Kompetenzbereichen und Ländern sogar von einer deutlichen Verbesserung des Kompetenzniveaus gesprochen werden. Fünf Länder können in allen Kompetenzbereichen signifikante Zuwächse verzeichnen, darunter sind beeindruckend große Punktgewinne gegenüber PISA 2000. Bemerkenswert sind die deutlichen Verbesserungen in einigen Ländern, die 2000 relativ schlecht abgeschnitten hatten. Beträchtliche Gewinne erzielen jedoch nicht nur vormals leistungsschwächere Länder, sondern auch zwei Länder, die 2000 im Verhältnis gut abgeschnitten hatten. Dieser Befund widerspricht der Annahme, die größten Zugewinne könnten nur in Ländern mit einem bislang niedrigen Leistungsniveau erreicht werden. Das Bild wird ergänzt durch zahlreiche weitere Länder, die

sich in drei, zwei oder einem Kompetenzbereich verbessern konnten. Insgesamt beschreiben diese Befunde positive Entwicklungen während der letzten drei Jahre.

Auch die Ergebnisse des Ländervergleichs bestätigen, dass die Schülerinnen und Schüler in Deutschland über ein *kognitives Potential* verfügen, das an sich höhere Leistungen in den fachlichen Kompetenzen erwarten lässt. Mit Ausnahme eines Landes liegen die Leistungen im Bereich Problemlösen über dem internationalen Durchschnitt. Wie die Analysen in den Länderkapiteln zeigen, werden die im Problemlösen erkennbaren kognitiven Fähigkeiten in einigen Ländern und Schularten angemessen umgesetzt, in anderen jedoch nicht. Bemerkenswert sind nicht nur die diskrepanten Leistungsniveaus, wenn man die analytische Problemlösekompetenz mit der „verwandten“ mathematischen und naturwissenschaftlichen Kompetenz vergleicht. Auch die Problemlösekompetenz und das Leseverständnis klaffen oft weiter auseinander als theoretisch zu vermuten ist.

Betrachtet man nicht nur das durchschnittlich erreichte Kompetenzniveau, sondern die *Verteilungen der Kompetenzen*, dann findet man für die drei inhaltlichen Domänen in fast allen Ländern eine hohe Leistungsstreuung. Problematisch wird diese Streuung in den Kompetenzwerten deshalb, weil sie in allen Ländern insbesondere von sehr großen Anteilen außerordentlich leistungsschwacher Schülerinnen und Schüler bestimmt wird. Die Länder unterscheiden sich zwar beträchtlich in den Anteilen von Jugendlichen, die auf oder unter der niedrigsten Kompetenzstufe eingeordnet werden. Die Größenordnungen reichen von ca. 12 bis 30 Prozent. Doch zeigt der Vergleich mit vielen OECD-Staaten, dass die Anteile von Jugendlichen mit Risikoprognosen für ihre weitere schulische und berufliche Zukunft noch sehr viel kleiner ausfallen können.

Sowohl beim internationalen wie beim nationalen Vergleich zeichnet sich ab, dass die Anteile von Jugendlichen auf beziehungsweise unter der ersten Kompetenzstufe in den Staaten beziehungsweise Ländern sehr klein ausfallen, die insgesamt sehr gut abschneiden. Bessere Ergebnisse im internationalen Vergleich können und müssen über Verbesserungen im unteren Leistungsbereich erzielt werden.

Der Ländervergleich bestätigt ebenfalls einen Befund des internationalen Vergleichs: Die *Kopplung der Kompetenz mit Merkmalen der sozialen Herkunft* ist meist in den Ländern geringer, die ein hohes Kompetenzniveau erzielen. Bei der PISA-Erweiterung können in Deutschland 2003 drei Länder identifiziert werden, die ein sehr hohes Kompetenzniveau bei einer relativ geringen Kopplung mit Merkmalen der sozialen und soziokulturellen Herkunft (ESCS) erreichen, eine Konstellation, die von der OECD als wünschenswert bezeichnet wird.

Betrachtet man die Befunde über die Kompetenzen, die in den einzelnen Ländern an den verschiedenen Schularten erzielt werden, dann sind die *Leistungsunterschiede* interessant, die *innerhalb eines Landes* gefunden werden. Bemerkenswert sind jedoch auch die sehr hohen Quoten von *verzögerten Schullaufbahnen*, die allerdings nennenswert zwischen den Ländern variieren. Die Zurückstellungen und Wiederholungen stehen in einem engen Zusammenhang mit den hohen Anteilen von Schülerinnen und Schülern im unteren Leistungsbereich. Offensichtlich wurden auch in der 2003 getesteten

Kohorte leistungsschwächere Schülerinnen und Schüler häufig in Wiederholungsschleifen geschickt beziehungsweise schrittweise an andere Schulformen weitergereicht. Diese Maßnahmen lassen keinen förderlichen Effekt auf die Kompetenzentwicklung erkennen, denn als Ergebnis finden wir große Anteile von Fünfzehnjährigen mit einer verzögerten Schullaufbahn *und* einem niedrigen Kompetenzniveau.

Die Analysen zu Merkmalen der *sozialen Herkunft und Migration* zeigen auch 2003 straffe Zusammenhänge mit dem Kompetenzniveau. Die Befunde lassen erkennen, dass die materiellen und kulturellen Ressourcen der Elternhäuser eine bedeutsame Rolle bei der Kompetenzentwicklung spielen. Dies bedeutet auch, dass Potentiale und Talente übersehen oder nicht ausreichend gefördert werden. Besondere Aufmerksamkeit fanden bei der nationalen Erweiterung von PISA Jugendliche mit Migrationshintergrund. Die Analysen zeigen, dass unter den Jugendlichen mit Migrationshintergrund relativ viele in ihren Leistungen auf den untersten Kompetenzstufen einzuordnen sind. Andererseits unterstreichen vergleichbare Anteile von Jugendlichen mit und ohne Migrationshintergrund im oberen Leistungsbereich, dass Migrationshintergrund keineswegs zwangsläufig mit einem niedrigen Kompetenzniveau verbunden ist. Die detaillierten Analysen von Typen der Akkulturation weisen auf die Bedeutung der frühzeitigen Aneignung der deutschen Sprache und des häufigen Gebrauchs dieser Sprache im Alltag für das Lernen und den Schulerfolg hin. Die Differenzierung von Herkunftsgruppen belegt, dass die Tendenz zur Aneignung und zum Gebrauch der deutschen Sprache je nach kultureller Herkunft durchaus verschieden ist und damit unterschiedliche Strategien zur Sprachförderung und Integration erfordert.

Die Analysen zu Merkmalen der sozialen Herkunft belegen auch, dass die Schulen in den Ländern der Bundesrepublik Deutschland mit unterschiedlichen Ausgangsbedingungen beziehungsweise unterschiedlichen Kompositionen der Schülerschaft umzugehen haben. Diese Unterschiede sind nicht zu unterschätzen, wie die statistischen Analysen zeigen. Allerdings führen *Adjustierungen nach Merkmalen der sozialen Herkunft und Migration* zu keinen größeren Veränderungen in den relativen Positionen der Länder zueinander. In einigen Kompetenzbereichen führt die Adjustierung zu einer Verringerung der Varianz zwischen den Ländern.

Die Kompetenzunterschiede, die wir bei den *Ländervergleichen* der Gesamtwerte sowie der Leistungen der Gymnasien finden, sind auch 2003 substantiell. Die maximalen Leistungsdifferenzen können in Abstände von bis zu einem Schuljahr umgerechnet werden. Allerdings haben sich diese Abstände zwischen 2000 und 2003 keineswegs vergrößert; sie sind gleich geblieben beziehungsweise haben sich in einem Bereich verringert.

Kompetenzunterschiede in der beobachteten Größenordnung sind unter dem Aspekt vergleichbarer Lebensverhältnisse und gerechter Chancen problematisch. Allerdings zeigen die Vergleiche von *regionalen Gliederungen* in anderen Staaten ebenfalls Differenzen in ähnlicher Größe. Diese Unterschiede können zu einem Teil auf Standortfaktoren zurückgeführt werden. Die Analysen zu regionalen Disparitäten belegen jedoch den großen Stellenwert von Unterricht, Schule und schulischen Rahmenbedingungen für



Kompetenzentwicklung. Somit dürften die Entwicklungen, die seit PISA 2000 zu verzeichnen sind, eher auf professionelle Maßnahmen zur Verbesserung von Unterricht und Schule zurückzuführen sein als auf Veränderungen in Rahmenbedingungen und Standortfaktoren.

Die Auswertungen der Schulfragebogen lassen (in Übereinstimmung mit den Lehrer- und Schülerfragebögen) erkennen, dass ein beträchtlicher Teil der Schulen in den Ländern sich aktiv und konsequent um eine Verbesserung der professionellen Zusammenarbeit und der Lernergebnisse bemüht, und dies auch unter zum Teil schwierigen, belastenden Rahmenbedingungen. Auf der anderen Seite weisen die Analysen von Schultypen darauf hin, dass ebenfalls ein erheblicher Teil von Schulen auch unter relativ günstigen Bedingungen noch wenig Anstrengungen unternimmt, die Qualität von Unterricht und Schule weiter zu entwickeln. Bemerkenswert sind hier die unterschiedlichen Verteilungen dieser Schultypen nach Ländern und Schularten.

Unsere Analysen belegen, dass kurzfristige *Trainingseffekte und selektive Ausschöpfungen* der Stichproben die in PISA 2003 gemessenen Testleistungen nicht beeinflusst haben. Es finden sich allerdings Hinweise, dass die Schülerinnen und Schüler die Tests 2003 nach der Bekanntheit von PISA in Deutschland etwas ernster genommen haben. Allerdings spricht die empirische Evidenz aufgrund der Auswertungen der Anstrengungsthermometer gegen die Annahme, die Zuwächse seien auf die größere *Testmotivation* zurückzuführen. Für die Kompetenzentwicklung, deren Ergebnisse bei PISA verglichen werden, ist nach wie vor in erster Linie die Qualität des Lehrens und Lernens im schulischen Unterricht ausschlaggebend.

## Literatur

- Adams, R. & Wu, M. (Hrsg.) (2002). *PISA 2000 Technical Report*. Paris: OECD.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.) (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.) (2002a). *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Artelt, C., Carstensen, C.H., Sibberns, H. & Stanat, P. (2002b). Untersuchungsgegenstand, Fragestellungen und technische Grundlagen der Studie. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich* (S. 11-38). Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.) (2003). *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.
- BFS / EDK (Hrsg.) (2005). *PISA 2003: Kompetenzen für die Zukunft – Zweiter nationaler Bericht*. Neuchâtel: BFS.

- Brunner, M., Artelt, C., Krauss, S. & Baumert, J. (in Druck). Coaching for the PISA test. *Learning and Instruction*.
- Carstensen, C.H., Knoll, S., Rost, J. & Prenzel, M. (2004). Technische Grundlagen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 371-387). Münster: Waxmann.
- Kienzl, A. (2005). *Überprüfung der Schulnotenverteilungen im Rahmen von PISA 2003 (Diplomarbeit)*. Universität Kiel: Psychologisches Institut.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H.J. (Hrsg.) (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. (BMBF Bildungsreform 1)*. Bonn: BMBF.
- Klieme, E., Baumert, J., Köller, O. & Bos, W. (2000). Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1 – Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit* (S. 85-133). Opladen: Leske + Budrich.
- Köller, O. (2003). Qualitätsmonitoring PISA-E 2003. Unveröffentlichtes Manuskript, Erlangen.
- Mislevy, R.J., Beaton, A.E., Kaplan, B. & Sheehan, K.M. (1992). Estimating population characteristics from sparse matrix samples of items. *Journal of Educational Measurement*, 29(2), 133-164.
- OECD (2001). *Knowledge and skills for life. First results from PISA 2000*. Paris: OECD.
- OECD (2003). *The PISA 2003 Assessment framework – Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.
- OECD (Hrsg.) (2005). *PISA 2003 Technical Report*. Paris: OECD.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J. & Schiefele, U. (2004a). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Prenzel, M., Drechsel, B., Carstensen, C.H. & Ramm, G. (2004b). PISA 2003 – eine Einführung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 13-46). Münster: Waxmann.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2., vollst. überarb. u. erw. Aufl.). Bern: Hans Huber.
- Ständige Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland (Hrsg.) (2005). *Schüler, Klassen und Absolventen der Schulen 1994 bis 2003* (Vol. 174). Bonn: KMK.
- Tenorth, H.-E. (2004). Bildungsstandards und Kerncurriculum – Systematischer Kontext, bildungstheoretische Probleme. *Zeitschrift für Pädagogik*, 50(5), 650-661.
- Turner, R. (2002). Proficiency scales construction. In R. Adams & M. Wu (Hrsg.), *PISA 2000: Technical Report*. (S. 195-216). Paris: OECD.
- Weinert, F.E. (1999). *Concepts of competence (Contribution within the OECD project. Definition and selection of competencies: Theoretical and conceptual foundations (DeSeCo))*. Neuchâtel: Definition and Selection of Competencies (DeSeCo).