

Schwippert, Knut [Hrsg.]; Kasper, Daniel [Hrsg.]; Köller, Olaf [Hrsg.]; McElvany, Nele [Hrsg.]; Selter, Christoph [Hrsg.]; Steffensky, Mirjam [Hrsg.]; Wendt, Heike [Hrsg.]

TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich

Münster ; New York : Waxmann 2020, 365 S.



Quellenangabe/ Reference:

Schwippert, Knut [Hrsg.]; Kasper, Daniel [Hrsg.]; Köller, Olaf [Hrsg.]; McElvany, Nele [Hrsg.]; Selter, Christoph [Hrsg.]; Steffensky, Mirjam [Hrsg.]; Wendt, Heike [Hrsg.]: TIMSS 2019. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Münster ; New York : Waxmann 2020, 365 S. - URN: urn:nbn:de:0111-pedocs-213539 - DOI: 10.25656/01:21353

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-213539>

<https://doi.org/10.25656/01:21353>

in Kooperation mit / in cooperation with:



WAXMANN
www.waxmann.com

<http://www.waxmann.com>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt unter folgenden Bedingungen vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen: Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen. Dieses Werk bzw. der Inhalt darf nicht für kommerzielle Zwecke verwendet werden. Die neu entstandenen Werke bzw. Inhalte dürfen nur unter Verwendung von Lizenzbedingungen weitergegeben werden, die mit denen dieses Lizenzvertrages identisch oder vergleichbar sind.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.en> - You may copy, distribute and transmit, adapt or exhibit the work in the public and alter, transform or change this work as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to make commercial use of the work. If you alter, transform, or change this work in any way, you may distribute the resulting work only under this or a comparable license.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung

E-Mail: pedocs@dipf.de

Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft



Knut Schwippert
Daniel Kasper
Olaf Köller
Nele McElvany
Christoph Selter
Mirjam Steffensky
Heike Wendt
(Hrsg.)

TIMSS 2019

**Mathematische und
naturwissenschaftliche
Kompetenzen von
Grundschulkindern
in Deutschland
im internationalen
Vergleich**

WAXMANN

Knut Schwippert, Daniel Kasper, Olaf Köller,
Nele McElvany, Christoph Selter,
Mirjam Steffensky, Heike Wendt (Hrsg.)

TIMSS 2019

Mathematische und naturwissenschaftliche
Kompetenzen von Grundschulkindern
in Deutschland im internationalen Vergleich



Waxmann 2020
Münster · New York

Bibliografische Informationen der Deutschen Nationalbibliothek

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

Print-ISBN 978-3-8309-4319-8

E-Book-ISBN 978-3-8309-9319-3

doi: 10.31244/9783830993193

© 2020, Waxmann Verlag GmbH,
Steinfurter Straße 555, 48159 Münster

www.waxmann.com
info@waxmann.com

Umschlaggestaltung: Christian Aeverbeck, Münster
Titelfoto: © Christian Schwier – [adobe.stock.com](https://www.adobe.com/stock)

Satz: Stoddart Satz- und Layoutservice, Münster

Dieses Werk ist unter der Lizenz *CC BY-NC-SA 4.0* veröffentlicht:
Namensnennung – Nicht-kommerziell – Weitergabe unter gleichen
Bedingungen 4.0 International (*CC BY-NC-SA 4.0*)



Inhalt

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

Kapitel 1

TIMSS 2019: Wichtige Ergebnisse im Überblick	13
---	-----------

Knut Schwippert, Daniel Kasper, Olaf Köller, Nele McElvany,
Christoph Selter, Mirjam Steffensky und Heike Wendt

1.1 Einleitung	13
1.2 Zentrale Ergebnisse	15
1.3 Bildungspolitische und didaktische Folgerungen	22
Literatur	24

Kapitel 2

Ziele, Anlage und Durchführung der <i>Trends in International Mathematics and Science Study</i> (TIMSS 2019)	25
---	-----------

Knut Schwippert, Luise A. Scholz, Christin Beese, Daniel Kasper, Katrin Schulz-Heidorf
und Anna-Lena Girelli

2.1 TIMSS – Schule und Unterricht im internationalen Vergleich	25
2.2 TIMSS – ein internationales und nationales Forschungsnetzwerk	26
2.3 Ziele der TIMSS-Untersuchung	28
2.4 Anlage von TIMSS	35
2.5 Vorbereitung und Durchführung von TIMSS 2019	46
2.6 Technische und empirische Voraussetzungen für die Auswertung von TIMSS-Daten	51
2.7 Hinweise für die Ergebnisdarstellungen	52
Literatur	53

Kapitel 3

Mathematische Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse	57
--	-----------

Christoph Selter, Daniel Walter, Aiso Heinze, Johanna Brandt und Armin Jentsch

3.1 Einleitung	57
3.2 Mathematische Grundbildung in Deutschland	59
3.3 TIMSS-Rahmenkonzeption zur Erfassung mathematischer Kompetenzen	64
3.4 Der TIMSS-Test zur Erfassung der mathematischen Kompetenzen	70
3.5 Ergebnisse	80
3.6 Zusammenfassung	106
Literatur	112

Kapitel 4

Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse 115

Mirjam Steffensky, Luise A. Scholz, Daniel Kasper und Olaf Köller

4.1	Einleitung	115
4.2	Naturwissenschaftliche Grundbildung in Deutschland	117
4.3	TIMSS-Rahmenkonzeption zur Erfassung der naturwissenschaftlichen Kompetenzen	121
4.4	TIMSS-Test	127
4.5	Ergebnisse zu den naturwissenschaftlichen Leistungen in Deutschland am Ende der vierten Jahrgangsstufe	137
4.6	Zusammenfassung	162
	Literatur	165

Kapitel 5

Leistungsveränderungen in TIMSS zwischen 2015 und 2019: Die Rolle des Testmediums und des methodischen Vorgehens bei der Trendschätzung 169

Alexander Robitzsch, Oliver Lüdtke, Knut Schwippert, Frank Goldhammer, Ulf Kroehne und Olaf Köller

5.1	Einleitung	169
5.2	Trendschätzungen in internationalen Schulleistungstudien	170
5.3	Methodisches Vorgehen	175
5.4	Ergebnisse	178
5.5	Zusammenfassung	181
	Literatur	183

Kapitel 6

Einblicke in die Gestaltung des Mathematik- und naturwissenschaftsbezogenen Sachunterrichts an Grundschulen in Deutschland 187

Justine Stang, Chantal Lepper, Mirjam Steffensky und Nele McElvany

6.1	Einleitung	187
6.2	Basisdimensionen der Unterrichtsqualität	188
6.3	Aktueller Stand TIMSS 2019 – internationaler und nationaler Vergleich	189
6.4	Einschätzung weiterer Gestaltungsaspekte des Unterrichts	195
6.5	Zusammenfassung	202
	Literatur	205

Kapitel 7

Privater Nachhilfeunterricht und Lehrkräftefortbildung am Ende der Grundschulzeit209

Karin Guill und Heike Wendt

7.1	Einleitung.....	209
7.2	Verbreitung von Nachhilfeunterricht.....	209
7.3	Nutzung von privatem Nachhilfeunterricht in verschiedenen Schülerinnen- und Schülergruppen.....	214
7.4	Wissen der Lehrkräfte um den Besuch von privatem Nachhilfeunterricht.....	216
7.5	Teilnahme der Lehrkräfte an Fortbildungen.....	217
7.6	Zusammenfassung.....	220
	Literatur.....	221

Kapitel 8

Geschlechterunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen223

Sonja Nonte, Ricarda Steinmayr und Luise A. Scholz

8.1	Einleitung.....	223
8.2	Geschlechterunterschiede in schulischen Leistungen.....	224
8.3	Erklärungsansätze zu Geschlechterunterschieden in schulischen Leistungen.....	227
8.4	Ergebnisse zu geschlechterspezifischen Leistungsunterschieden in Mathematik.....	230
8.5	Ergebnisse zu geschlechterspezifischen Leistungsunterschieden in den Naturwissenschaften.....	241
8.6	Ergebnisse zu geschlechterspezifischen Unterschieden in Einstellungen und Selbstkonzepten.....	251
8.7	Zusammenfassung.....	256
	Literatur.....	259

Kapitel 9

Soziale Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen von Viertklässlerinnen und Viertklässlern263

Tobias C. Stubbe, Maria Krieg, Christin Beese und Donieta Jusufi

9.1	Einleitung.....	263
9.2	Soziale Herkunft und Bildungserfolg.....	264
9.3	Indikatoren der sozialen Herkunft.....	266
9.4	Soziale Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen im internationalen Vergleich.....	269
9.5	Soziale Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen in Deutschland.....	279
9.6	Zusammenfassung.....	284
	Literatur.....	285

Kapitel 10

Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund291

Heike Wendt, Knut Schwippert, Tobias C. Stubbe und Donieta Jusufi

10.1	Einleitung.....	291
10.2	Migration und Schulleistungen.....	292
10.3	Indikatoren des Migrationshintergrundes.....	295
10.4	Mathematische und naturwissenschaftliche Kompetenzen der Schülerinnen und Schüler nach Migrationshintergrund in den Familien im internationalen Vergleich	296
10.5	Lebensbedingungen von Schülerinnen und Schülern mit und ohne Migrationshintergrund in Deutschland.....	300
10.6	Kompetenzunterschiede von Schülerinnen und Schülern in Deutschland im Trend.....	301
10.7	Leistungen auf unterschiedlichen Kompetenzstufen im Trend.....	304
10.8	Zusammenhänge des Migrationshintergrundes mit Sprache und Merkmalen der sozialen Herkunft	306
10.9	Fachbezogene Selbstkonzepte, Motivation und soziale Eingebundenheit.....	308
10.10	Zusammenfassung	309
	Literatur.....	310

Kapitel 11

Schullaufbahnpräferenzen am Übergang in die Sekundarstufe und der Zusammenhang mit leistungsrelevanten und sozialen Merkmalen315

Tobias C. Stubbe, Daniel Kasper und Armin Jentsch

11.1	Einleitung.....	315
11.2	Schullaufbahnpräferenzen von Lehrkräften und Erziehungsberechtigten: Verteilung und Ausmaß der Übereinstimmung.....	318
11.3	Schullaufbahnpräferenzen von Lehrkräften: Leistungsrelevante Merkmale der Schülerinnen und Schüler.....	320
11.4	Schullaufbahnpräferenzen von Lehrkräften und Erziehungsberechtigten: Soziale Merkmale der Schülerinnen und Schüler.....	323
11.5	Zusammenfassung	327
	Literatur.....	328

Kapitel 12

Trends in Schülerzusammensetzungen und mathematischen und naturwissenschaftlichen Kompetenzen331

Daniel Kasper, Heike Wendt, Knut Schwippert und Olaf Köller

12.1	Einleitung.....	331
12.2	Charakteristika der Viertklässlerinnen und Viertklässler in TIMSS 2007, 2011, 2015 und 2019.....	333
12.3	Trends in den Leistungen der Schülerinnen und Schüler	334
12.4	Veränderungen in den Schülerzusammensetzungen.....	342
12.5	Zusammenfassung	346
	Literatur.....	348

Anhang	351
Abbildungsverzeichnis	358
Tabellenverzeichnis.....	363

Vorwort

Im Rahmen der *Trends in Mathematics and Science Study* (TIMSS) von 2019 wurden wir Autorinnen und Autoren bei vorbereitenden Informationsveranstaltungen an verschiedenen Schulen immer wieder gefragt, warum wir denn jetzt noch eine Grundschulstudie zur Erfassung der Kompetenzen von Schülerinnen und Schülern durchführen wollen – es gäbe doch schon so viele andere Untersuchungen.

Ja, es gibt eine ganze Reihe von Untersuchungen – auch ein paar, die in Deutschland ähnlich groß angelegt sind wie TIMSS, aber dennoch ist TIMSS etwas Besonderes. Jede Studie ist jeweils für die Untersuchung bestimmter Fragestellungen optimiert, auf die andere Studien eben keine oder nur ausschnitthafte Antworten geben können. Das Besondere an TIMSS ist, dass hier Schülerinnen und Schüler, ihre Erziehungsberechtigten, Lehrkräfte und schließlich auch die Schulleitungen befragt werden, sodass ein Bild über die Umstände des Lehrens und Lernens gezeichnet werden kann. Insbesondere beschreibt TIMSS, welche Kompetenzen die Kinder in Mathematik und den Naturwissenschaften am Ende der Grundschulzeit in Begleitung ihrer Lehrkräfte und Erziehungsberechtigten erreichen. Dabei sind die Kinder, die Klassen, in denen sie unterrichtet werden, und die Schule, in die sie gehen, so ausgewählt worden, dass sie ein repräsentatives Bild für Deutschland vermitteln. TIMSS ist in diesem Sinne eine Bestandsaufnahme eines Bereiches des deutschen Bildungssystems.

Beurteilen zu können, ob die gesetzten Bildungsziele gut oder weniger gut erreicht wurden, hängt von Erwartungen und Vergleichsmöglichkeiten ab – und hier bietet TIMSS weitergehende Informationen, die in dieser Form keine andere Untersuchung für das Lernen mathematischer und naturwissenschaftlicher Kompetenzen in der Grundschule bietet: Den internationalen Vergleich, der einen Blick über den nationalen ‚Bildungs-Tellerrand‘ erlaubt und der für eine Auseinandersetzung mit der Frage, wo das deutsche Bildungssystem steht, hilfreich ist. So kann aufgrund der internationalen Vergleiche festgestellt werden, wo das deutsche Bildungssystem Stärken hat beziehungsweise was im Vergleich mit anderen Staaten gut gelingt. Darüber hinaus erlaubt der internationale Vergleich aber auch einen Blick darauf, was im deutschen Bildungssystem noch besser gelingen könnte. Dieser Vergleich kann auch neue oder andere Ziele im Bildungssystem vor Augen führen. Dies sind somit nicht nur Fragen einer Orientierung an Spitzenleistungen, es sind auch Fragen der Gerechtigkeit, der Eröffnung von Bildungschancen oder auch der Möglichkeit, Vorurteilen zu begegnen. Dabei werden eigene Einstellungen, Erfahrungen und auch Erwartungen der Befragten erhoben.

Über diese multiperspektivischen Betrachtungen von Zielen und Rahmenbedingungen des Bildungssystems hinaus bietet TIMSS auch die Möglichkeit, das Bildungssystem im zeitlichen Verlauf zu betrachten. Durch die regelmäßige Wiederholung der Untersuchung alle vier Jahre wird TIMSS zum Gradmesser von Veränderungen im und durch das Bildungssystem, die Studien verborgen bleiben, die vertiefte Momentaufnahmen von Lehr-Lernsituationen in den Blick nehmen oder aber die Schülerinnen und Schüler über eine gewisse Zeit begleitend befragen und testen.

Neben diesen besonderen Perspektiven, welche die TIMS-Studie einerseits im internationalen Vergleich und andererseits im Trend untersucht, stellt gerade die TIMSS-Untersuchung von 2019 eine zusätzliche Besonderheit dar: Es ist die letzte groß angelegte internationale Systemmonitoringstudie, die noch nicht unter dem Einfluss der Covid-19-Pandemie stand. Vor diesem Hintergrund sind die hier vorgestellten Analysen, Interpretationen und Empfehlungen auf Basis einer Bestandsaufnahme vor der Covid-19-Pandemie zu verstehen. Folgen durch Covid-19 sind und werden sicherlich noch umfassend Gegenstand aktueller und auch zukünftiger Studien sein. Insofern macht die Bestandsaufnahme durch TIMSS 2019 schon jetzt neugierig auf die Ergebnisse des nächsten TIMSS-Zyklus in 2023, da in dieser Studie dann die Grundschülerinnen und Grundschüler am Ende der vierten Jahrgangsstufe in den Blick genommen werden, die 2019 eingeschult wurden und somit einen Teil ihrer Schulzeit unter dem Eindruck der Pandemie er- und durchlebt haben. Es bleibt zu hoffen, dass dann im Trendvergleich festgestellt werden darf, dass die Pandemie nur eine Episode war, die keine bleibenden negativen Folgen im Bildungssystem hinterlassen hat. Es ist schon heute zu beobachten, dass alle Akteurinnen und Akteure im Bildungssystem, die sich um das Wohl und die positive Entwicklung unserer zukünftigen Generationen bemühen, das Beste dafür geben.

Kapitel 1

TIMSS 2019: Wichtige Ergebnisse im Überblick

Knut Schwippert, Daniel Kasper, Olaf Köller, Nele McElvany, Christoph Selter, Mirjam Steffensky und Heike Wendt

1.1 Einleitung

Wissenschaftliche Untersuchungen wie die *International Trends in Mathematics and Science Study* (TIMSS) 2019 verfolgen das Ziel, Rahmenbedingungen, Voraussetzungen und Ergebnisse des Lehrens und Lernens zu untersuchen.

Die TIMS-Studie legt den Fokus darauf, langfristige Entwicklungen zu mathematischen und naturwissenschaftlichen Kompetenzen der Viertklässlerinnen und Viertklässler im internationalen Vergleich zu dokumentieren und zu interpretieren. Durch die gleichzeitige Betrachtung von mathematischen und naturwissenschaftlichen Kompetenzen, curricularen Vorgaben sowie schulischen und familiären Rahmenbedingungen wird es möglich, Wissen über die Zusammenhänge von Einflussfaktoren und Lernerfolgen der Grundschülerinnen und Grundschüler zu untersuchen und Entscheidungshilfen für bildungspolitische Fragen zur Weiterentwicklung der Bildungssysteme zu geben. Deutschland beteiligt sich seit 2007 an der alle vier Jahre stattfindenden Grundschuluntersuchung und kann daher mittlerweile die Entwicklungen im deutschen Bildungssystem über einen Zeitraum von zwölf Jahren abbilden.

Eine Untersuchung wie TIMSS lässt sich nur durch den Zusammenschluss von Expertinnen und Experten unterschiedlicher internationaler und nationaler Fachrichtungen realisieren. Initiator ist die *International Association for the Evaluation of Educational Achievement* (IEA). Die IEA hat mehr als 60 Jahre Erfahrung in der Konzeption, der Durchführung und der Auswertung von internationalen Schulleistungsstudien. Untersuchungen der IEA zur Erfassung mathematischer und naturwissenschaftlicher Kompetenzen von Schülerinnen und Schülern gibt es seit den 1960er-Jahren. Mit den ersten internationalen Mathematik- beziehungsweise Naturwissenschaftsstudien der IEA wurden Bestandsaufnahmen der Kompetenzen in diesen Fächern initiiert. Die erste TIMSS-Untersuchung im Jahr

1995 bildete dann den Grundstein für längerfristige Trendanalysen bezüglich mathematisch-naturwissenschaftlicher Kompetenzen.

Die IEA hat das internationale Management der Studie dem *TIMSS & PIRLS International Study Center (ISC)* am Boston College anvertraut. Das ISC ist verantwortlich für das Design und die Implementierung der Studie, für die internationale Koordination der Entwicklung der Test- und Befragungsinstrumente und der Erhebungsprozeduren sowie für die Überprüfung der Qualität der Datenerhebung. Darüber hinaus wird am ISC die Skalierung der Daten aus den Tests durchgeführt und der internationale Ergebnisbericht verfasst. Auf nationaler Ebene wurde die TIMSS-Untersuchung durch das interdisziplinär besetzte TIMSS-2019-Konsortium inhaltlich verantwortet.

Zur Untersuchung der mathematischen und naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler wurde in Anlehnung an die Curricula der beteiligten Staaten und Regionen eine Rahmenkonzeption entwickelt. In Mathematik korrespondiert die Konzipierung der standardisierten Leistungstests somit weitgehend mit den 2004 von der *Kultusministerkonferenz (KMK)* verabschiedeten Bildungsstandards. Diese legen fest, welche mathematischen Kompetenzen Schülerinnen und Schüler am Ende ihrer Grundschulzeit erworben haben sollten. Für die naturwissenschaftlichen Kompetenzen, die in Deutschland im Rahmen des Sachunterrichts vermittelt werden, gibt es bislang keine Bildungsstandards. Jedoch werden im Perspektivrahmen der *Gesellschaft für Didaktik des Sachunterrichts (GDSU)* multikriteriale Ziele des (naturwissenschaftlichen) Unterrichts beschrieben. Dieser Perspektivrahmen hat zwar keinen bindenden Charakter, stellt aber einen wichtigen Orientierungsrahmen für die Weiterentwicklung von Bildungsplänen und didaktischen Materialien dar.

Sowohl für den Mathematik- als auch für den Sachunterricht werden in TIMSS drei Inhalts- und Anforderungsbereiche formuliert. Die Inhalte für Mathematik umfassen die Bereiche *Arithmetik*, *Messen und Geometrie* sowie *Daten*, die Inhalte für den Sachunterricht *Biologie*, *Physik/Chemie* und *Geografie*. In den Inhaltsbereichen werden drei kognitive Anforderungsbereiche unterschieden, die die für die Aufgabenbearbeitung notwendigen Denkprozesse der Schülerinnen und Schüler beschreiben. Dabei handelt es sich um:

- *Reproduzieren* von Wissen, Fertigkeiten und Grundvorstellungen,
- *Anwenden* von Wissen, Fertigkeiten und Grundvorstellungen beim Bearbeiten von Standardaufgaben sowie
- *Problemlösen* von komplexeren, anspruchsvolleren Aufgabenstellungen.

Im Rahmen von TIMSS 2019 kann in Mathematik eine zufriedenstellende Passung der TIMSS-Rahmenkonzeption 2019 mit den curricularen Vorgaben durch die KMK-Bildungsstandards von 2004 festgestellt werden – wie es auch bei den bisherigen TIMSS-Zyklen der Fall war. Für den Sachunterricht zeigt sich eine große Überlappung zwischen den Inhalten und dazugehörigen Themenbereichen aus TIMSS 2019 mit den Empfehlungen beziehungsweise Vorgaben aus dem Perspektivrahmen der GDSU beziehungsweise den Bildungsplänen.

Neben der Auseinandersetzung mit den Inhaltsbereichen und Anforderungsniveaus der Aufgaben spielt inzwischen auch das Testformat eine wichtige Rolle. Dabei kommt der Sicherstellung der Trendinformationen eine besondere Bedeutung zu, wenn sich die untersuchten Gegebenheiten in den Bildungssystemen im Laufe des langen Beobachtungszeitraums verändern und dies bei der Gestaltung der Testinstrumente berücksichtigt werden muss, um eine zeitgemäße Erfassung der im Fokus liegenden Kompetenzen sicherzustellen. In

TIMSS ist dieser Punkt in 2019 erreicht. Digitale Medien haben in den letzten Jahren zunehmend Einzug in die Lebenswelten der Schülerinnen und Schüler in vielen Staaten gehalten und spielen eine selbstverständliche Rolle im schulischen und privaten Leben. Während immer mehr Staaten digitale Medien auch im Unterricht einsetzen, gibt es in anderen Staaten nur eingeschränkt Möglichkeiten, diese überhaupt zu nutzen. Um diesen unterschiedlichen Entwicklungsständen Rechnung zu tragen, hat die internationale Studienleitung beschlossen, für die Staaten, bei denen die Digitalisierung in der Schule und im Unterricht vorangeschritten ist, neben den bewährten papierbasierten Tests auch digitale Testformate anzubieten. In 2019 entschieden sich neben Deutschland insgesamt 29 weitere Staaten, dieses neue Testformat einzusetzen.

Mit diesem Band werden die ersten nationalen und internationalen Befunde der TIMSS-2019-Untersuchung für Deutschland vorgelegt. Im Folgenden werden die Ergebnisse kapitelweise zusammengefasst.

1.2 Zentrale Ergebnisse

Der vorliegende Berichtsband bietet eine nationale Vertiefung und Ausdifferenzierung der internationalen Berichtslegung vom Boston College (Kelly, Centurino, Martin, & Mullis, 2020; Mullis & Martin, 2017; Mullis, Martin, Foy, Kelly, & Fishbein, 2020). Wesentliche internationale Befunde finden sich in beiden Berichten, jedoch werden in dem hier vorgelegten nationalen Band Ergebnisse mit besonderer Bedeutung und mit wichtigen Perspektiven für das deutsche Bildungssystem herausgearbeitet und um vertiefende Analysen ergänzt. Dabei werden nationale Hintergrundinformationen berücksichtigt, die den Rahmen einer internationalen Berichtslegung übersteigen. Zudem werden in den einzelnen Kapiteln die Befunde in den wissenschaftlichen Diskurs eingeordnet.

1.2.1 Mathematische Kompetenzen im internationalen Vergleich

Mit TIMSS 2019 liegen Befunde aus den vergangenen zwölf Jahren vor, die es ermöglichen, mathematische Kompetenzen von Viertklässlerinnen und Viertklässlern aus Deutschland im internationalen Vergleich und im Trend in den Blick zu nehmen. Die Analyse der erfassten Daten erlaubt es, auf Systemebene differenzierte Aussagen über spezifische Stärken und Schwächen der Schülerinnen und Schüler in den Inhaltsbereichen und den kognitiven Anforderungsbereichen des Mathematikunterrichts am Ende ihrer Grundschulzeit zu machen. Die Analyse der mathematischen Kompetenzen sowie der Einstellungen und der Selbstkonzepte von Schülerinnen und Schülern steht im Mittelpunkt des Kapitels von Selter, Walter, Heinze, Brandt und Jentsch.

Deutschland liegt im Bereich der mathematischen Kompetenzen über dem internationalen Mittelwert. Beschränkt man die Vergleiche Deutschlands auf die teilnehmenden EU- und OECD-Staaten, so zeigen sich für die Schülerinnen und Schüler aus Deutschland leicht unterdurchschnittliche Leistungen. Die Leistungsstreuung der Kompetenzen fällt im internationalen Vergleich eher gering aus. Gleichwohl ist der Anteil der Viertklässlerinnen und Viertklässler, die lediglich rudimentäre oder niedrige mathematische Kompetenzen entwickeln konnten (Kompetenzstufe I und II), mit etwa einem Viertel substantiell. Hinsichtlich des relativen Anteils auf der höchsten Kompetenzstufe V können nur

marginale Veränderungen im Vergleich der vier TIMSS-Zyklen festgestellt werden. Nach wie vor ist der Anteil an Schülerinnen und Schülern in Deutschland, die der höchsten Kompetenzstufe V zugeordnet werden können, mit maximal 6 Prozent im internationalen Vergleich gering.

Dabei zeigen die Viertklässlerinnen und Viertklässler relative Stärke im Bereich *Messen und Geometrie*. Für die beiden Bereiche *Arithmetik* und *Daten* zeigen sich hingegen aktuell relative Schwächen. Es wird sichtbar, dass die Anteile von Kindern mit positiven Einstellungen zur Mathematik beziehungsweise mit hohem mathematischem Selbstkonzept geringer geworden sind, auch wenn diese sich nach wie vor auf einem hohen Niveau befinden.

1.2.2 Naturwissenschaftliche Kompetenzen im internationalen Vergleich

In Deutschlands Grundschulen sind die Naturwissenschaften keine eigenständigen Fächer – sie gehören zu dem mehrperspektivischen Fach Sachunterricht. Während in den 1980/90er-Jahren Naturwissenschaften und insbesondere Themen der Chemie und Physik kaum eine Rolle im Sachunterricht spielten, deuten die aktuellen Befunde sowie die Ergebnisse aus TIMSS 2015 darauf hin, dass die Naturwissenschaften im Sachunterricht wieder stärker in den Vordergrund getreten sind. Der Untersuchung der Kompetenzen im Allgemeinen, aber auch unter besonderer Berücksichtigung der naturwissenschaftlichen Inhaltsbereiche und der kognitiven Anforderungsbereiche, widmen sich in Kapitel 4 Steffensky, Scholz, Kasper und Köller. In dem Kapitel werden neben der Betrachtung der Kompetenzen der Grundschülerinnen und Grundschüler auch deren Einstellung und Selbstkonzept in Bezug auf den Sachunterricht analysiert.

Der Mittelwert der naturwissenschaftlichen Kompetenzen der Grundschülerinnen und Grundschüler liegt in Deutschland 10 Punkte unter dem Mittelwert der vorherigen Studienzyklen, der seit 2007 stabil bei 528 Punkten lag. Neben dem niedrigeren Kompetenzstand ist eine Zunahme der Streuung bei den naturwissenschaftlichen Kompetenzen zu beobachten. Etwas genauer betrachtet, zeigt sich die Zunahme der Streuung durch ein Absinken am unteren Ende des Leistungsspektrums. Hieraus ergibt sich, dass der Anteil der Schülerinnen und Schüler, die lediglich die unteren beiden Kompetenzstufen erreichen, in TIMSS 2019 leicht gestiegen ist. Bei den höheren Kompetenzstufen zeigen sich nur leichte Verschiebungen, sodass die Verteilung hier im Trend als relativ stabil bezeichnet werden kann. Das im Vergleich zu früheren TIMSS-Erhebungen festgestellte schwächere Abschneiden der Schülerinnen und Schüler in 2019 kann nicht an einem bestimmten Inhaltsbereich festgemacht werden. Vielmehr zeigt sich, dass sich die Abnahme der Kompetenzstände über alle drei Inhaltsbereiche verteilt.

Erfreulich erscheint der Befund, dass rund drei Viertel der Schülerinnen und Schüler sehr positive Einstellungen zum Sachunterricht berichten und darüber hinaus auch absolut betrachtet ein positives sachunterrichtsbezogenes Selbstkonzept aufweisen. Einschränkend ist jedoch im Vergleich zu den vorherigen Studienzyklen 2007 bis 2015 festzustellen, dass diese Werte geringfügig aber dennoch statistisch signifikant gesunken sind.

1.2.3 Leistungsveränderungen in TIMSS zwischen 2015 und 2019: Die Rolle des Testmediums

Im fünften Kapitel untersuchen Robitzsch, Lüdtke, Schwippert, Goldhammer, Kroehne und Köller den sogenannten Moduseffekt, der mit der Umstellung von einer papierbasierten auf eine computerbasierte Testung in TIMSS 2019 einhergeht. Der Wechsel des Testmediums ist für Trendschätzungen insbesondere in internationalen Schulleistungsstudien empirisch herausfordernd. In einer Vielzahl von Studien wurde gezeigt, dass sich Testitems in ihren Eigenschaften (Schwierigkeit und Trennschärfe) verändern können, wenn das Testmedium gewechselt wird. Diese Beobachtung wird als Moduseffekt beschrieben. So konnte in Deutschland in früheren *Programme for International Student Assessment* (PISA)-Erhebungen gezeigt werden, dass Aufgaben, die am Computer gelöst wurden, im Mittel schwieriger waren als bei der Bearbeitung im gedruckten Testheft. In TIMSS 2019 hat die internationale Studienleitung ein Testdesign umgesetzt, das für die internationalen Datensätze, aber auch für den nationalen Datensatz Analysen erlaubt, um Moduseffekte zu schätzen und damit auch zu untersuchen, in welchem Ausmaß für die Trendschätzungen in Mathematik und den Naturwissenschaften der Effekt des Testmediums berücksichtigt werden muss. In diesem Kapitel wird der Frage nach potenziellen Moduseffekten mit vertiefenden Analysen nachgegangen. Jenseits der Aufdeckung dieser Effekte wird auch untersucht, ob die Art der gewählten Trendschätzung basierend auf den internationalen Datensätzen oder spezifische Trendschätzungen für Deutschland basierend auf den nationalen Datensätzen (originaler vs. marginaler Trend) Effekte auf die Interpretation der Veränderungen der Leistungen in Deutschland hat. Da die zur Untersuchung der Moduseffekte eingesetzten Verfahren empirisch sehr voraussetzungsreich sind, richtet sich dieses Kapitel vornehmlich an Lesende mit Vorkenntnissen im Bereich der *Item-Response-Theory*.

Ebenso wie für die Fünfzehnjährigen in PISA 2015 und PISA 2018 ergab sich für die Viertklässlerinnen und Viertklässler in TIMSS 2019, dass die mit dem Computer administrierten Aufgaben im Mittel in beiden getesteten Domänen schwieriger sind. Die im internationalen Bericht vorgestellten originalen (also international durchgeführten) Trendschätzungen für Deutschland, für die eine Adjustierung der mittels digitaler Medien eingesetzten Items bereits vorgenommen wurde, können durch die Analysen der nationalen Daten weitgehend gestützt werden.

Die international berichteten Befunde weisen in Deutschland auf stabile Leistungen in Mathematik zwischen 2015 und 2019 hin, in den Naturwissenschaften sinken die Leistungen leicht ab. Die deutschlandspezifischen Trendschätzungen stützen diese internationale Befundlage, wenngleich sie das Absinken der Leistungen in den Naturwissenschaften etwas geringer erscheinen lassen. Die Abweichungen dürften vermutlich vor allem dadurch zustande kommen, dass der nationale Moduseffekt in Deutschland in den Naturwissenschaften stärker ausfällt als der internationale. Zusammengefasst zeigt sich, dass sich die Leistungen zwischen 2015 und 2019 nur geringfügig verändert haben.

1.2.4 Einblicke in die Gestaltung des Mathematik- und naturwissenschaftsbezogenen Sachunterrichts an Grundschulen in Deutschland

Unter dem Begriff *Unterrichtsqualität* werden verschiedene Unterrichtsmerkmale zusammengefasst, die in einem positiven Zusammenhang mit Unterrichts- und Erziehungszielen stehen. Im deutschsprachigen Raum hat sich eine dreigliedrige Konzeptualisierung etabliert, welche zentrale Tiefenstrukturmerkmale in die sogenannten Basisdimensionen Klassenführung, kognitive Aktivierung und konstruktive Unterstützung differenziert. Diese übergeordneten Tiefenstrukturmerkmale sowie die Unterrichtsmerkmale erlebte Motivierungsqualität, fachbezogene Gestaltungsmerkmale des Unterrichts, adaptives Unterrichten und Differenzierungsmöglichkeiten für leistungsstärkere und leistungsschwächere Lernende stehen im Fokus des Kapitels von Stang, Lepper, Steffensky und McElvany.

Es zeigt sich, dass in TIMSS 2019 die Lernenden sowohl die Klassenführung, die kognitive Aktivierung als auch die konstruktive Unterstützung in beiden Unterrichtsfächern positiv bewerten, wenn auch die Einschätzungen im Vergleich zu 2015 niedriger ausfallen. Ebenfalls in beiden Fächern wird in TIMSS 2019 die Motivierungsqualität von den Schülerinnen und Schülern als hoch bewertet. Während sich die Bewertungen der drei Basisdimensionen in Abhängigkeit vom Leistungsniveau der Grundschulkinder unterscheiden, erleben die Lernenden die Motivierungsqualität unabhängig ihres Leistungsstandes als positiv. Für die Wahrnehmung der Basisdimensionen ergibt sich entsprechend, dass leistungsstärkere Kinder die Klassenführung und konstruktive Unterstützung im Vergleich zu leistungsschwächeren Kindern höher wahrnehmen, während letztere die kognitive Aktivierung höher einschätzen.

In TIMSS 2019 wurden zudem Lehrkräfte befragt. Sie gaben Auskunft über fachspezifische und adaptive Gestaltungsaspekte des Unterrichts sowie über die leistungsbezogene Differenzierung im Mathematik- und naturwissenschaftsbezogenen Sachunterricht. In Anbetracht heterogener Klassen ist insbesondere die leistungsbezogene Differenzierung von Relevanz. Mehr als die Hälfte der Lernenden wird von Mathematiklehrkräften unterrichtet, die die Differenzierung für leistungsstärkere und -schwächere Schülerinnen und Schüler in ihrem eigenen Unterricht als hoch ausgeprägt einschätzen. Im naturwissenschaftsbezogenen Sachunterricht fällt der Anteil an Lernenden, deren Sachunterrichtslehrkräfte die leistungsbezogene Differenzierung im eigenen Unterricht als hoch ausgeprägt einstufen, geringer aus.

1.2.5 Privater Nachhilfeunterricht und Lehrkräftefortbildungen am Ende der Grundschulzeit

Neben dem regulären Schulunterricht erhält ein großer Anteil von Schülerinnen und Schülern zusätzlich privaten Nachhilfeunterricht. Diese Verbreitung von Nachhilfeunterricht kann einerseits positiv als Engagement für gute Schulleistungen gewertet werden, andererseits aber auch kritisch als Zeichen eines Vertrauensverlustes in das reguläre Schulsystem gesehen werden. Mit Blick auf internationale Trends wird seit vielen Jahren auf eine zunehmende Verbreitung von Nachhilfeunterricht hingewiesen. In Deutschland erhielt in TIMSS 2015 und auch in TIMSS 2019 jedoch weniger als jedes zehnte Kind Nachhilfeunterricht in Mathematik. Mit Blick auf den Sachunterricht spielt

Nachhilfe praktisch keine Rolle in Deutschland. Guill und Wendt untersuchen in Kapitel 7 die aktuelle Situation zur Nutzung von privatem Nachhilfeunterricht in Deutschland sowie das Wissen der Lehrkräfte darüber. Sie geben darüber hinaus auch einen Einblick in die Nutzung von Lehrkräftefortbildungen und wie diese in Deutschland und im internationalen Vergleich in Anspruch genommen werden.

In TIMSS 2019 zeigt sich, dass insbesondere Schülerinnen und Schüler mit nur ausreichenden oder schlechteren Mathematiknoten Nachhilfeunterricht bekommen. Auch wenn privater Nachhilfeunterricht einen finanziellen Aufwand für die Familien bedeutet, konnte festgestellt werden, dass die Anteile der Kinder aus armutsgefährdeten Haushalten, die Nachhilfe erhalten, nicht niedriger sind als die aus nicht armutsgefährdeten Familien. Bemerkenswert erscheint, dass die Erziehungsberechtigten die Lehrkräfte nicht immer über den Nachhilfeunterricht ihrer Kinder informieren. Dies schränkt eine zusammenwirkende Förderung von Kindern durch Nachhilfe- und Regelunterricht ein – eine mögliche Synergie wird in diesen Fällen nicht genutzt. In Bezug auf das Fortbildungsverhalten von Lehrkräften lässt sich feststellen, dass Grundschullehrerinnen und -lehrer in Deutschland durchaus regelmäßig fachspezifische Fortbildungsveranstaltungen wahrnehmen. Im internationalen Vergleich liegt die Teilnahmeintensität allerdings eher unterhalb des Niveaus des europäischen Durchschnitts. Zudem zeigt sich ein stärkeres Interesse an fachlichen, didaktischen und lehrplanbezogenen Fortbildungsangeboten vor jenen, die diagnostische Aspekte in den Fokus rücken. Fortbildungen zu Aspekten der individuellen Förderung von Schülerinnen und Schülern wird stärker von Mathematik- als von Sachunterrichtslehrkräften nachgefragt. Fortbildungen zum Einsatz digitaler Medien werden auffällig selten besucht.

1.2.6 Geschlechterunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen

Geschlechterunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen sind seit langer Zeit Gegenstand wissenschaftlicher Untersuchungen, gerade auch in Hinblick auf das Wahlverhalten junger Erwachsener betreffend einer Ausbildung in den Bereichen *Mathematik, Informatik, Naturwissenschaft und Technik* (MINT). Dabei zeigt sich, dass Kompetenzunterschiede zwischen Mädchen und Jungen in Mathematik und den Naturwissenschaften im Durchschnitt eher gering sind. Des Weiteren konnte herausgefunden werden, dass für Geschlechterunterschiede in Leistungen auch die Selbsteinschätzung der eigenen Fähigkeiten in der jeweiligen Domäne sowie der Wert, der diesem Bereich beigemessen wird, bedeutsam sind. Diese Aspekte stehen im Fokus des Kapitels von Nonte, Steinmayr und Scholz.

Die generellen Kompetenzunterschiede zwischen Mädchen und Jungen in Deutschland zeigen für Mathematik keinen eindeutigen Trend. In dieser Domäne konnten die im Vergleich zu 2007 ermittelten abnehmenden Geschlechterdifferenzen in 2011 und 2015 als Trendentwicklung in 2019 jedoch nicht mehr festgestellt werden. Die Mathematikleistungen von 2007 unterscheiden sich statistisch nicht signifikant zu denen in 2019. In den Naturwissenschaften zeigt sich hingegen ein abnehmender Trend in den Differenzen zwischen Mädchen und Jungen, die in 2007 bei 15 Punkten, in 2011 bei 12 Punkten und schließlich deutlich geringer in 2015 mit 3 und 2019 mit 4 Punkten ausfielen. Der Unterschied in den Differenzen hat von 2007 zu 2019 somit signifikant abgenommen. Der

Kompetenzanstieg, den die Mädchen in 2015 zeigen konnten, setzt sich jedoch nicht fort. Die Differenz zwischen Jungen und Mädchen bleibt im Vergleich von 2015 und 2019 nahezu unverändert – wenn auch insgesamt auf einem niedrigeren Niveau. Differenzierte Analysen für die Inhaltsbereiche in Mathematik und den Naturwissenschaften zeigen ebenfalls dieses Bild. Insgesamt haben Mädchen und Jungen hoch ausgeprägte positive Einstellungen zur Mathematik und zum Sachunterricht – auch wenn dieses positive Bild im Trend etwas abnimmt. Unterschiede im Selbstkonzept zugunsten der Jungen sind in Mathematik festzustellen, im Sachunterricht liegen keine signifikanten Unterschiede in TIMSS 2019 vor.

1.2.7 Soziale Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen

Die Frage nach Chancengleichheit im Bildungssystem steht im öffentlichen und wissenschaftlichen Interesse. Vor dem Hintergrund eines umfangreichen empirisch fundierten Wissens aus wiederholt durchgeführten Schulleistungsstudien ist es ein kontinuierliches bildungspolitisches Ziel, den Bildungserfolg von Schülerinnen und Schülern von ihrer sozialen Herkunft zu entkoppeln und für mehr Chancengleichheit zu sorgen.

Die Ungleichheit von Bildungschancen resultiert aus institutionellen Bedingungen des Bildungssystems, dem Lernverhalten der Schülerinnen und Schüler sowie aus den in den Familien der Schülerinnen und Schüler getroffenen Bildungsentscheidungen. Dabei zeigt sich für den Grundschulbereich in Deutschland nach wie vor eine deutliche Verknüpfung zwischen der sozialen Herkunft und dem Bildungserfolg der Schülerinnen und Schüler am Ende der vierten Jahrgangsstufe, die im Zeitverlauf relativ stabil ausfällt. Der Untersuchung dieser Zusammenhänge im Rahmen von TIMSS 2019 widmen sich Stubbe, Krieg, Beese und Jusufi in Kapitel 9 dieses Bandes.

Wie bereits für TIMSS 2007, 2011 und 2015 zeigt sich auch in TIMSS 2019 eine Kopplung zwischen dem sozioökonomischen Status von Familien und dem Bildungserfolg ihrer Kinder in Mathematik und den Naturwissenschaften. International betrachtet sind zwischen den Teilnehmerstaaten deutliche Unterschiede bei den Zusammenhängen zwischen den sozialen Hintergrundmerkmalen und den Kompetenzen der Viertklässlerinnen und Viertklässler zu beobachten. Im Trend betrachtet ist für Deutschland festzustellen, dass keine signifikanten Veränderungen des Zusammenhangs über die Zeit vorliegen.

Ein besonderes Maß des sozialen Hintergrundes ist der Indikator der *Armutsgefährdung*. Diese Gefährdung besteht insbesondere für Kinder aus Familien mit Migrationshintergrund und aus Familien mit einem niedrigen Bildungsniveau. Werden die Lernunterschiede von armutsgefährdeten mit nicht armutsgefährdeten Kindern verglichen, zeigt sich in TIMSS 2019 in Mathematik eine Differenz von knapp einem Lernjahr. Bei den naturwissenschaftlichen Kompetenzen beträgt diese Differenz bis zu zwei Lernjahre. Im Trend zeigen sich zwischen TIMSS 2007, 2011, 2015 und 2019 keine signifikanten Veränderungen bei den festgestellten sozialen Disparitäten.

1.2.8 Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund

Es gilt ausdrücklich als bildungspolitisches Ziel, sowohl zuwanderungsbezogene Disparitäten in der Bildungsbeteiligung als auch im Kompetenzerwerb möglichst gering zu halten und zu reduzieren. In TIMSS 2019 sind die Analysen von Wendt, Schwippert, Stubbe und Jusufi für die Betrachtung von zuwanderungsbezogenen Disparitäten in folgender Hinsicht besonders interessant: Angesichts der im letzten Jahrzehnt gestiegenen Zahl neu zugewanderter Kinder und Jugendlicher im schulpflichtigen Alter ist zu untersuchen, welche Bedeutung dies für die Leistungsstände von Kindern mit einem Migrationshintergrund hat.

Leistungsunterschiede, die zuungunsten von Schülerinnen und Schülern aus Familien mit Zuwanderungsgeschichte ausfallen, sind durch ein komplexes Zusammenspiel individueller Merkmale und Benachteiligungen sowie Strukturen und Praktiken des Bildungssystems zu erklären.

In dem seit 2007 alle vier Jahre nachgezeichneten Trend zeigen sich keine Befunde, die darauf hindeuten, dass migrationsbedingte Disparitäten in den letzten Jahren abgenommen haben. Seit 2007 ist die Befundlage somit praktisch stabil. Die Befunde aus TIMSS 2015, die auf eine mögliche systematische Reduktion der Leistungsdisparitäten über die Zeit hinwiesen, ließen sich in 2019 nicht replizieren. Zur Erklärung von Disparitäten zwischen Kindern aus Familien mit und ohne Migrationshintergrund trägt insbesondere die Betrachtung des familiären Sprachgebrauches und des sozioökonomischen Hintergrundes bei. Darüber hinaus belegen Schülerinnen- und Schülerangaben zur sozialen Eingebundenheit, dass sich Kinder mit Migrationshintergrund in den deutschen Grundschulen nicht im gleichen Maße eingebunden fühlen wie ihre Mitschülerinnen und Mitschüler aus Familien ohne Migrationshintergrund. Zudem sind sie nach eigenen Auskunft weniger selbstbewusst und haben auch weniger Freude am Lernen, insbesondere bezogen auf naturwissenschaftliche Inhalte.

1.2.9 Schullaufbahnpräferenzen am Übergang in die Sekundarstufe

Stubbe, Kasper und Jentsch befassen sich in Kapitel 11 mit den Übergangsentscheidungen nach der vierten Jahrgangsstufe. Die Bedeutung des Übergangs in die Sekundarstufe ist für den zukünftigen Bildungserfolg von Schülerinnen und Schülern in Deutschland unumstritten. Fraglich ist jedoch, wie durchlässig das Bildungssystem ist, wenn sich herausstellt, dass Schülerinnen und Schüler nicht auf der für sie geeigneten Schulform sind. Hierbei kommt der immer wieder in Untersuchungen festgestellten Abhängigkeit der Lehrkräfteempfehlungen beziehungsweise der Entscheidungen der Erziehungsberechtigten von der sozialen Lage der Schülerinnen und Schüler eine besondere Aufmerksamkeit zu.

Die vergangenen Jahre waren in Deutschland durch länderspezifische Reformen der Sekundarschulstruktur geprägt. Aufgrund des Bildungsföderalismus ist es in Deutschland nicht gelungen, sich auf einheitliche Namen für die neuen Schulformen zu einigen. Aktuell lassen sich 15 Schulform-Bezeichnungen (inkl. Hauptschule, Realschule und Gymnasium) finden, wobei mit gleichem Namen in verschiedenen Ländern unterschiedliche Schultypen bezeichnet werden. Es ist davon auszugehen, dass diese Uneinheitlichkeit bei der Wahl einer weiterführenden Schule insbesondere in bildungsferneren Haushalten beim Wechsel in andere Länder der Bundesrepublik Deutschland zu Verunsicherungen führt.

Die mit TIMSS 2019 seit 2011 nachgezeichneten Trendanalysen zum Übergang von der Grundschule in die Sekundarstufe I zeigen ein insgesamt gleichbleibendes Bild: Veränderungen ergeben sich vor allem durch die erwähnten Schulstrukturen. Positiv kann festgestellt werden, dass leistungsrelevante Merkmale der Schülerinnen und Schüler entsprechend der bildungspolitischen Vorgaben einen engen Zusammenhang mit der Schullaufbahnpräferenz der Lehrkräfte aufweisen. Als problematisch ist jedoch anzusehen, dass auch unter Kontrolle der Kompetenzen weiterhin ein deutlicher Zusammenhang zwischen der sozialen Herkunft der Schülerinnen und Schüler und den Schullaufbahnpräferenzen sowohl der Lehrkräfte als auch der Erziehungsberechtigten besteht.

1.2.10 Trends in Schülerzusammensetzungen und mathematischen und naturwissenschaftlichen Kompetenzen

Wiederholte Querschnittsuntersuchungen im Rahmen von *Large-Scale-Assessments* lassen vergleichende Bestandsaufnahmen von Bildungssystemen zu. Diese gewähren einen Einblick, welchen Stand und welche Veränderungen neben den erfassten Kompetenzständen der Schülerinnen und Schüler auch Merkmale des Lehrens und Lernens aufweisen. Für eine globale, beschreibende Gegenüberstellung von Bildungssystemen oder für die Betrachtung von Bildungssystemen über die Zeit ist diese Möglichkeit der Berichterstattung gut geeignet. Wenn sich jedoch verschiedene Entwicklungslinien überschneiden, stellt sich die Frage, wie sich bestimmte Merkmale verändert beziehungsweise gehalten hätten, wenn sich die Begleitumstände oder Rahmenbedingungen nicht geändert hätten. Dieser Frage wird in dem Kapitel von Kasper, Wendt, Schwippert und Köller nachgegangen.

In dem Kapitel kann zusammenfassend festgestellt werden, dass sich die Zusammensetzung der Schülerschaft im Betrachtungszeitraum von TIMSS 2007 bis 2019 gewandelt hat. Auf diese Veränderungen, die unter anderem durch globale Entwicklungen (Migration) oder nationale Entwicklungen (Entscheidung für ein inklusives Bildungssystem) zurückzuführen sind, musste reagiert werden beziehungsweise wurde im Bildungssystem reagiert. Dabei lässt sich festhalten, dass sich trotz der insbesondere für die Lehrkräfte gestiegenen Anforderungen 2019 keine zusätzlichen oder sich verstärkenden Problemlagen identifizieren lassen. Die Schülerinnen und Schüler zeigen gleichbleibende Kompetenzniveaus. Allerdings konnten im zeitlichen Verlauf auch soziale beziehungsweise migrationsbedingte Disparitäten nicht abgebaut werden.

1.3 Bildungspolitische und didaktische Folgerungen

Die Befundlage in TIMSS 2019 lässt sich, insbesondere mit Blick auf den TIMSS-Trend seit 2007, aus zwei Perspektiven betrachten: Die Befunde für Deutschland verweisen auf ein hohes Maß an Kontinuität. Dies kann durchaus positiv gewertet werden, da das Niveau sich im internationalen Vergleich als befriedigend darstellt. Die gleiche Kontinuität kann aber in Bezug auf die seit Jahren herausgearbeiteten Herausforderungen im deutschen Bildungssystem auch als beunruhigend wahrgenommen werden.

Welche Perspektive – also die der zufriedenstellenden oder die der beunruhigenden Bilanzierung – eingenommen wird, hängt von ganz unterschiedli-

chen Wertgewichtungen ab, wobei jede Forderung nach einer Verbesserung oder Veränderung auch immer Nebenwirkungen mit sich bringt, da den verschiedenen im Bildungssystem tätigen Personen insbesondere zeitliche und materielle Ressourcen nur begrenzt zur Verfügung stehen.

Forderungen nach einer intensiveren Auseinandersetzung mit bestimmten oder neuen Inhaltsbereichen im Unterricht ziehen die Frage nach sich, auf welche anderen Inhalte dann gegebenenfalls weniger eingegangen werden soll; oder wenn intensivere Förderungen von Kindern mit besonderen Voraussetzungen – was explizit nicht nur auf defizitorientierte Ausgangslagen fokussiert – erwartet werden, wäre mit zu überlegen, um welche insbesondere zeitlichen Ressourcen zum Beispiel die betreuenden Lehrkräfte entlastet werden können, um einer solchen neuen Forderung nachzukommen. Dies gilt auch für die Ausgestaltung und Schwerpunktsetzung in der Ausbildung von Lehrkräften im Studium oder deren Fortbildung. Welche Inhalte können weniger intensiv behandelt werden, wenn neue Anforderungen zu bewältigen sind?

Ziel muss es daher sein, abzuwägen, wie – insbesondere mit Blick auf die Befunde aus TIMSS 2019 – Grundschülerinnen und Grundschüler auf ihre Zukunft vorbereitet werden können, um ihnen die Gelegenheit zu eröffnen, sich zu informierten, selbstständigen, kritischen und auch sozialen und konstruktiven Mitgliedern der Gesellschaft entwickeln zu können. Werden diese langfristig zu erreichenden Ziele tatsächlich allen Kindern zugesprochen, bedarf es nicht nur einer Haltung, niemanden zurückzulassen, sondern auch das Beste für alle Kinder und damit perspektivisch für alle Mitglieder der zukünftigen Generationen zu wollen.

Um diesen Anforderungen gerecht zu werden, liegt es auf der Hand, dass – mit Blick auf Schule und Unterricht – der Schlüssel in einer angemessenen Diagnostik (feststellen, wo jedes Kind gerade steht), einer optimalen Förderung (um den Anschluss zu finden oder über sich hinauszuwachsen) und einer zukunftsorientierten Ausrichtung von Lehrkräften in einer sich gerade auch medial stark wandelnden Welt liegt. So einfach und global diese Forderungen erscheinen, so anspruchsvoll und – vor dem Hintergrund der Befunde von TIMSS 2019 – so aktuell und notwendig erscheinen sie.

Im Berichtsband lassen sich Handlungsbedarfe erkennen, die es auch weiterhin im Blick zu behalten gilt:

- Ein relativ hoher Anteil von Viertklässlerinnen und Viertklässlern droht in Mathematik beziehungsweise in den Naturwissenschaften den Anschluss aufgrund unzureichender Grundkenntnisse, auf die sie im Rahmen ihres weiteren Lernprozesses aufbauen können sollten, zu verlieren. Der ebenfalls in Mathematik beziehungsweise den Naturwissenschaften sehr geringe Anteil von Kindern auf der höchsten Kompetenzstufe verweist auf fehlende flächendeckende und systematische Förderprogramme für Schülerinnen und Schüler, die bereits über ein sehr ausgeprägtes Vorwissen verfügen oder sehr effiziente Lernstrategien haben. Beide Befundlagen weisen darauf hin, wie notwendig es ist, sich im Rahmen der Lehrkräfteaus- und insbesondere auch der Lehrkräftefortbildung stärker mit Fragen der inhaltsbezogenen Diagnostik und der fokussierten Förderung auseinanderzusetzen.
- Grundschülerinnen und Grundschüler kommen mit inhaltlich gleichen, aber anhand digitaler Medien präsentierten Testformaten nicht so gut zurecht, wie mit herkömmlichen gedruckten Test- und Fragebogenformaten. In einer zunehmend durch digitale Medien geprägten Gesellschaft müssten Grundschülerinnen und Grundschüler systematisch an den konstruktiven und kri-

tischen Gebrauch dieser Medien – auch im Kontext des Fachunterrichts – herangeführt werden. Lehrkräfte müssten für die Nutzung digitaler Medien umfangreicher und verbindlicher fortgebildet werden.

- Durch außerschulischen Nachhilfeunterricht versuchen Erziehungsberechtigte Defizite, die ihre Kinder beim Lernen zeigen, auf privatem Wege zu kompensieren. Die Möglichkeit schulischen und außerschulischen (Nachhilfe-) Unterricht systematisch aufeinander zu beziehen und so Synergien zu erzeugen, sollte insbesondere durch kommunikative Initiativen der Lehrkräfte den Erziehungsberechtigten erschlossen werden. Langfristig sollte in den Blick genommen werden, dass problematische schulische Lernprozesse gemeinsam durch die Lehrkräfte und Erziehungsberechtigten begleitet werden sollten.
- Mädchen erreichen im Durchschnitt in Grundschulen in Deutschland nicht das gleiche mathematische Kompetenzniveau wie Jungen. In der Forschung und auch in der Praxis sollte diesen Disparitäten, die in einer emanzipierten modernen Gesellschaft nicht akzeptabel sind, weiter nachgegangen werden. Dies gilt auch für systematische Nachteile der Jungen in den sprachlichen Fächern. Ein möglicher Ansatzpunkt könnte dabei eine Auseinandersetzung mit gesellschaftlich geprägten und gegebenenfalls im Unterricht verfestigten Stereotypen sein.
- Trotz fortdauerndem Bemühen Chancenunterschieden im Bildungserwerb zu begegnen, die im Zusammenhang mit dem sozioökonomischen Hintergrund der Grundschülerinnen und Grundschüler beziehungsweise deren Migrationshintergrund stehen, gelingt es mit Blick auf TIMSS seit 2007 nicht, diese Scheren zu schließen. Dies gilt auch in Hinblick auf Chancenunterschiede beim Übergang von der Grundschule in die weiterführende Schule. Dabei wäre – um ungerechtfertigte verallgemeinernde Zuschreibungen im Zusammenhang mit bestimmten Gruppen von Schülerinnen und Schülern zu vermeiden – eine zielgerichtete inhaltsbezogene und alltagstaugliche Diagnostik und ein handhabbares und wirkungsvolles Förderrepertoire bei den Lehrkräften notwendig, damit die Unterstützung so zielgerichtet erfolgen kann, wie sie erforderlich ist.

Literatur

- Kelly, D., Centurino, V. A. S., Martin, M. O. & Mullis, I. V. S. (2020). *TIMSS 2019 encyclopedia: Education policy and curriculum in mathematics and science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center. Boston College. Verfügbar unter: <https://timssandpirls.bc.edu/timss2019/encyclopedia/>
- Mullis, I. V. S. & Martin, M. O. (Hrsg.). (2017). *TIMSS 2019 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center. Boston College. Verfügbar unter: <http://timssandpirls.bc.edu/timss2019/frameworks/>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center. Boston College. Verfügbar unter: <http://timssandpirls.bc.edu/timss2019/international-results/>

Kapitel 2

Ziele, Anlage und Durchführung der *Trends in International Mathematics and Science Study* (TIMSS 2019)

Knut Schwippert, Luise A. Scholz, Christin Beese, Daniel Kasper, Katrin Schulz-Heidorf und Anna-Lena Girelli

2.1 TIMSS – Schule und Unterricht im internationalen Vergleich

Wissenschaftliche Untersuchungen in Schule und Unterricht haben unter anderem das Ziel, Rahmenbedingungen, Voraussetzungen und Prozesse des Lehrens und Lernens zu optimieren. Die Forschung in diesem Bereich muss ständig mit sich wandelnden Bedingungen, unter denen gelehrt und gelernt wird, Schritt halten. Ursachen für diese Veränderungen liegen auf ganz unterschiedlichen Ebenen: So verändert sich unsere Gesellschaft und deren Zusammensetzung fortwährend und stellt das Bildungssystem vor neue Herausforderungen. Beispielhaft kann für Deutschland das Bemühen in Schule und Unterricht um eine gleichberechtigte Förderung von Mädchen und Jungen oder auch von Kindern unterschiedlicher sozialer Herkunft angeführt werden. Es ergeben sich aber auch Herausforderungen durch die fortschreitende Internationalisierung und Globalisierung. Ab den 1960er-Jahren entstanden diese aus der Schulpflicht von Kindern aus Familien mit Migrationsgeschichte und stellen sich aktuell – unter ganz anderen weltpolitischen Umständen und mit anderen humanitären Ansprüchen – wieder neu. Weitere Aufgaben, die zu lösen sind, ergeben sich auch durch die Umsetzung der anerkannten UN-Behindertenrechtskonvention zur Inklusion (Beauftragte der Bundesregierung für die Belange von Menschen mit Behinderungen, 2017). Und schließlich stellt die stetige Bildungsexpansion, die zum Beispiel an beständig wachsenden Gymnasialquoten abzulesen ist, eine weitere neue Herausforderung dar. Für die Akteurinnen und Akteure im Bildungssystem, die auf den unterschiedlichen Ebenen entscheiden müssen, wie mit diesen Herausforderungen umgegangen werden soll, sind Beispiele hilfreich, wie in anderen Bildungssystemen auf gleiche oder ähnliche Herausforderungen reagiert wurde. Mithilfe wissenschaftlicher (Vergleichs-)Untersuchungen werden den Handlungsträgerinnen und Handlungsträgern dem aktuellen Stand der Forschung entsprechende relevante

Informationen bereitgestellt, auf deren Basis diese konkrete Pläne und Strategien zur Bewältigung von Herausforderungen im Bildungssystem entwickeln können.

Die *Trends in International Mathematics and Science Study* (TIMSS) legt den Schwerpunkt auf die Untersuchung der mathematischen und naturwissenschaftlichen Kenntnisse von Schülerinnen und Schülern verschiedener Staaten unter Berücksichtigung curricularer Vorgaben und zentraler Rahmenbedingungen schulischer und familiärer Lernumgebungen. Dabei geht es vor allem darum, langfristige Entwicklungen in den teilnehmenden Staaten zu dokumentieren, zu interpretieren und Entscheidungshilfen für mögliche zukünftige Weiterentwicklungen der Bildungssysteme zu bieten. Die TIMSS-Untersuchung wird daher auch als Systemmonitoringstudie (oder ‚Bildungsmonitoring-‘ bzw. ‚großangelegte Schulvergleichsstudie‘) bezeichnet.

Die TIMSS-Untersuchung wurde als Vergleichsuntersuchung für Schülerinnen und Schüler unterschiedlicher Altersgruppen konzipiert. TIMSS bietet die Möglichkeit, Schülerinnen und Schüler am Ende der Grundschulzeit (i.d.R. 4. Klasse) und gegen Ende der Pflichtschulzeit (i.d.R. 8. Klasse) auf ihre mathematischen und naturwissenschaftlichen Kompetenzen hin alle vier Jahre zu untersuchen. Der zyklische Abstand von vier Jahren wurde gewählt, damit Staaten, die beide Altersgruppen untersuchen möchten, die gleiche Schülerinnen- und Schülerkohorte in der achten Klassenstufe befragen, die vier Jahre zuvor die Grundschule besucht hat und dort die Zielgruppe der TIMSS-Grundschuluntersuchung war. In Deutschland wurde entschieden, sich ab 2007 lediglich an der TIMSS-Grundschuluntersuchung in der vierten Klassenstufe zu beteiligen, da die älteren Schülerinnen und Schüler im Wesentlichen die Zielgruppe der ebenfalls regelmäßig durchgeführten Untersuchung des *Programme for International Student Assessment* (PISA) darstellen und eine doppelte Belastung der Bildungseinrichtungen durch praktisch parallel durchgeführte Untersuchungen vermieden werden sollte. Als zusätzliche TIMSS-Zielgruppe wurden in ausgewählten Jahren (1995, 2008 und 2015) auch Schülerinnen und Schüler untersucht, die am Ende der Regelschulzeit standen. An dieser Untersuchung gab es nur im Jahr 1995 eine deutsche Beteiligung.

In diesem Kapitel werden zunächst die Organisation von TIMSS sowie die beteiligten Institutionen vorgestellt, um anschließend auf Ziele, Anlage und Durchführung der TIMSS-2019-Untersuchung einzugehen und wesentliche Merkmale der Studie zu dokumentieren. Im Abschluss dieses Kapitels werden technische und empirische Voraussetzungen für die Auswertung sowie einige Hinweise zur Ergebnisdarstellung in diesem Band vorgestellt.

2.2 TIMSS – ein internationales und nationales Forschungsnetzwerk

Eine Untersuchung wie TIMSS lässt sich nur durch den Zusammenschluss von Expertinnen und Experten mit ganz unterschiedlichen internationalen und nationalen Fachkenntnissen realisieren. Initiator der TIMSS-Untersuchung ist die *International Association for the Evaluation of Educational Achievement* (IEA). Die IEA ist ein gemeinnütziger Verein, der durch einen Kreis von Bildungsexpertinnen und -experten gegründet wurde. Diese haben schon vor Jahrzehnten das Potenzial erkannt, aus dem internationalen Vergleich von Bildungssystemen im Sinne eines „natürlichen Experiments“ (Husén, 1967, S. 27) mit wissenschaftlichen Mitteln Informationen über die Gestaltung und schließlich zur Weiterentwicklung von Bildungssystemen zu gewinnen. Die IEA hat inzwischen mehr

als 60 Jahre Erfahrung in der Konzeption, der Durchführung und der Auswertung von internationalen Systemmonitoringstudien. Untersuchungen der IEA zur Erfassung mathematischer und naturwissenschaftlicher Kompetenzen von Schülerinnen und Schülern gab es schon vor der ersten TIMSS-Untersuchung 1995, allerdings wurden erst mit TIMSS die wiederholten Untersuchungen im Sinne der Abbildung eines Trends systematisch aufeinander bezogen.

Die IEA hat das internationale Management der Studie dem *TIMSS & PIRLS International Study Center* (ISC) am Boston College anvertraut, welches unter Leitung von Ina V. S. Mullis, Professorin an der Lynch School of Education, Boston College, sowie Michael O. Martin, Research Professor an der Lynch School of Education, Boston College, steht. Das ISC ist verantwortlich für das Design und die Implementierung der Studie, für die internationale Koordination der Entwicklung der Instrumente und der Erhebungsprozeduren sowie für die Überprüfung der Qualität der Datenerhebung. Darüber hinaus wird am ISC die internationale Skalierung der Tests durchgeführt und der internationale Ergebnisbericht verfasst. Für die Stichprobenziehung, die Dokumentation der nationalen Stichproben und die Berechnung der internationalen Stichprobengewichte kooperiert die internationale Studienleitung mit Statistics Canada (Sylvie LaRoche) in Ottawa, Ontario und der Abteilung für Stichprobenziehung der IEA Hamburg (Dr. Sabine Meinck) (Mullis & Martin, 2017).

Auch auf nationaler Ebene wäre die TIMSS-Untersuchung ohne den Zusammenschluss von verschiedenen Expertinnen und Experten nicht möglich. In Deutschland war Prof. Dr. Wilfried Bos (Technische Universität Dortmund) für die TIMS-Studien 2007, 2011 und 2015 verantwortlich, die dieser 2011 und 2015 gemeinsam mit Prof. Dr. Heike Wendt (damals Technische Universität Dortmund – aktuell Karl-Franzens-Universität Graz) durchgeführt hat. Die wissenschaftliche Gesamtleitung und die Koordinierung des nationalen TIMSS-2019-Konsortiums liegt bei Prof. Dr. Knut Schwippert (Universität Hamburg), unterstützt durch die nationale Projektleitung für Methoden Dr. Daniel Kasper und für Administration Christin Beese, Dr. Armin Jentsch und Dr. Katrin Schulz-Heidorf (bis 2019). Dabei haben die wissenschaftlichen Mitarbeiterinnen Donieta Jusufi, Anna-Lena Girelli (geb. Kremer; bis 2019) und Luise A. Scholz wesentlich zum Gelingen des Projekts und der nationalen Berichtslegung beigetragen. Eine solche Berichtslegung ist ohne die tatkräftige Unterstützung von vielen engagierten studentischen Hilfskräften nicht möglich.¹

Das nationale TIMSS-2019-Konsortium besteht aus einschlägig ausgewiesenen Wissenschaftlerinnen und Wissenschaftlern, die ihre Expertise aus Untersuchungen von mathematischen und naturwissenschaftlichen Kompetenzen, der Gestaltung von Lehr- und Lerngelegenheiten und schließlich auch der Nutzung von digitalen Medien in Schule und Unterricht in die Konzeption, Durchführung und Auswertung der TIMSS-Untersuchung eingebracht haben. An der Vorbereitung und Durchführung von TIMSS 2019 sind als Verbundpartnerinnen und Verbundpartner Prof. Dr. Olaf Köller (Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik – IPN Kiel), Prof. Dr. Mirjam Steffensky (seit 2020 Universität Hamburg, davor IPN Kiel), Prof. Dr. Nele McElvany und Prof. Dr. Christoph Selter (Technische Universität Dortmund) beteiligt. Darüber hinaus wirken in dem Konsortium mit: Prof. Dr. Frank Goldhammer (DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation in Frankfurt),

1 Unser Dank gilt hierbei Katharina Boße, Jannis Burkhard, Jan Höpfner, Melanie Jagst, Celestine Kleinesper, Julia Koplin, Alina Kothe, Neele Krüger, Svenja Krümel, Martje Lott, Liska Oncken, Elena Zavjalov.

Prof. Dr. Aiso Heinze (IPN Kiel), Dr. Ulf Kröhne (DIPF), Prof. Dr. Tobias C. Stubbe (Georg-August-Universität Göttingen) und Prof. Dr. Heike Wendt (Karl-Franzens-Universität Graz). Damit sind im Konsortium Vertreterinnen und Vertreter der Erziehungswissenschaft, der Empirischen Bildungsforschung, der Pädagogischen Psychologie, der Mathematikdidaktik und der Didaktik des Sachunterrichts berücksichtigt.

Das Konsortium wurde bei der Vorbereitung der Untersuchungsinstrumente und den Genehmigungsverfahren in den Ländern der Bundesrepublik Deutschland durch die IEA Hamburg als Sub-Kontraktor unterstützt, die schließlich auch die Testorganisation und -durchführung in den Grundschulen realisiert hat. Ermöglicht wurde die Beteiligung Deutschlands an der internationalen TIMSS-Untersuchung mithilfe der Finanzierung durch das *Bundesministerium für Wissenschaft und Forschung* (BMBF) und die *Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland* (KMK). Ebenso übernahmen auch die in Deutschland an der Durchführung der TIMSS-Untersuchung beteiligten Universitäten Anteile der Finanzierung der Untersuchung.

2.3 Ziele der TIMSS-Untersuchung

TIMSS liefert Informationen über das nationale Bildungssystem, aber auch über Bildungssysteme im internationalen Vergleich. Im Fokus stehen dabei die Kompetenzen von Schülerinnen und Schülern in Mathematik und den Naturwissenschaften. Als Kompetenz wird die Grundbildung der Schülerinnen und Schüler bezeichnet, über die sie verfügen, um in verschiedenen Situationen die an sie gestellten Anforderungen zu erfüllen (Robitzsch et al., 2017; Weinert, 2001). Neben den Kompetenzständen der Schülerinnen und Schüler werden auch Hintergrundinformationen ermittelt, zum Beispiel zum persönlichen Lernumfeld, zu den Lehrkräften und zum Unterricht oder zu den organisatorischen Rahmenbedingungen der Schulen. Diese werden sowohl einzeln als auch im Zusammenhang mit den Kompetenzen der Schülerinnen und Schüler betrachtet. Ziel von TIMSS ist es dabei ausdrücklich nicht, möglichst präzise Informationen über einzelne Schülerinnen und Schüler zu erhalten (Individualdiagnostik), sondern die befragten Schülerinnen und Schüler, die in Klassen, Schulen und Bildungsgängen organisiert sind, als Repräsentantinnen und Repräsentanten der gesamten Schülerschaft eines Jahrgangs zu betrachten.

2.3.1 Theoretischer Hintergrund

Die von der IEA durchgeführten internationalen Schulleistungsstudien folgen in ihrer theoretischen Rahmung traditionell einem Input-Prozess-Output-Modell, welches anhand der Beschreibung unterschiedlicher Curricula für das schulische System konkretisiert wird. Unterschieden wird hierbei:

- Input: Das intendierte Curriculum entspricht den generellen curricularen Vorgaben seitens Politik beziehungsweise Administration.
- Prozess: Das potenzielle Curriculum beschreibt die möglichen Lehrinhalte, die zum Beispiel in Schulbüchern dokumentiert sind, während das implementierte Curriculum die tatsächlichen Inhalte des Unterrichts umfasst.
- Output: Das realisierte Curriculum schließlich beschreibt, was von den Schülerinnen und Schülern tatsächlich gelernt wurde.

Mit diesem Modell wird angestrebt, die Teilnehmerstaaten hinsichtlich curricularer Vorgaben und ihrer *Unterrichtsqualität* zu vergleichen. Studien der letzten Jahre haben die Rolle unterschiedlicher instrukionaler Merkmale und Kompetenzen der Lehrenden für die positive Lernentwicklung von Schülerinnen und Schülern intensiv thematisiert. Aufbauend auf wichtigen Vorarbeiten (Helmke, 2009; Klieme, Schümer & Knoll, 2001; Stigler & Hiebert, 1999; Tschannen-Moran & Woolfolk-Hoy, 2001) hat sich insbesondere im deutschen Sprachraum eine Konzeptualisierung von Unterrichtsqualität anhand von drei Basisdimensionen etabliert: (1) Effiziente Klassenführung und Strukturierung, (2) kognitive Aktivierung und (3) konstruktive Unterstützung (Fauth, Decristan, Rieser, Klieme & Büttner, 2014; Kunter, 2016). Im Kontext heterogener Lerngruppen ist dabei das adaptive Unterrichten ein Bereich, der besonderer Aufmerksamkeit bedarf (Corno, 2008; Helmke, 2009).

Die drei Basisdimensionen der Unterrichtsqualität wurden bereits in TIMSS 2015 durch Einschätzungen der Schülerinnen und Schüler sowie der Lehrkräfte erhoben. Eine analoge und in Teilen erweiterte Erfassung der zugehörigen Aspekte der Unterrichtsqualität bietet wichtige Informationen über die Gestaltung des Unterrichts in Mathematik und dem naturwissenschaftsbezogenen Sachunterricht in deutschen Grundschulen im Jahr 2019 und ermöglicht zudem Aussagen über Unterschiede der Unterrichtsqualitätsmerkmale zwischen 2015 und 2019. Darüber hinaus ermöglicht die Erfassung von Unterrichtsqualitätsmerkmalen die Analyse von Zusammenhängen mit kognitiven und motivational-affektiven Lernergebnissen im Grundschulunterricht in den Fächern Mathematik und Sachunterricht. Durch die repräsentative Datengrundlage leisten Analysen im Rahmen von TIMSS 2019 einerseits einen wichtigen Beitrag zur Theoriebildung bezüglich unterrichtlicher Lehr-Lernprozesse und liefern andererseits wichtige Hinweise für die Verbesserung pädagogischer Praxis im Bereich Mathematik und Naturwissenschaften.

Wie in den früheren Erhebungen sind auch in TIMSS 2019 aussagekräftige Befunde über den mathematischen und naturwissenschaftsbezogenen Unterricht in der Grundschule möglich. Statistische Zusammenhänge zwischen Unterrichtsmerkmalen und Leistungen bieten Hinweise darauf, wie Unterricht weiterentwickelt werden kann, um noch erfolgreichere mathematisch-naturwissenschaftliche Bildungsverläufe aufseiten der Schülerinnen und Schüler zu ermöglichen.

2.3.2 Hauptanliegen von TIMSS

Für die beteiligten Staaten bietet die TIMSS-Untersuchung im Wesentlichen drei Untersuchungsperspektiven: Durch die regelmäßige Wiederholung der Untersuchung können (1) Aussagen über Entwicklungen in den untersuchten Bildungssystemen getroffen werden und (2) durch die internationale Beteiligung von 64 Staaten und Regionen in 2019 können Vergleiche der verschiedenen Bildungssysteme vorgenommen werden. Hierbei zeigt sich, dass einerseits einige Bildungssysteme ähnlich gestaltet sind, andererseits aber auch Staaten an TIMSS teilnehmen, deren Bildungssysteme sehr unterschiedlich konzipiert sind. Dies ermöglicht zum einen kontrastreiche Vergleiche, zum anderen können diese Bildungssysteme als Ideengeber für das eigene Bildungssystem dienen. Und schließlich können (3) auf nationaler Ebene spezifische aktuelle Herausforderungen im eigenen Bildungssystem in den Blick genommen werden und damit denjenigen Informationen zur Verfügung gestellt werden, die für dessen Gestaltung zuständig sind.

Internationale Vergleiche

Wenn im Rahmen von internationalen Vergleichsstudien die Bildungssysteme von Staaten miteinander verglichen werden sollen (2. Untersuchungsperspektive), ist sicherzustellen, dass die zum Vergleich herangezogenen Informationen auch tatsächlich vergleichbar sind. Zur allgemeinen Bestandsaufnahme ist es sicherlich hilfreich, so viele Staaten wie möglich mit in den Blick zu nehmen. Sollen jedoch Vergleiche mit Staaten mit ähnlichem historischen, kulturellen oder ökonomischen Hintergrund vorgenommen werden, ist eine entsprechende Auswahl zu treffen. Wenn schließlich ein besonderes Interesse an Veränderungen im Bildungssystem besteht, sollten die Staaten in den Blick genommen werden, die zu gleichen Zeitpunkten wiederholt an der TIMS-Studie teilgenommen haben.

Im vorliegenden Berichtsband zu TIMSS 2019 werden dazu entsprechende Vergleichsgruppen (VG) gebildet. Es finden sich Vergleiche mit all jenen Staaten, die an TIMSS 2019 beteiligt waren, Vergleiche mit den OECD-Staaten, die ähnliche ökonomische Hintergründe und Zielsetzungen teilen, und schließlich gibt es die Vergleichsgruppe der an TIMSS 2019 teilnehmenden EU-Staaten. In Tabelle 2.1 sind die in dem Bericht zusammengefassten Vergleichsgruppen beschrieben und deren konstituierende Staaten aufgeführt.

Bei einigen internationalen Vergleichen werden nicht alle in TIMSS 2019 teilnehmenden Staaten aufgeführt. Diese verkürzten Teilnehmerlisten umfassen nur Staaten und *Benchmark*-Teilnehmer, die EU- oder OECD-Mitglied sind oder sich in der jeweiligen Domäne im Mittelwert nicht signifikant von Deutschland unterscheiden oder signifikant besser sind.

Trendvergleiche

Neben dem Vergleich von Bildungssystemen ist ein weiteres Hauptanliegen von TIMSS 2019 die Fortsetzung des bereits bestehenden TIMSS-Trends im Kontext aktueller internationaler und nationaler Forschungsdiskurse. Langfristiges Ziel ist die Bereitstellung von Auswertungen der Kenntnisstände verschiedener Schülerinnen- und Schülergruppen und deren Entwicklung im Trend. Darüber hinaus werden beschreibende Statistiken zur Gestaltung von Schule und Unterricht und deren Zusammenhang mit den erfassten Kenntnisständen der Grundschülerinnen und Grundschüler dokumentiert.

Tabelle 2.1: Vergleichsgruppen in TIMSS 2019

VG _{EU} (n = 24)		VG _{OECD} (n = 29)	
Belgien (Fläm. Gem.)	Österreich	Australien	Niederlande
Bulgarien	Polen	Belgien (Fläm. Gem.)	Nordirland
Dänemark	Portugal	Chile	Norwegen
Deutschland	Schweden	Dänemark	Österreich
England*	Slowakei	Deutschland	Polen
Finnland	Spanien	England	Portugal
Frankreich	Tschechische Republik	Finnland	Republik Korea (Südkorea)
Irland	Ungarn	Frankreich	Schweden
Italien	Zypern	Irland	Slowakei
Kroatien		Italien	Spanien
Lettland		Japan	Tschechische Republik
Litauen		Kanada	Türkei
Malta		Lettland	Ungarn
Niederlande		Litauen	USA
Nordirland*		Neuseeland	

* = Zum Zeitpunkt der Haupterhebung in 2019 waren England und Nordirland noch Teil der EU.

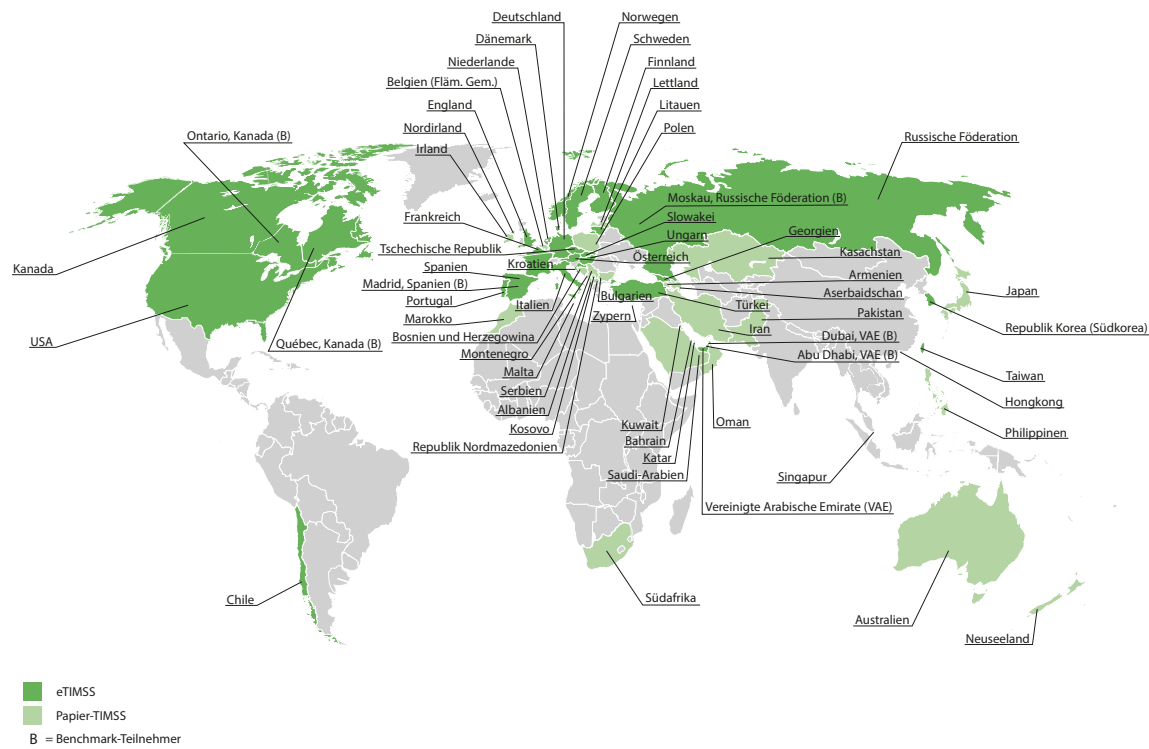
IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Eine solche langfristige Perspektive ist bei der Betrachtung eines so komplexen Systems, wie es die Schule darstellt, notwendig, wenn kurzfristig bedingte Unterschiede in Schule und Unterricht von langfristig angelegten Reformen und Förderlinien analytisch getrennt werden sollen. Hierbei bietet gerade auch die in TIMSS gebotene internationale Vergleichsperspektive wichtige Anhaltspunkte, da eine Vielzahl von Staaten mitunter gleichermaßen von globalen oder regionalen politischen, ökonomischen und humanitären Veränderungen und Herausforderungen betroffen war, diesen jedoch im Bildungssystem mit unterschiedlichen Maßnahmen begegnete.

Mit der Teilnahme Deutschlands an der TIMSS-2019-Grundschuluntersuchung bietet sich die Möglichkeit, die sich zwischen den Erhebungen von 2007, 2011 und 2015 ergebenden Unterschiede in den mathematischen und naturwissenschaftlichen Kompetenzen von Viertklässlerinnen und Viertklässlern bis 2019 im Trend zu betrachten. Durch die Anlage der Untersuchung als wiederholte Querschnitterhebung können die individuellen Entwicklungen von Schülerinnen und Schülern – anders als bei Panelstudien, in denen Probandinnen und Probanden im zeitlichen Verlauf mehrfach befragt werden – nicht nachgezeichnet werden. Jedoch können Aussagen über generelle Unterschiede und auch Entwicklungen in Bezug auf Merkmale der Bildungssysteme getroffen werden. So stellen beispielsweise der demografische Wandel (mit sinkenden Schülerinnen- und Schülerzahlen), regionale Bevölkerungsverschiebungen (Konzentration in Ballungsgebieten) und auch der ansteigende Anteil von Schülerinnen und Schülern mit Migrationshintergrund ständig neue Anforderungen an das Bildungssystem. Eine sich wandelnde Zusammensetzung der Schülerschaft und unterschiedliche Zeitpunkte des Eintritts in das Bildungssystem erfordern konkret eine darauf reagierende Schul- und Unterrichtsentwicklung. Aspekte dieser Entwicklungen, soweit sie über die Instrumentarien von TIMSS zu erfassen sind, gilt es

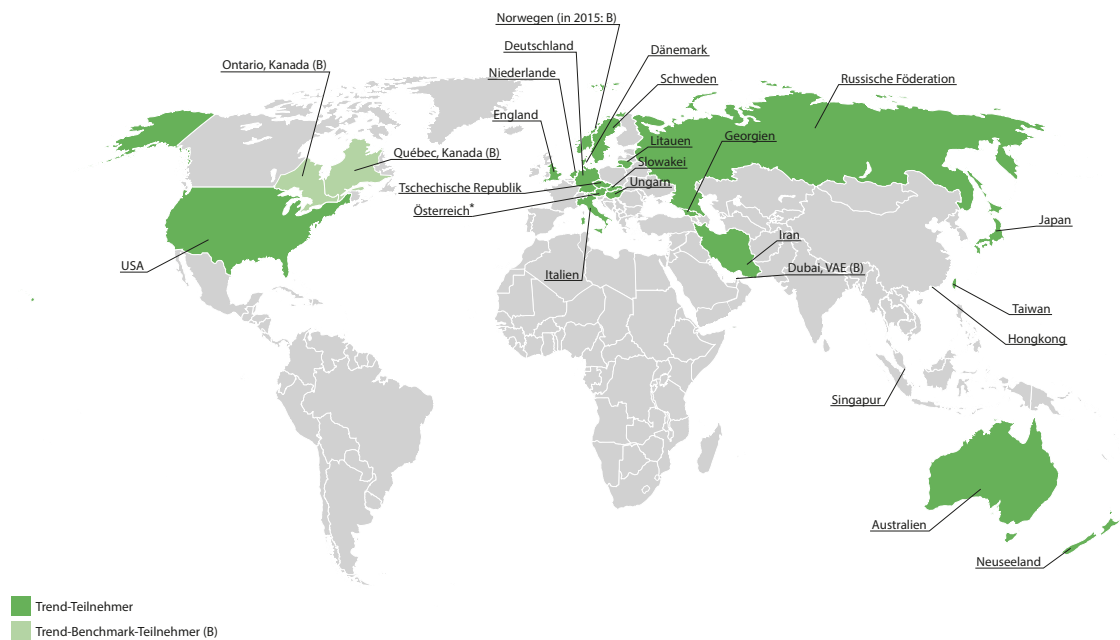
Abbildung 2.1: Staaten und Regionen, die in 2019 an eTIMSS und Papier-TIMSS teilgenommen haben



IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Abbildung 2.2: Teilnehmer an TIMSS 2007, 2011, 2015 und 2019



IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

nachzuzeichnen und daraus Impulse für Weiterentwicklungen zu gewinnen. In diesem Band werden bei den Analysen der Trenduntersuchungen die Staaten, die sich – wie auch Deutschland – an TIMSS 2007, 2011, 2015 und 2019 beteiligt haben, einbezogen. Zusätzlich wurde Österreich mitberücksichtigt, obwohl in 2015 die Teilnahme nicht realisiert werden konnte. Grund hierfür ist, dass es sich bei Österreich aktuell um den einzigen anderen deutschsprachigen an TIMSS teilnehmenden Staat handelt, der auch in der jüngeren Geschichte Bildungsstandards eingeführt hat. In Abbildung 2.1 sind die an TIMSS 2019 teilnehmenden Staaten aufgeführt und in Abbildung 2.2 die Staaten, die seit 2007 wiederholt teilgenommen haben.

2.3.3 eTIMSS als Reaktion auf eine zunehmende Digitalisierung

Als Reaktion auf sich wandelnde Lehr- und Lerngelegenheiten hat sich die IEA für den Zyklus 2019 erstmals dazu entschieden, TIMSS mithilfe digitaler Medien durchzuführen (eTIMSS). Die Umstellung der Testung von Papier und Bleistift auf digitale Medien² unterliegt der Anforderung, dass die im Rahmen von TIMSS begonnenen Trenduntersuchungen auch zukünftig in gleicher Qualität weitergeführt werden können. Dies bedeutet, dass die Umstellung für alle Staaten, die bisher an TIMSS teilgenommen haben, anhand einer erweiterten Stichprobe von Schülerinnen und Schülern empirisch abzusichern ist. Das ISC am Boston College bietet den Staaten mit eTIMSS im Jahr 2019 die erste und im Jahr 2023 die letzte reguläre Möglichkeit, diese Umstellung vorzubereiten. Ohne eine solche Übergangsstudie kann für die beteiligten Staaten die Fortschreibung des TIMSS-Trends nicht gewährleistet werden. Deutschland hat sich bereits bei der ersten Möglichkeit der Umstellung an der digitalen Testung im Rahmen von TIMSS 2019 beteiligt und so die Gelegenheit genutzt, nationale Erfahrungen, Anforderungen und Interessen bereits frühzeitig in die Umsetzung der digitalen Testung einfließen zu lassen. Dies bedeutet auch, dass bereits in 2019 in Deutschland vorliegende Moduseffekte – also Leistungsunterschiede, die auf die unterschiedlichen Testformate zurückzuführen sind – in den Grundschulen untersucht werden können.

Die Umstellung der papierbasierten Testung auf die Testung mithilfe digitaler Medien bedeutet nicht nur eine Übertragung der bisherigen Papier-Bleistift-Tests auf ein elektronisches Medium, sie bietet vielmehr auch die Möglichkeit, innovative Aufgabenformate zu administrieren, die komplex und authentisch sind, Multimediaelemente enthalten und einen hohen Interaktionsgrad aufweisen können. Auch wenn eine flächendeckende Digitalisierung der schulischen Lernumwelt in Deutschland noch nicht erreicht ist (Eickelmann et al., 2019), zeigt sich doch anhand aktueller Untersuchungen, dass die Nutzung digitaler Medien bei Grundschülerinnen und Grundschülern durchaus bereits angekommen ist, was bedeutet, dass eine Testung unter Nutzung von digitalen Medien für

2 In der Literatur finden sich für papier- beziehungsweise computerbasierte Testformate auch die Abkürzungen PBA (*paper-based-assessment*) und CBA (*computer-based-assessment*). In der Medienforschung in Deutschland hat sich jedoch etabliert, globaler von ‚digitalen Medien‘ statt nur von ‚Computern‘ zu sprechen, da inzwischen digitale Medien im Alltag genutzt werden, die nicht unter dem Begriff ‚Computer‘ subsummiert werden (z.B. Desktop-Computer, Laptops, Netbooks, Tablet-Geräte und Smartphones – wenn diese nicht ausschließlich zum Telefonieren oder zum Schreiben von Textnachrichten genutzt werden) (Eickelmann, Bos & Labusch, 2019). Dieser Begriffsverwendung wird außer im methodisch orientierten Kapitel 5 auch in diesem Band gefolgt.

die in TIMSS 2019 befragte Schülerpopulation keine generellen Probleme der Handhabung mit sich bringen sollte (Medienpädagogischer Forschungsverband Südwest, 2019, 2020).

Durch die Nutzung von digitalen Medien kann die Vielfältigkeit der Kompetenztests erweitert werden, da mit digitalen Medien Bereiche der mathematischen und naturwissenschaftlichen Kompetenz ermittelt werden können, die mit papierbasierten Tests nicht erfassbar sind: So erlauben es die bisherigen Papier-Bleistift-Tests nicht, alle Anforderungen zu stellen, mit denen sich Schülerinnen und Schüler beim schulischen Lernen beschäftigen, sei es digital (z.B. Onlinelearnmaterialien mit Simulationen von naturwissenschaftlichen Phänomenen im Hausaufgabenkontext) oder auch nichtdigital (z.B. praktische Übungen im naturwissenschaftlichen Laborunterricht). Die Nutzung von digitalen Medien kann hier eine Angleichung von Testinhalten an den schulischen Alltag ermöglichen und somit einen wichtigen Fortschritt hinsichtlich der realitätsbezogenen Konstruktion der TIMSS-Tests erzielen. TIMSS 2019 bietet somit die Möglichkeit, im Rahmen großangelegter Schulvergleichsstudien Testformate zu entwickeln, die stärker an die aktuelle Lebens- und Erfahrungswelt von Schülerinnen und Schülern angepasst sind (Mullis & Martin, 2017).

Durch die Umstellung der TIMSS-Testung auf digitale Medien wird außerdem relevant, wie vertraut die Schülerinnen und Schüler in den teilnehmenden Staaten mit diesen Medien sind. In den Bildungssystemen, in denen die Nutzung digitaler Medien schon seit Längerem selbstverständlich ist und in den Bildungssystemen, in denen aktuell erste Erfahrungen gesammelt werden, wird die Handhabung und Nutzung von digitalen Medien für die Schülerinnen und Schüler unterschiedliche Herausforderungen darstellen. Für die Entscheidung zur Nutzung digitaler Medien in TIMSS sprach und spricht die Erkenntnis, dass die Handhabung digitaler Medien eine notwendige Voraussetzung für eine sich wandelnde Gesellschaft darstellt, da sich abzeichnet, dass die Nutzung von digitalen Medien so selbstverständlich wie das Lesen von Printmedien ist.

Da diese Modusumstellung den Kindern zusätzliche Kompetenzen abverlangt und so Auswirkungen auf die eigentliche inhaltliche Bearbeitung der Mathematik- und Naturwissenschaftstests hat, kommt der Kontrolle dieses Übergangs von papierbasierten Tests zu Tests mit digitalen Medien eine besondere Aufmerksamkeit zu. Vorbereitet durch eine der Hauptstudie zwei Jahre vorgelegte Moduseffekt-Studie zur Abschätzung dieser Wirkungen in verschiedenen Bildungssystemen wird in der 2019 durchgeführten TIMSS-Hauptstudie eine Brückenstudie realisiert. In den Staaten, die 2019 digitale Medien beim TIMSS-Test einsetzen, wird neben dieser regulären Hauptstichprobe eine zusätzliche Stichprobe von Schülerinnen und Schülern ausschließlich im herkömmlichen papierbasierten Format getestet. Dieses Design lässt zum einen die Dokumentation des Trendverlaufs über die TIMSS-Zyklen hinweg und zum anderen auch die Untersuchung des Leistungsunterschieds zwischen papierbasierten und digitalen medienbasierten Tests zu.

Zur Wahrung größtmöglicher Transparenz hat sich das ISC entschieden, die sich aus den beiden Testmodi ergebenden Befunde für die Staaten, die in TIMSS 2019 digitale Testmedien genutzt und auch an der Brückenstudie teilgenommen haben, getrennt zu dokumentieren. Hierbei ist zu berücksichtigen, dass es bei dem vom ISC gewählten Vorgehen nicht um die staatenpezifisch optimierte Analyse des Übergangs von papierbasierten zu digitalen Testmedien geht, sondern um die optimierte Modellierung eines internationalen Trends unter Berücksichtigung von papierbasierten und digitalen Testmedien. Bei staatenspezifi-

fischen Analysen des Wechsels in den Testmodi lassen sich die nationalen Effekte bei der Bearbeitung herkömmlicher (papierbasierter) und digitaler Testformate empirisch klarer herausarbeiten. Diese Modelle erlauben jedoch keinen Vergleich des Übergangs von papierbasierten zu digitalen Testformaten zwischen den Staaten. In dem vorliegenden Band wird beiden Analysestrategien (optimierter internationaler Vergleich und optimierte nationale Analyse des Testmodus) gefolgt, da sich diese im Wesentlichen nicht widersprechen, sondern zu den jeweils spezifischen Fragestellungen optimierte Antworten bieten (siehe Kapitel 5 in diesem Band).

2.3.4 Forschungsfragen in TIMSS 2019

Aus dem aktuellen Forschungsstand und den Erfahrungen aus vorangegangenen TIMSS-Zyklen ergeben sich für TIMSS 2019 folgende zentrale Forschungsfragen:

- Welche Kompetenzniveaus zeigen Schülerinnen und Schüler in den Bereichen Mathematik und Naturwissenschaften am Ende der vierten Jahrgangsstufe in Deutschland im Jahr 2019? Wie lassen sich die Ergebnisse im internationalen Vergleich bewerten? Wie haben sich die Ergebnisse seit 2007, 2011 und 2015 verändert? Welche Trends zeichnen sich ab?
- Wie lassen sich die TIMSS-Leistungskennwerte auf den Kompetenzstufen einordnen? Wie groß sind die Gruppen der auffällig leistungsschwachen und leistungsstarken Schülerinnen und Schüler? Wie unterscheiden sich die Ergebnisse von TIMSS 2007, 2011, 2015 und 2019?
- Welche Ergebnisse erzielen die Viertklässlerinnen und Viertklässler in Deutschland in den domänenspezifischen Inhaltsbereichen und den kognitiven Anforderungsbereichen der TIMS-Studie?
- Welche Bedeutung haben die individuellen Lernvoraussetzungen und Kontextfaktoren für Kompetenzen in den Bereichen Mathematik und Naturwissenschaften? Welche Veränderungen von Lehr- und Lernbedingungen lassen sich seit 2007 beobachten?
- Wie ist die Unterrichtsqualität in den Fächern Mathematik und Sachunterricht in Deutschland ausgeprägt? Welche Veränderungen sind im Trend festzustellen? Welche Zusammenhänge zeigen sich zwischen der Unterrichtsqualität in den jeweiligen Fächern und der mathematischen beziehungsweise naturwissenschaftlichen Kompetenz? Gibt es eine differenzielle Wirksamkeit von Unterrichtsqualitätsmerkmalen für unterschiedliche Schülerinnen- und Schülergruppen?

2.4 Anlage von TIMSS

Um großangelegte Schulvergleichsstudien durchführbar, für viele Staaten finanzierbar und darüber hinaus auch die Inanspruchnahme im Zusammenhang mit der Testdurchführung in den Schulen so gering wie möglich zu halten, haben sich insbesondere zwei Herangehensweisen bei der Konzeption der Studie bewährt: Zum einen wird eine Schülerinnen- und Schülerstichprobe für jedes Bildungssystem gezogen, die groß genug ist, um die Zielsetzung des internationalen Vergleichs zu erfüllen. Zum anderen wird eine systematisch entwickelte Auswahl von Testaufgaben zur Feststellung mathematischer und naturwissen-

schaftlicher Kompetenzen der Viertklässlerinnen und Viertklässler eingesetzt. Somit werden sowohl in Bezug auf die Schülerinnen und Schüler als auch bei den Testaufgaben jeweils Stichproben aus deutlich größeren Grundgesamtheiten ausgewählt. Den Vorteil, den eine nach wissenschaftlichen Kriterien geplante und durchgeführte Untersuchung wie TIMSS bietet, ist, dass die mit den Stichprobenziehungen verbundenen Unsicherheiten (Stichprobenfehler und Messfehler) bei der Generalisierung der Befunde empirisch abgeschätzt und dokumentiert werden können. Im Folgenden werden die beiden eben skizzierten Herangehensweisen nacheinander näher beleuchtet.

2.4.1 Sampling

Definition der Zielpopulation

Da sich die Klassenstufenbeschreibungen in den beteiligten Staaten unterscheiden, wurde durch die internationale Studienleitung eine Definition der Zielpopulation für die Untersuchungen vorgegeben, die sich an der Anzahl der formalen Unterrichtsjahre gemäß der *International Standard Classification of Education* (ISCED) nach der Definition des UNECSO-Instituts von 2011 orientiert (UNESCO Institute for Statistics, 2012). Elementar- und vorschulische Betreuungen zählen hierbei nicht mit. Die in TIMSS realisierte Grundschuluntersuchung nimmt die Kinder in den Blick, die im vierten regulären Unterrichtsjahr sind – dies entspricht in Deutschland der vierten Klassenstufe. Zwar werden im Rahmen von TIMSS auch Schülerinnen und Schüler der achten Klasse befragt, da sich Deutschland jedoch nur an der Grundschuluntersuchung beteiligt, werden die Aspekte von TIMSS, die sich auf die Untersuchung der Sekundarstufe beziehen, nicht näher beleuchtet.

Stichprobenziehung

Eine wichtige Voraussetzung für generalisierte Aussagen über Bildungssysteme ist, dass die in den teilnehmenden Staaten erhobenen Daten repräsentativ für die Zielpopulationen der Staaten sind. Daher kommt der Stichprobenziehung eine besondere Bedeutung zu. Je nach Besonderheit der nationalen Bildungssysteme werden die Stichprobenverfahren entsprechend angepasst – in Deutschland betrifft dies zum Beispiel die Struktur des föderalen Bildungssystems und der 16 beteiligten Länder der Bundesrepublik Deutschland, für die jeweils individuelle Stichproben gezogen und dann in einem Gesamtdatensatz zusammengeführt werden. Diese beiden Besonderheiten führen in Deutschland dazu, dass die Untersuchung in 16 verschiedenen Ländern beantragt und genehmigt werden muss, um eine repräsentative Stichprobe für Deutschland zu generieren. Daraus resultiert gleichzeitig, dass sich die Verpflichtungsgrade zur Teilnahme an TIMSS 2019 zwischen den Ländern in der Bundesrepublik Deutschland unterscheiden können.

In Tabelle 2.2 sind die Verpflichtungsgrade für die TIMSS-Untersuchungen nach den Ländern der Bundesrepublik Deutschland aufgeführt. Dieser ist zu entnehmen, dass sich die Kultusministerinnen und Kultusminister der Länder der Bundesrepublik Deutschland auf die verpflichtende Teilnahme an den Kompetenztests verständigt haben, sodass die mit TIMSS erfassten Kompetenzstände als repräsentativ für Deutschland gelten können. Darüber hinaus ist Tabelle 2.2 auch zu entnehmen, dass die verpflichtende beziehungsweise freiwillige Teilnahme bei der Beantwortung der Hintergrundfragebögen zwischen den Ländern variiert.

Tabelle 2.2: Länderspezifische Modalitäten der Teilnahme an TIMSS 2007, 2011, 2015 und 2019

Länder	Verpflichtungsgrade der Teilnahme bei Erhebungen an öffentlichen Schulen				
	Leistungstest	Schülerfragebogen	Lehrerfragebogen	Schulleitungsfragebogen	Elternfragebogen
Baden-Württemberg	verpflichtend	verpflichtend ¹	freiwillig	freiwillig	freiwillig
Bayern	verpflichtend	freiwillig	freiwillig ²	freiwillig ²	freiwillig
Berlin	verpflichtend	verpflichtend ³	teilverpflichtend ³	teilverpflichtend ³	freiwillig
Brandenburg	verpflichtend	verpflichtend	verpflichtend	verpflichtend	freiwillig
Bremen	verpflichtend	verpflichtend ⁴	teilverpflichtend	teilverpflichtend	freiwillig
Hamburg	verpflichtend	freiwillig	freiwillig ⁵	teilverpflichtend	freiwillig
Hessen	verpflichtend	verpflichtend	teilverpflichtend	teilverpflichtend	freiwillig
Mecklenburg-Vorpommern	verpflichtend	verpflichtend ⁶	verpflichtend	verpflichtend	freiwillig
Niedersachsen	verpflichtend	verpflichtend	verpflichtend	verpflichtend	freiwillig
Nordrhein-Westfalen	verpflichtend	freiwillig	verpflichtend	verpflichtend	freiwillig
Rheinland-Pfalz	verpflichtend	freiwillig	teilverpflichtend	teilverpflichtend	freiwillig
Saarland	verpflichtend	freiwillig	teilverpflichtend ⁷	teilverpflichtend ⁷	freiwillig
Sachsen	verpflichtend	freiwillig	freiwillig	freiwillig	freiwillig
Sachsen-Anhalt	verpflichtend	verpflichtend	verpflichtend	verpflichtend	freiwillig
Schleswig-Holstein	verpflichtend	freiwillig	freiwillig ⁸	freiwillig ⁸	freiwillig
Thüringen	verpflichtend	verpflichtend	verpflichtend	verpflichtend	freiwillig

1 = In Baden-Württemberg war im Jahr 2015 die Teilnahme am Schülerfragebogen noch freiwillig.

2 = In Bayern war im Jahr 2007 die Teilnahme am Lehrer- und Schulleitungsfragebogen noch verpflichtend.

3 = In Berlin war in den Jahren 2007 und 2011 die Teilnahme am Schüler-, Lehrer- und Schulleitungsfragebogen noch verpflichtend. Im Jahr 2015 war die Teilnahme am Schülerfragebogen teilverpflichtend.

4 = In Bremen war in den Jahren 2007 und 2011 die Teilnahme am Schülerfragebogen noch freiwillig.

5 = In Hamburg war in den Jahren 2007 und 2011 die Teilnahme am Lehrerfragebogen noch teilverpflichtend.

6 = In Mecklenburg-Vorpommern war im Jahr 2007 die Teilnahme am Schülerfragebogen noch freiwillig.

7 = In Saarland war in den Jahren 2007 und 2011 die Teilnahme am Lehrer- und Schulleitungsfragebogen noch verpflichtend.

8 = In Schleswig-Holstein war in den Jahren 2007 und 2011 die Teilnahme am Lehrer- und Schulleitungsfragebogen noch teilverpflichtend.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Bei der Stichprobenziehung wurde in Deutschland mehrstufig vorgegangen: Zunächst wurden innerhalb Deutschlands je Land Schulen per Zufallsverfahren unter Wahrung der Repräsentativität unterschiedlicher Schulgrößen (PPS: *probability proportional by size* kombiniert mit dem Auswahlverfahren *random start – equal interval*, Särndal, Swensson & Wretman, 1992; Martin, Mullis & Hooper, 2016) ausgewählt. Wenn es innerhalb der Schulen mehr als eine vierte Klasse gab, wurde eine davon per Zufall ausgewählt. Unabhängig vom spezifischen nationalen Design war durch die internationale Studienleitung vorgegeben, dass pro Staat mindestens 150 Schulen (entspricht ca. 4000 Schülerinnen und Schülern) für die Untersuchung gezogen werden sollten. Hatte sich ein Staat für die Teilnahme an eTIMSS entschieden, war die Erweiterung um eine sogenannte Brückenstudie (bei der zusätzlich Kinder papierbasiert getestet wurden) mit mindestens 1500 weiteren Schülerinnen und Schülern obligatorisch.

Da sich die Ziehungswahrscheinlichkeiten der einzelnen Schülerinnen und Schüler je Land oder Schule unterscheiden können und damit die angestrebte repräsentative Stichprobe leicht verzerrt werden würde, wird für jede Schülerin und jeden Schüler für die späteren Analysen ein Stichproben-Korrekturfaktor (Gewicht) berechnet. Gewichtete Analysen sind relevant, um die in der Grundgesamtheit bestehenden Verteilungen auch tatsächlich bei der Auswertung abzu-

Tabelle 2.3: Beteiligungszahlen und zentrale Stichprobencharakteristika in TIMSS 2007, 2011, 2015 und 2019

	TIMSS 2007	TIMSS 2011	TIMSS 2015	TIMSS 2019	
				eTIMSS	Brücken- studie
Schulen					
Anzahl	246	198	204	203	78
Beteiligung am Schulleitungsfragebogen (%)	95	95	90	93	96
Klassen					
Anzahl	246	198	204	211	80
Lehrkräfte					
Anzahl (Mathematik/Sachunterricht)	240/231	208/204	224/226	216/218	80/82
Beteiligung am Lehrerfragebogen (%)	95	96	90	91	95
Schülerinnen und Schüler					
Anzahl	5200	3998	3948	3437	1505
Durchschnittsalter	10.4	10.4	10.4	10.4	10.4
Mädchen (%)	49	49	49	50	52
Kinder aus armutsgefährdeten Familien (%)	34	25	29	22	21
Kinder mit Migrationshintergrund (%) (beide Elternteile im Ausland geboren)	17	16	20 ¹	22	20
Testteilnahme					
Anzahl (Mathematik/Naturwissenschaften)	5194/5193	3995/3998	3948/3948	3437/3437	1505/1505
Teilnahmequote (%)	96	95	95	97	96
Beteiligung am Schülerfragebogen (%)	87	90	85	87	83
Beteiligung am Elternfragebogen (%)	77	80	63	67	65

1 = Abweichungen zu den Ergebnissen von TIMSS 2015 (Wendt, Schwippert & Stubbe, 2016) ergeben sich aus einem veränderten Umgang mit fehlenden Werten.

bilden. Auch wenn länderspezifische Stichproben gezogen wurden, so sind diese – bis auf für das bevölkerungsreichste Land der Bundesrepublik Deutschland Nordrhein-Westfalen – zu klein, um empirisch gehaltvolle Aussagen im Rahmen von Ländervergleichen vornehmen zu können. Solche Vergleiche sind aktuell der hierfür ausgelegten Bildungstrendstudie des *Instituts für Qualitätsentwicklung im Bildungswesen* (IQB) in Berlin vorbehalten, die jedoch anders als TIMSS keine internationalen Vergleiche zulassen.

Optimalerweise nehmen bei der Durchführung der TIMSS-Untersuchung auch tatsächlich alle gezogenen Schülerinnen und Schüler teil (volle Ausschöpfung der Stichprobe). Abweichungen können sich dann ergeben, wenn zum Beispiel eine Schule kurzfristig geschlossen wurde oder einzelne Kinder am Tag der Untersuchung krank sind. Solchen Ausfällen wird im Vorwege durch die Auswahl von Ersatzschulen beziehungsweise durch die Möglichkeit der Nachtestung zu einem späteren Zeitpunkt begegnet. Geringe Abweichungen vom Stichprobenplan

sind in einem vorgegebenen Rahmen akzeptabel, wenn die Repräsentativität der Untersuchung nicht gefährdet wird (Martin et al., 2016). Sind diese Abweichungen jedoch zu groß, ist dies bei den Analysen und der Berichtslegung entsprechend zu berücksichtigen. Tabelle 2.3 sind die in Deutschland erzielten Ausschöpfungsquoten bei der Durchführung der TIMSS-2019-Untersuchung zu entnehmen. Alle Abweichungen vom internationalen Idealwert (= 100 %) liegen im Bereich der akzeptablen Abweichungen. Bei der Anzahl der Schulen ist zu beachten, dass 36 Schulen sowohl an der eTIMSS- als auch der Brückenstudie (mit unterschiedlichen Klassen) teilgenommen haben. Hinzu kommen 167 Schulen, die nur an eTIMSS und 42, die nur an der Brückenstudie beteiligt waren. Dass die Anzahl der Klassen der Schulen übersteigt, obwohl ursprünglich nur eine Klasse je Schule gezogen werden sollte, liegt daran, dass bei besonders kleinen Klassen weitere Parallelklassen hinzugezogen wurden, um den notwendigen Stichprobenumfang zu erreichen.

Abweichungen von den durch die internationale Studienleitung definierten Stichprobenumfängen werden nachvollzogen und dokumentiert. Diese Dokumentation ist notwendig, damit bei vertiefenden Vergleichen von Staaten nur solche einander gegenübergestellt werden, die tatsächlich aufgrund gleicher Teilnahmebedingungen vergleichbar sind. Abweichungen vom international vorgegebenen Stichprobenplan sind in diesem Band jeweils in den Abbildungen dokumentiert, sodass diese nachvollziehbar sind und bei der Interpretation der Befunde berücksichtigt werden können. Nachfolgend sind die Besonderheiten der Teilnahmebedingungen in Form von Fußnoten aufgeführt, die in den Auswertungen im vorliegenden Band berichtet werden. Diese wurden bereits in TIMSS 2011 und 2015 auf gleiche Weise dokumentiert.

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

Die Fußnote 4 gilt nur für die Studienzyklen 2011, 2015 und 2019. Wenn notwendig wird sie zudem in Fußnote 4 (nicht skalierbare Leistungswerte in Mathematik) und Fußnote 5 (nicht skalierbare Leistungswerte in den Naturwissenschaften) differenziert. Darüber hinaus gibt es noch drei weitere Fußnoten. Die Fußnote 6 verweist darauf, dass die Testung zeitlich verzögert durchgeführt wurde. Für die Kennzeichnung der Teilnahme von Staaten beziehungsweise *Benchmark*-Teilnehmern an *TIMSS Numeracy* (in 2019: *Less Difficult Mathematics*) wurde in TIMSS 2015 die Fußnote 7 eingeführt. Besonderheiten in den Trendvergleichen mit den Zyklen 2007, 2011 und 2015 werden durch die Fußnote 8 gekennzeichnet (siehe Anhang).

Darüber hinaus wird durch die Kursivschreibung von Staaten- und Regionennamen kenntlich gemacht, dass Besonderheiten in den Erhebungsbedingungen vorliegen, sodass eine Interpretation der Ergebnisse nur unter deren Berücksichtigung erfolgen sollte. Erläuternd ist dazu unter den Abbildungen

die Fußnote „Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.“ ergänzt (siehe Anhang).³

Neben der Ausschöpfungsquote ist im Stichprobenverfahren eine Möglichkeit angelegt, bestimmte Kinder nicht in die Untersuchung mit einzubeziehen. Durch die Möglichkeit der Nicht-Teilnahme wird sichergestellt, dass Schülerinnen und Schüler, die aufgrund spezieller schulorganisatorischer Gründe nicht an der Testung teilnehmen können oder die der Testsituation aufgrund von Behinderungen oder Beeinträchtigungen nicht gewachsen sind oder nicht über hinreichende Sprachkenntnisse verfügen, von der Testung befreit werden beziehungsweise bei der Testauswertung nicht berücksichtigt werden können. Diese Testbefreiungsquoten unterscheiden sich aufgrund verschiedener geografischer Herausforderungen und unterschiedlicher Traditionen integrativer und inklusiver Beschulung zwischen den Staaten.

Die Möglichkeit der Testbefreiung führt immer wieder zu der Frage, ob sich in Deutschland das Ergebnis der Kompetenztests verändern würde, wenn sich eine höhere Anzahl von Schülerinnen und Schülern nicht an der Untersuchung beteiligen würde. Bei dieser Frage schwingt das Misstrauen mit, dass die beobachteten Stichprobenausfälle nicht aus den zuvor genannten Gründen erfolgen würden, sondern dass zum Beispiel leistungsschwächere Schülerinnen und Schüler ohne berechtigte Befreiungsgründe nicht mit in die Untersuchung einbezogen würden, damit die Staaten im Gesamtergebnis besser dastünden. Auch wenn solche systematischen Stichprobenmanipulationen durch die fundierte Kontrolle der internationalen Studienleitung bei der Untersuchungsdurchführung praktisch ausgeschlossen werden können, wurde der hypothetische Effekt auf den Rangplatz Deutschlands untersucht. In Tabelle 2.4 sind die Mittelwerte für Deutschland dargestellt, die sich ergeben würden, wenn leistungsschwächere Schülerinnen und Schüler ausgeschlossen werden würden, bis zum Erreichen einer fiktiv höheren Ausschlussquote von 12.8 Prozent (Singapur), 7.2 Prozent (USA), und 5.4 Prozent (Schweden) für die Gesamtskala Mathematik und in Tabelle 2.5 für die Gesamtskala Naturwissenschaften. Als Vergleichsstaaten wurden in den Tabellen jeweils die Bildungssysteme mit aufgenommen, die signifikant höhere oder nicht signifikant von Deutschland abweichende Kompetenzwerte aufweisen. Die *Benchmark*-Teilnehmer werden dabei nicht berücksichtigt.

Dabei zeigen sich sowohl in Mathematik als auch in den Naturwissenschaften in 2019 zwar numerische Verschiebungen und dadurch auch kleinere Veränderungen in der Rangreihe der berichteten Staaten. Insbesondere der Unterschied zu den Staaten, deren Schülerinnen und Schüler substanziell höhere Kompetenzen aufwiesen, wie Singapur, Hongkong, die Republik Korea, Taiwan, Japan und die Russische Föderation, bleibt aber auf signifikantem Niveau erhalten. Veränderungen ergeben sich lediglich bei den mittleren Rangplätzen der Staatenmittelwerte. Diese sind jedoch als nicht aussagekräftig einzustufen. Die gleichen Befunde zeigten sich auch in TIMSS 2015.

3 Die hier vorgestellten Ergebnisse basieren auf dem Stand vom 1.11.2020. Die Befunde für die Inhaltsbereiche in Mathematik sind für Chile, die kognitiven Anforderungsbereiche in den Naturwissenschaften für die Republik Nordmazedonien und den Oman als vorläufig anzusehen.

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

Teilnehmer	Ausschlussquote		(SE)	Signifikanz zu Deutschland bei einer Ausschlussquote in Prozent von			
	in %	M		3.9	5.4	7.2	12.8
² <i>Singapur</i>	12.8	625	(3.9)	+	+	+	+
³ <i>Hongkong</i>	3.5	602	(3.3)	+	+	+	+
Republik Korea (Südkorea)	2.3	600	(2.2)	+	+	+	+
Taiwan	2.0	599	(1.9)	+	+	+	+
³ <i>Japan</i>	2.2	593	(1.8)	+	+	+	+
² <i>Russische Föderation</i>	6.3	567	(3.3)	+	+	+	+
³ <i>Nordirland</i>	2.8	566	(2.7)	+	+	+	+
¹² <i>England</i>	5.8	556	(3.0)	+	+	+	+
Irland	3.0	548	(2.5)	+	+	+	+
² <i>Lettland</i>	6.9	546	(2.6)	+	+	+	+
¹³ <i>Norwegen (5. Jgst.)</i>	4.7	543	(2.2)	+	+	+	+
² <i>Litauen</i>	6.7	542	(2.8)	+	+	+	+
² <i>Österreich</i>	5.4	539	(2.0)	+	+	+	+
³ <i>Niederlande</i>	3.5	538	(2.2)	+	+	+	+
²³ <i>USA</i>	7.2	535	(2.5)	+	+	+	+
Tschechische Republik	4.7	533	(2.5)	+	+	+	+
Deutschland ^A	12.8	533	(1.9)	+	+	+	+
³ <i>Belgien (Fläm. Gem.)</i>	3.0	532	(1.9)	+	+	+	+
Zypern	4.6	532	(2.9)	+	+	+	+
Finnland	3.3	532	(2.3)	+	+	+	+
Deutschland ^A	7.2	526	(2.0)				-
² <i>Portugal</i>	7.8	525	(2.6)				-
³ <i>Dänemark</i>	3.1	525	(1.9)				-
Deutschland ^A	5.4	523	(2.1)				-
Ungarn	4.1	523	(2.6)				-
¹² <i>Türkei (5. Jgst.)</i>	7.0	523	(4.4)				-
² <i>Schweden</i>	5.4	521	(2.8)				-
Deutschland	3.9	521	(2.3)				-
Polen	3.1	520	(2.7)				-
Australien	4.8	516	(2.8)		-	-	-
Aserbaidshan	2.6	515	(2.7)		-	-	-
Bulgarien	3.4	515	(4.3)			-	-
Italien	4.9	515	(2.4)		-	-	-

+ = Statistisch signifikant positive Abweichungen ($p \leq .05$) zu dem für Deutschland ermittelten Leistungswert unter Berücksichtigung der in der Spaltenüberschrift angegebenen Ausschlussquote

- = Statistisch signifikant negative Abweichungen ($p \leq .05$) zu dem für Deutschland ermittelten Leistungswert unter Berücksichtigung der in der Spaltenüberschrift angegebenen Ausschlussquote

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

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

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

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

A = Berechnung für Deutschland nach einem höheren Ausschluss von Schülerinnen und Schülern mit geringen Kompetenzwerten

2.4.2 Entwicklung der Befragungsinstrumente in TIMSS

In TIMSS werden zwei Arten von Befragungsinstrumenten eingesetzt: Hintergrundfragebögen, die Merkmale der Schülerinnen und Schüler, der Erziehungsberechtigten, der Lehrkräfte und der Schulen erfragen, und Kompetenztests, die die mathematischen und naturwissenschaftlichen Fähigkeiten der Schülerinnen und Schüler erfassen. Für beide Instrumente gilt, dass Informationen, die im Trend dokumentiert werden sollen, auch entsprechend zu erfassen sind. Das be-

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

Teilnehmer	Ausschlussquote			Signifikanz zu Deutschland bei einer Ausschlussquote in Prozent von			
	in %	M	(SE)	3.9	5.4	7.2	12.8
² <i>Singapur</i>	12.8	595	(3.4)	+	+	+	+
Republik Korea (Südkorea)	2.3	588	(2.1)	+	+	+	+
² <i>Russische Föderation</i>	6.3	567	(3.0)	+	+	+	+
³ <i>Japan</i>	2.2	562	(1.8)	+	+	+	+
Taiwan	2.0	558	(1.8)	+	+	+	+
Finnland	3.3	555	(2.6)	+	+	+	+
² <i>Lettland</i>	6.9	542	(2.4)	+	+	+	+
^{1 3} <i>Norwegen (5. Jgst.)</i>	4.7	539	(2.2)	+	+	+	+
^{2 3} <i>USA</i>	7.2	539	(2.7)	+	+	+	+
² <i>Litauen</i>	6.7	538	(2.5)	+	+	+	+
² <i>Schweden</i>	5.4	537	(3.3)	+	+	+	
¹² <i>England</i>	5.8	537	(2.7)	+	+	+	
Tschechische Republik	4.7	534	(2.6)	+	+	+	
Australien	4.8	533	(2.4)	+	+	+	
Deutschland^A	12.8	531	(2.0)	+	+	+	
³ <i>Hongkong</i>	3.5	531	(3.3)	+	+		
Polen	3.1	531	(2.6)	+	+	+	
Ungarn	4.1	529	(2.7)	+	+		
Irland	3.0	528	(3.2)	+			
¹² <i>Türkei (5. Jgst.)</i>	7.0	526	(4.2)				
Deutschland^A	7.2	524	(2.0)				-
Kroatien	4.2	524	(2.2)				-
² <i>Kanada</i>	7.0	523	(1.9)				-
³ <i>Dänemark</i>	3.1	522	(2.4)				-
² <i>Österreich</i>	5.4	522	(2.6)				-
Bulgarien	3.4	521	(4.9)				-
Deutschland^A	5.4	521	(2.0)				-
² <i>Slowakei</i>	5.5	521	(3.7)				-
³ <i>Nordirland</i>	2.8	518	(2.3)				-
³ <i>Niederlande</i>	3.5	518	(2.9)				-
Deutschland	3.9	518	(2.2)				-
² <i>Serbien</i>	8.2	517	(3.5)				-
Zypern	4.6	511	(3.0)		-	-	-

+ = Statistisch signifikant positive Abweichungen ($p \leq .05$) zu dem für Deutschland ermittelten Leistungswert unter Berücksichtigung der in der Spaltenüberschrift angegebenen Ausschlussquote

- = Statistisch signifikant negative Abweichungen ($p \leq .05$) zu dem für Deutschland ermittelten Leistungswert unter Berücksichtigung der in der Spaltenüberschrift angegebenen Ausschlussquote

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

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

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

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

A = Berechnung für Deutschland nach einem höheren Ausschluss von Schülerinnen und Schülern mit geringen Kompetenzwerten

deutet, dass die Fragen, die in den bisherigen TIMSS-Zyklen gestellt und berichtet wurden, in identischer Weise auch in TIMSS 2019 zu erfassen waren. Dort, wo aufgrund deutlicher Entwicklungen eine Anpassung der Fragen notwendig wurde, um für 2019 überhaupt valide Antworten erhalten zu können, wurde die Anpassung vorgenommen – wissentlich, dass dann mögliche Trendaussagen nur noch eingeschränkt oder nicht mehr möglich sind.

2.4.3 Konzeption der TIMSS-Tests

Die TIMSS-Untersuchung hat zum Ziel, die Kompetenzen von Schülerinnen und Schülern in den Domänen Mathematik und Naturwissenschaften, die in deutschen Grundschulen im Sachunterricht⁴ unterrichtet werden, zu erfassen. Für beide Domänen ist somit eine Auswahl von Testaufgaben bereitzuhalten, welche die entsprechenden Kompetenzen der Grundschülerinnen und Grundschüler möglichst angemessen erfassen.

Tabelle 2.6 gibt Auskunft darüber, wie viele Aufgaben in den TIMSS-Untersuchungen von 2007 bis 2019 genutzt wurden. Dabei wurde nach einem Zyklus von TIMSS jeweils ein Teil der Aufgaben (in 2019 jeweils zwei Testblöcke pro Domäne) veröffentlicht und ein größerer Teil für die Anwendung im nächsten Zyklus zunächst unter Verschluss gehalten. Die wiederholte Nutzung von Aufgaben in verschiedenen TIMSS-Zyklen erlaubt den direkten Vergleich der beiden Untersuchungen miteinander – sie bilden also die notwendige Voraussetzung für Trendanalysen. Dieses System wurde über alle Zyklen hinweg bis 2019 so realisiert. Somit können die Testergebnisse in Mathematik und den Naturwissenschaften von 2007 über 2011 und 2015 bis 2019 miteinander verglichen werden.

Tabelle 2.6: Anzahl von Testaufgaben für Mathematik und Naturwissenschaften in TIMSS 2007, 2011, 2015 und 2019

Domänen	2007	2011	2015	2019
	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>
Mathematik	179	175	169	171
Naturwissenschaften	174	168	168	169
Gesamt	353	343	337	340

Die Werte geben die Anzahl der Items an, die für die internationale Skalierung der Leistungsergebnisse berücksichtigt wurden. In einzelnen Staaten und Zyklen standen zum Teil geringfügig mehr Items für nationale Auswertungen zur Verfügung.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Da die Bearbeitung aller in TIMSS eingesetzten Aufgaben durch alle Schülerinnen und Schüler zu zeitaufwendig wäre und darüber hinaus auch über der Belastbarkeit der Grundschülerinnen und Grundschüler liegen würde, wurden die Aufgaben in kleinere Aufgabenblöcke aufgeteilt, von denen die Kinder jeweils immer nur eine Auswahl bearbeiten. Durch die systematische Zusammenstellung und Durchmischung dieser Aufgabenblöcke kann sichergestellt werden, dass der Mathematik- und Naturwissenschaftstest in TIMSS aufgrund der hohen Zahl an Aufgaben insgesamt hinreichend valide ist, die einzelnen Schülerinnen und Schüler jedoch nicht überlastet werden und schließlich durch die Testung von ausreichend vielen Schülerinnen und Schülern auch Aussagen über das Bildungssystem vorgenommen werden können. Die Einschränkung, die man bei dieser Art von Systemmonitoringstudien in Kauf nimmt, ist, dass keine differenzierten Befunde für die einzelnen Kinder – im Sinne einer Individualdiagnostik – vorliegen, dass jedoch die Summe aller Informationen gut geeignet ist, Aussagen

4 In Bayern und Thüringen wird das Fach als Heimat- und Sachunterricht beziehungsweise Heimat- und Sachkunde bezeichnet. Der Einfachheit halber wird in diesem Kapitel von dem Fach Sachunterricht gesprochen.

über die nationale Grundgesamtheit der Untersuchung treffen zu können (Populationsschätzer). Diese sind auch dafür geeignet, Analysen zu Lehr- und Lernbedingungen vorzunehmen und Aussagen auf Ebene der Schule und des Unterrichts machen zu können.

Obwohl die Schülerinnen und Schüler jeweils unterschiedliche Teile des TIMSS-Gesamttests bearbeiten, ist es durch die systematische Variation der Testblöcke möglich, für jede beziehungsweise jeden einen entsprechenden Kompetenzwert zu ermitteln. Etwas technischer ausgedrückt werden in dem rotierten *Multi-Matrix-Design* anhand von Modellen der *Item-Response-Theory* basierend auf den Tests und den Hintergrunddaten die Schülerinnen- und Schülerkompetenzen für die beiden Domänen bestimmt. Hierbei wird bei *Multiple-Choice-Fragen* das *3-PL-Model* angewendet, bei offenen Fragen mit einer richtigen Antwortmöglichkeit das *2-PL-* und bei mehr als zwei möglichen Antwortabstufungen das *Partial-Credit-Model* genutzt (siehe Kapitel 5 in diesem Band).

Bei der Bestimmung der Leistungsdaten werden auch Informationen aus den Hintergrundfragebögen hinzugezogen, um unter Berücksichtigung der Aufgabenauswahl (Testblöcke) gemeinsam mit den Hintergrundinformationen die bestmöglichen Schätzer für die Kompetenzwerte in der Grundgesamtheit (Population) zu erhalten. Um den durch das Untersuchungsdesign notwendigen stichprobenbedingten Einschränkungen bei der Befragung der Kinder Rechnung zu tragen und angemessene Schätzer für Aussagen über das Bildungssystem vornehmen zu können, werden je teilnehmendem Kind und Domäne fünf Kompetenzwerte (*plausible values*) bestimmt, die im Rahmen simultaner Berechnungen zusammengeführt werden und die bestmöglichen Schätzer für Aussagen über das Bildungssystem erlauben. Durch die gemeinsame Bestimmung der Aufgabenschwierigkeiten in den vorangegangenen TIMSS-Zyklen können darüber hinaus auch vergleichende Analysen über die Bildungssysteme im Trend vorgenommen werden.

2.4.4 Konzeption der TIMSS- und eTIMSS-Testformate in TIMSS 2019

Bei der Entwicklung von Tests besteht ein großes Bestreben darin, den Schülerinnen und Schülern Aufgaben vorzulegen, die idealerweise auch genau das erfassen (messen), was sie erfassen sollen. Bei TIMSS betrifft dies insbesondere die mathematischen und naturwissenschaftlichen Kompetenzen von Grundschülerinnen und Grundschulern. Aufgaben, die aber exakt nur diese kognitiven Kompetenzen für die Lösung voraussetzen, sind so idealtypisch nicht zu entwickeln. Bei der Bearbeitung der Aufgaben sind immer auch andere kognitive, motivationale, affektive oder auch technische Ansprüche zu erfüllen. Bei Papier-Bleistift-Tests müssen die Kinder lesen können, sie müssen den Stift führen können und sie müssen sich generell auch der Aufgabenbearbeitung selber stellen. Da im Allgemeinen diese notwendigen Voraussetzungen für die fachliche Lösung der Aufgaben als weit verbreitet vorausgesetzt werden können, erfüllen die Mathematik- und Naturwissenschaftstests bei TIMSS ihren Zweck. Die Aufgaben sind dabei so gestaltet, dass – wie auch sonst im Unterricht – ein einfaches Leseverständnis ausreicht, die Aufgaben zu verstehen, dass bei der schriftlichen Beantwortung der fachliche Inhalt und nicht die richtige Schreibweise gewertet wird und dass die Kinder im Rahmen der Testinstruktion für die Mitarbeit motiviert werden. Daher darf – auch vor dem Hintergrund der langjährigen Erfahrungen bei der Entwicklung und Durchführung – bei TIMSS von der vali-

den Feststellung von mathematischen und naturwissenschaftlichen Kompetenzen ausgegangen werden.

2.4.5 Konzeption der Hintergrundfragebögen

Anspruch von TIMSS ist es, international vergleichbare Informationen aus verschiedenen Staaten bereitzustellen. Deshalb gibt es für alle teilnehmenden Staaten einen gemeinsamen internationalen Teil in den Hintergrundfragebögen. Dieser wird vom ISC vorgegeben, um die Erfassung wichtiger Informationen für die Trendweiterführung in allen Staaten zu sichern. Darüber hinaus bietet TIMSS auch die Möglichkeit, nationale Erweiterungen an die internationalen Teile anzuschließen. Diese Möglichkeit hat Deutschland seit der ersten Durchführung von TIMSS 2007 intensiv genutzt, um auf aktuelle Fragestellungen im Kontext der Schul- und Unterrichtsforschung zu reagieren und Merkmale und Wirkungen aktueller Herausforderungen, die sich durch gesellschaftliche Veränderungen ergeben haben, zu erfassen. So wurde eine in dieser Form einmalige Datenbasis geschaffen, die es erlaubt, jeweils spezifische Fragen zum deutschen Schulsystem sowohl im internationalen Vergleich als auch im nationalen Kontext zu beantworten (Hornberg, Bos, Buddeberg, Potthoff & Stubbe, 2007; Lankes et al., 2003; Schwippert, Stubbe & Wendt, 2015; Wendt, Bos, Kasper, Walzebug, Goy & Jusufi, 2016; Wendt, Tarelli, Bos, Frey & Vennemann, 2012).

Um einen möglichst umfassenden Eindruck von den Rahmenbedingungen des Lehrens und Lernens bei Grundschulkindern zu erhalten, werden unterschiedliche Informationsebenen durch die Befragungen adressiert. Hierbei orientiert sich die IEA an dem bereits beschriebenen Curriculum-Modell. Um abschätzen zu können und somit analysierbar zu machen, wie sich geplante Lehrangebote auf den Lernerfolg der Schülerinnen und Schüler auswirken, hat sich eine mehrstufige Betrachtung bewährt (Blömeke, 2014; Bos & Postlethwaite, 2005): Hierbei wird zunächst in den Blick genommen, welche Ziele im Bildungssystem durch ministerielle Vorgaben erreicht werden sollen (intendiertes Curriculum). Das implementierte Curriculum beschreibt, was durch eine individualisierte Vorbereitung durch Lehrende, unter der Nutzung von Unterrichtsmaterialien (insbesondere Lehrbücher, Forschungsberichte etc.) realisiert wird. Bedingt durch schulische Vorgaben und Rahmenbedingungen und individuelle Voraussetzungen beziehungsweise Dispositionen der Schülerinnen und Schüler werden die vermittelten Inhalte ganz individuell verarbeitet und somit auch unterschiedlich aufgenommen und als realisiertes Curriculum sichtbar.

So nachvollziehbar diese Wirkungskette vom intendierten über das implementierte bis hin zum realisierten Curriculum ineinandergreift, so schwierig ist es auch, die Wirkungen von Veränderungen am Anfang der Wirkungskette (im Sinne der Veränderung von Zielen durch die Ministerien) über unterschiedliche Rahmenbedingungen in den Schulen und die individuellen Umsetzungen durch die Lehrkräfte bis zu den konkreten Kompetenzen der Schülerinnen und Schüler nachzuvollziehen. In Tabelle 2.7 sind die Adressatinnen und Adressaten der Hintergrundfragebögen den jeweils ermittelten Informationen zugeordnet. Vor diesem Hintergrund werden bei der Konzeption der Hintergrundfragebögen zum Teil komplexe Abwägungsprozesse durchlaufen, um eine möglichst gute Balance zwischen Zielsetzung und aktuellen Herausforderungen zu finden.

Die Reaktion auf aktuelle Herausforderungen bedeutet in der Regel die Berücksichtigung neuer Fragen beziehungsweise Frageblöcke in den Hintergrund-

Tabelle 2.7: Hintergrundfragebögen und Charakterisierung der erfragten Informationen in TIMSS 2019

Fragebogen	Adressaten	Informationen	Informationsebene
Curriculum-Fragebogen	Nationale Studienleitung	Inhalte und Entwicklung der ministeriellen Vorgaben für den Unterricht	Intendiertes Curriculum
TIMSS-Test	Für TIMSS 2019 ausgewählte Schülerinnen und Schüler	Kompetenzen in Mathematik und in den Naturwissenschaften	Realisiertes Curriculum
Schülerfragebogen	Für TIMSS 2019 ausgewählte Schülerinnen und Schüler	Selbst- und fachbezogene Einstellungen, bildungsrelevante Ressourcen, individuelles inner- und außerschulisches Lernverhalten	
Elternfragebogen	Erziehungsberechtigte der ausgewählten Schülerinnen und Schüler	Häusliche Lernumgebung und Ressourcen, Erziehungs- und Unterstützungsverhalten, familiäre Merkmale	
Lehrerfragebogen	Lehrkräfte der ausgewählten Schülerinnen und Schüler	Unterrichtsqualität und -gestaltung, Qualifizierung und Fortbildung	Implementiertes Curriculum
Schulleitungsfragebogen	Schulleitungen der ausgewählten Schulen	Strukturelle, organisatorische und personelle Rahmenbedingungen der Schule	Implementiertes Curriculum

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

fragebögen, für deren Beantwortung aber nur eine begrenzte Zeit zur Verfügung steht. Um die Befragten und insbesondere die Kinder dadurch zeitlich nicht über Gebühr zu belasten, müssen, um neue Fragen aufnehmen zu können, ältere Fragen gestrichen werden, auch wenn das das Ende eines mit diesen Fragen bisher dokumentierten Trends bedeutet. Die hierzu notwendigen Entscheidungen wurden im deutschen TIMSS-2019-Konsortium gemeinsam getroffen.

2.5 Vorbereitung und Durchführung von TIMSS 2019

Auch wenn bereits viele Test- und Befragungsinstrumente vorliegen und umfassende Erfahrung in der Vorbereitung, Durchführung und Auswertung der Untersuchung vorhanden ist, sind in jedem neuen Zyklus Instrumente und Prozeduren weiterzuentwickeln, um auf gesellschaftliche und kulturelle Veränderungen zu reagieren. Diese Weiterentwicklungen sind dahingehend empirisch zu überprüfen, ob sie inhaltlich geeignet und auch praktisch anwendbar sind, um die Trenduntersuchung ohne Qualitätsverlust fortschreiben zu können. Um neue Instrumente und Abläufe zu entwickeln und zu erproben, sind daher in jedem TIMSS-Zyklus den jeweiligen Hauptstudien Pilotuntersuchungen vorangestellt, in denen die Güte und Funktionsweise der neu entwickelten Instrumente empirisch überprüft werden. Durch die in TIMSS 2019 im ersten Schritt vorgenommene Umstellung von papierbasiertem Testen auf das Testen mit digitalen Medien wurden zwei Pilotuntersuchungen notwendig, an denen sich auch Deutschland beteiligt hat.

Der Hauptstudie im Frühjahr 2019 gingen im Jahr 2017 in Deutschland eine Moduseffekt-Studie und in 2018 ein Feldtest voraus. In Tabelle 2.8 sind die Zeitfenster, die beteiligten Länder der Bundesrepublik Deutschland und die

Tabelle 2.8: Stichprobenumfänge und Durchführungszeitpunkte der TIMSS-2019-Vorbereitungsstudien und der Hauptstudie

	Moduseffekt-Studie	Feldtest	Hauptstudie
Zeitpunkt	Mai 2017	April 2018	Mai-Juni 2019
Anzahl der Länder der Bundesrepublik Deutschland	2 ²	5 ³	16 ⁴
Fallzahlen beteiligter Personen¹			
Schülerinnen und Schüler	847	1332	4942
Erziehungsberechtigte	-	876	3269
Lehrkräfte	-	79	397
Schulleitungen ⁵	-	57	229

1 = Die Angaben zu der Moduseffekt-Studie und dem Feldtest sind ungewichtet. Die Angaben zu der Hauptstudie sind gewichtet.

2 = An der Moduseffekt-Studie nahmen folgende Länder der Bundesrepublik Deutschland teil: Bayern und Schleswig-Holstein.

3 = Am Feldtest nahmen folgende Länder der Bundesrepublik Deutschland teil: Bayern, Nordrhein-Westfalen, Rheinland-Pfalz, Sachsen-Anhalt und Thüringen.

4 = An der Hauptstudie nahmen alle 16 Länder der Bundesrepublik Deutschland teil.

5 = 36 Schulen und Schulleitungen nahmen sowohl an eTIMSS als auch an der Brückenstudie teil. In diesen Schulen wurden 2 Parallelklassen getestet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Anzahlen der befragten Personengruppen und schließlich die Anzahl der beteiligten Schulen aufgeführt. Hierbei ist zu beachten, dass die in Tabelle 2.8 angegebenen Fallzahlen der Moduseffekt-Studie und des Feldtests die ungewichteten Fallzahlen dieser Studienabschnitte dokumentieren. Die Funktion dieser Voruntersuchungen war die Überprüfung technischer Aspekte der TIMSS-Untersuchung. Die Befunde dieser beiden Voruntersuchungen sind für die technische Entwicklung von TIMSS notwendig und wertvoll. Die hieraus gewonnenen Befunde können jedoch nicht für Deutschland verallgemeinert werden, da aus untersuchungspraktischen Gründen keine für Deutschland repräsentative Stichprobe realisiert werden konnte. Dies ist erst anhand der (hier gewichtet berichteten) Hauptstudie von 2019 möglich.

2.5.1 TIMSS 2019: Moduseffekt-Studie im Frühjahr 2017

Im Rahmen der Moduseffekt-Studie im Frühjahr 2017 wurden bewährte Instrumente aus der TIMSS-2015-Untersuchung genutzt, um diese für digitale Medien nutzbar zu machen. Da ein Gesamttest in TIMSS immer aus verschiedenen Testteilen besteht, konnten den an dieser Studie teilnehmenden Schülerinnen und Schülern Teile der Tests als papierbasierte Version und jeweils andere Teile in digitaler Version vorgelegt werden. Rund die Hälfte der Kinder fing mit der Papierversion an, um nach einer kindgerechten Pause den anderen Testteil am Laptop zu bearbeiten, während die andere Hälfte der Kinder mit dem Laptop anfang, um nach der Pause die Papierversion zu bearbeiten. Insgesamt wurden aus 16 Testblöcken (je acht für Mathematik und die Naturwissenschaften) acht Varianten von Aufgabenkombinationen erstellt. Jede Schülerin und jeder Schüler bearbeitete sowohl in der papierbasierten als auch in der digitalen medienbasierten Version jeweils Mathematik- und auch Naturwissenschaftsaufgaben, die sich für sie inhaltlich nicht wiederholten. Durch dieses balancierte Design

der Testzuweisung können Unterschiede bei den Lösungshäufigkeiten der Papierbeziehungsweise der digitalen Version auf die Präsentationsform zurückgeführt werden. Diese Informationen sind hilfreich und notwendig, um Veränderungen der Ergebnisse zwischen den bisherigen TIMSS-Zyklen und der aktuellen TIMSS-2019-Untersuchung entweder auf Veränderungen in den im Fokus liegenden mathematischen und naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler oder aber auf die Wirkung von Papier- und Computertestungen zurückführen zu können. Nur unter Berücksichtigung der Moduseffekt-Studie ist die Fortführung des in Deutschland im Jahr 2007 begonnenen Trends über das Jahr 2019 hinweg auch in Zukunft möglich. An dieser optionalen Moduseffekt-Studie haben insgesamt 28 Staaten teilgenommen.

2.5.2 TIMSS 2019: Feldtest im Frühjahr 2018

Ebenfalls zu Beginn des neuen Studienzyklus und damit schon mehr als zwei Jahre vor der eigentlichen Hauptstudie wird mit der Entwicklung der Test- und Befragungsinstrumente begonnen. Bei der (Weiter-)Entwicklung der Instrumente werden vier zentrale Zielsetzungen verfolgt: (1) Entwicklung neuer inhaltlicher Testitems, (2) Entwicklung neuer innovativer Testformate, (3) Entwicklung neuer Fragen zum Hintergrund des Lehrens und Lernens und schließlich als übergreifende Herausforderung (4) Sicherung des TIMSS-Trends. Damit sichergestellt wird, dass die unter 1 bis 3 genannten neu entwickelten Tests und Befragungen empirisch anschlussfähig an den bisher dokumentierten TIMSS-Trend sind, werden diese im Rahmen des Feldtests einer empirischen Prüfung unterzogen. Neben der psychometrisch-empirischen Überprüfung der Instrumente werden im Feldtest auch die Organisation und Durchführung der Untersuchung systematisch überprüft. Durch ein international vorbereitetes und begleitetes *Quality Control Monitoring* werden die international vorgegebenen Standards bei den Testungen und Befragungen in Vorbereitung auf die Hauptstudie ein Jahr später durch unabhängige Beobachterinnen und Beobachter der IEA Amsterdam kontrolliert. Am TIMSS-Feldtest nahmen alle an TIMSS 2019 beteiligten Staaten teil, damit die weiter- und neuentwickelten Tests und Befragungsinstrumente in allen nationalen Sprachvarianten geprüft werden konnten.

In Deutschland wurde der Feldtest insbesondere zur Adjustierung der Testinstruktionen und Kürzung der Hintergrundfragebögen genutzt. Mit Abschluss des TIMSS-2019-Feldtests konnten die Durchführungsmanuale und die Test- und Befragungsinstrumente für die Hauptstudie fertiggestellt werden.

2.5.3 TIMSS 2019: Hauptstudie

Die Hauptstudie von TIMSS 2019 wurde im Frühjahr 2019 realisiert. Das gewählte Testzeitfenster lag in allen Ländern der Bundesrepublik Deutschland vor den Sommerferien, sodass mit der TIMSS-Erhebung die Kompetenzen der Schülerinnen und Schüler in Mathematik und den Naturwissenschaften am Ende der Grundschulzeit erhoben wurden. In der Hauptstichprobe wurden die Schülerinnen und Schüler mithilfe von Laptops getestet. Die Hintergrundfragebögen, die sie, ihre Erziehungsberechtigten und die Lehrkräfte der Schulen und deren Schulleitungen ausfüllten, wurden ihnen in papierbasierter Form vorgelegt. Einer ergänzenden Stichprobe von Schülerinnen und Schülern wurden die

Testinstrumente ebenfalls in papierbasierter Form vorgelegt. Diese sogenannte Brückenstudie sollte zwei Jahre nach der Durchführung der Moduseffekt-Studie sich möglicherweise verändernde Moduseffekte feststellen und international den Trend sicherstellen; sie war für alle Staaten obligatorisch, die die TIMSS-Tests 2019 digital durchführten.

Die Fragebögen für die Lehrkräfte und Schulleitungen wurden vor dem gemeinsam vereinbarten Testtermin zugesandt, sodass diese in der Regel am Testtag von den Testleiterinnen und Testleitern bereits ausgefüllt eingesammelt werden konnten. Der Elternfragebogen hingegen wurde am Testtag allen teilnehmenden Schülerinnen und Schülern von den Testleiterinnen und Testleitern inklusive Anschreiben und Rückumschlag mit der Bitte ausgehändigt, die Unterlagen an ihre Erziehungsberechtigten weiterzugeben. In dem Anschreiben wurden die Erziehungsberechtigten darum gebeten, den Fragebogen zu bearbeiten und diesen ihrem Kind im verschlossenen Umschlag wieder mit in die Schule zu geben.

2.5.4 Übersetzung

Da die internationale Studiensprache in TIMSS Englisch ist, werden alle Materialien durch das ISC in dieser Sprache bereitgestellt. Dies sind neben den Tests und Hintergrundfragebögen auch die Manuale und Handreichungen. In den Staaten, die mit den englischsprachigen Unterlagen nicht direkt arbeiten können, sind diese zu übersetzen. Da der spätere Vergleich der Bildungssysteme insbesondere auf den getesteten Kompetenzen und erfragten Hintergrundinformationen basiert, werden die Übersetzungen besonderen Qualitätskontrollen im Rahmen der international vorgegebenen *Translation Verification* unterzogen. Zunächst sind die Staaten angehalten, die Tests und Fragebögen durch zwei Übersetzerinnen und Übersetzer unabhängig voneinander in die entsprechende Staatssprache zu übersetzen und diese im Rahmen eines Expertinnen- und Expertengesprächs in eine dann gültige Übersetzung zusammenzuführen. Die so übersetzten Tests und Befragungsinstrumente werden durch die IEA Amsterdam durch unabhängige Übersetzerinnen und Übersetzer in die englische Sprache zurückübersetzt und mit den Originalen verglichen. Abweichungen werden kommuniziert und im gegenseitigen Einvernehmen angepasst. Notwendige nationale Abweichungen werden durch das ISC dokumentiert und bei der anschließenden Berichtslegung berücksichtigt.⁵

2.5.5 Testdurchführung und Qualitätssicherung

Der Ablauf bei der Durchführung der Tests und Befragungen ist international standardisiert. Das bedeutet, dass in allen Staaten die gleichen Abläufe in den Testsitzungen und die gleichen zeitlichen Vorgaben einzuhalten sind. Hierzu werden detaillierte Manuale für die externen Testleiterinnen und Testleiter vorbereitet.

Am Testtag bearbeiten die Schülerinnen und Schüler die Tests in altersgerechten Abschnitten, unterbrochen durch regelmäßige und aktive Pausen (siehe Tabelle 2.9). Abgeschlossen wird die Testsitzung durch das Bearbeiten der

5 In den Naturwissenschaften wurde für Deutschland zusätzlich ein Item (aus den Bereichen Biologie und Problemlösen) aufgrund ungenügender Passung von englischsprachigem Original und deutscher Übersetzung aus der nationalen Skalierung ausgeschlossen.

Tabelle 2.9: Untersuchungsablauf der TIMSS-Brücken- und TIMSS-Hauptstudie 2019

Ablauf der Testsitzung	TIMSS Hauptstudie	TIMSS Brückenstudie
Vor Beginn der Testsitzung: Vorbereitung der Testsitzung zusammen mit der Testleitung		
Beginn der Testsitzung Begrüßung, Einweisung	ca. 30 Min.	ca. 10 Min.
Bearbeitung des Testhefts Teil I (Mathematik oder Naturwissenschaften)	36 Min.	36 Min.
Pause	10 Min.	10 Min.
Bearbeitung des Testhefts Teil II (Mathematik oder Naturwissenschaften)	36 Min.	36 Min.
Bearbeitung der Fragen zum Test	5 Min.	-
Pause (Austeilen der Schülerfragebögen)	-	15 Min.
Pause (Einsammeln der USB-Sticks; Austeilen der Schülerfragebögen)	15 Min.	-
Einweisung in die Bearbeitung des kognitiven Fähigkeitstests Teil I A+B	5 Min.	5 Min.
Bearbeitung des kognitiven Fähigkeitstests Teil I A (CFT 20-R) Matrizen	4 Min.	4 Min.
Bearbeitung des kognitiven Fähigkeitstests Teil I B (CFT 20-R) Reihenfortsetzen	3 Min.	3 Min.
Einweisung in die Bearbeitung des kognitiven Fähigkeitstests Teil II	3 Min.	3 Min.
Bearbeitung des kognitiven Fähigkeitstests Teil II (KFT, V3)	7 Min.	7 Min.
Einweisung in die Bearbeitung des Schülerfragebogens	5 Min.	5 Min.
Bearbeitung des Schülerfragebogens Teil I	25 Min.	25 Min.
Pause	10 Min.	10 Min.
Bearbeitung des Schülerfragebogens Teil II	20 Min.	20 Min.
Beenden der Testsitzung: Einsammeln der Testmaterialien		
Reine Bearbeitungszeit	136 Min.	131 Min.
Gesamtzeit	ca. 214 Min.	ca. 189 Min.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Schülerfragebögen. Um die zeitliche Belastung der Schülerinnen und Schüler an den für die TIMSS-Untersuchung ausgewählten Förderschulen zu verringern, wurden die Test- und Befragungsteile in Absprache mit der internationalen Studienleitung an diesen Schulen auf zwei Tage verteilt.

Um sicherzustellen, dass diesen Vorgaben in den Schulen gefolgt wird, werden durch die IEA Amsterdam geschulte unabhängige Prozessbeobachterinnen und -beobachter in die Schulen geschickt, in denen die TIMSS-Untersuchung durchgeführt wird. Auch der Ablauf dieser Qualitätssicherungsmaßnahme in den Schulen ist standardisiert und wird von den internationalen Prozessbeobachterinnen und -beobachtern im Rahmen der Qualitätssicherung schrittweise durchgeführt und dokumentiert. Bei zu starken Abweichungen von den Vorgaben kann es zu Ausschlüssen von Schülerinnen und Schülern oder sogar ganzen Klassen für nachfolgende Auswertungen kommen. Nicht nur die Übersetzung der Materialien und der Ablauf der Testung unterliegt zur Wahrung der Vergleichbarkeit internationalen Standards, auch die Dateneingabe durchläuft eine mehrstufige Qualitätskontrolle.

Neu entwickelte Testaufgaben werden in jedem teilnehmenden Staat übersetzt und im Feldtest von mindestens 200 Schülerinnen und Schülern je Aufgabe bearbeitet, um so die neuen Testaufgaben empirisch zu überprüfen und deren Passung an die Kompetenzmodelle in Mathematik und den Naturwissenschaften sicherzustellen. Aufwendige Rückübersetzungen und ein Vergleich mit den Originalvarianten der Tests und Befragungsinstrumente sorgen im

Rahmen der internationalen Übersetzungskontrolle für international vergleichbare Befragungsinstrumente. Im Anschluss an die Testung mit den mehrfach geprüften Instrumenten müssen die erhobenen Daten in ein elektronisches Format überführt werden. Dazu werden die Daten in den nationalen Testzentren eingegeben und auf Eingabefehler kontrolliert. Anschließend übernimmt die internationale Datenverarbeitungsabteilung der IEA Hamburg die Zusammenführung und Kontrolle der Daten aller teilnehmenden Staaten. Nach Kontrolle und Korrekturen von Eingaberoutinen werden darüber hinaus national notwendige Anpassungen so vorgenommen und dokumentiert, dass sie international in eine gemeinsame Datenbank einfließen können. Diese Datenbank kann anschließend genutzt werden, um die Kompetenzstände der Schülerinnen und Schüler zu bestimmen und damit die Basis für die internationalen und nationalen Berichtslegungen zu schaffen.

2.6 Technische und empirische Voraussetzungen für die Auswertung von TIMSS-Daten

Um den Besonderheiten in der Anlage und Durchführung der Untersuchung angemessen Rechnung zu tragen, müssen das Stichprobendesign und die Verwendung des Testdesigns bei den Analysen Berücksichtigung finden. Hierzu sind die Nutzung der Stichprobengewichte, die Bestimmung der Stichproben- und Messfehler durch die simultane Berücksichtigung der *plausible values* (im Rahmen von Kompetenzanalysen) und die stichprobengerechte Bestimmung der Stichprobenfehler (durch das sogenannte *Jackknifing*-Verfahren, Foy & LaRoche, 2016) bei den Analysen obligatorisch. Darüber hinaus ist die Angabe der Standardfehler beziehungsweise der sich daraus ergebenden Konfidenzintervalle in diesem Zusammenhang üblich. In Tabellen und Abbildungen, bei denen lediglich eine beschreibende Darstellung vorgenommen wird, kann auf diese Informationen verzichtet werden.

Wird bei Analysen einer dieser notwendigen Aspekte nicht berücksichtigt (z. B. Schätzung von Mittelwerten oder Zusammenhängen von Merkmalen), kann es zu Fehlinterpretationen sowohl bei der Beschreibung der Befunde im Trend als auch im internationalen Vergleich kommen. In einigen Untersuchungen werden für fehlende Antworten der Befragten plausible Werte empirisch geschätzt. Dies ist insbesondere im Kontext von großangelegten Schulvergleichsstudien ein übliches Verfahren, möglichen Verzerrungen durch Stichprobenausfälle zu begegnen und konsistentere Schätzer für die Grundgesamtheit zu erhalten. Aufgrund des sehr kurzen Zeitfensters zwischen Datenübergabe und Berichtslegung konnte dies in TIMSS 2019, wie auch bereits in TIMSS 2007 und 2011, nicht realisiert werden. Daher wurden, wo notwendig, einzelne Angaben aus TIMSS 2015 ebenfalls ohne Imputation berechnet, um einen direkten Vergleich über alle berichteten Studienzyklen zu ermöglichen. Eventuelle geringfügige Abweichungen zum TIMSS-2015-Bericht sind hier entsprechend kenntlich gemacht.

Schließlich sei in Bezug auf die Analysen von Daten der Schulleitungen und Lehrkräfte auf eine weitere Besonderheit hingewiesen: Da das gewählte Stichprobendesign für die Ziehung einer Schülerinnen- und Schülerstichprobe optimiert ist, können bei entsprechenden Analysen keine Aussagen über die Schulbeziehungsweise Lehrkräftestichprobe getroffen werden, da die Lehrkräfte nicht als Zufallsauswahl aller Lehrkräfte in der Schule ausgewählt wurden (und damit deren Ziehungswahrscheinlichkeit nicht bestimmt werden kann). In TIMSS

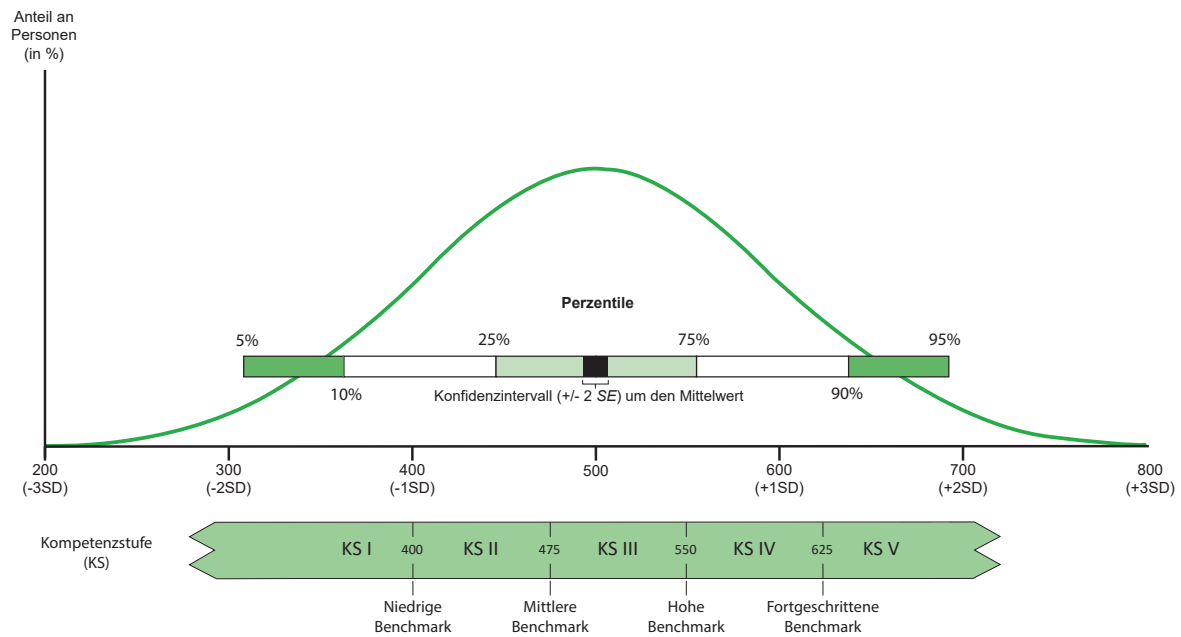
werden die Lehrkräfte befragt, die die befragten Schülerinnen und Schüler unterrichten. In der Logik der Stichprobenziehung können somit nur Aussagen über Schülerinnen und Schüler gemacht werden, nicht aber über die Lehrkräfte an sich. So ist die Formulierung, dass zum Beispiel ‚89 Prozent aller Schülerinnen und Schüler von weiblichen Lehrkräften unterrichtet werden‘ zwar umständlich, aber richtig – im Gegensatz zu der Aussage, dass ‚89 Prozent der Lehrkräfte weiblich sind‘. Bei der Analyse von spezifischen Fragen im Schulleitungsfragebogen gilt dieselbe Logik. Da jedoch die Ziehungswahrscheinlichkeiten der Schulen bestimmbar sind, können (ohne Bezug auf Variablen von Schülerinnen und Schülern) auch Aussagen über Schulen in Deutschland gemacht werden.

2.7 Hinweise für die Ergebnisdarstellungen

Neben der tabellarischen Dokumentation der Ergebnisse werden in dem vorliegenden Band auch Abbildungen genutzt, um die komplexe Befundlage möglichst anschaulich zu visualisieren. Dabei werden neben der Darstellung von prozentualen Anteilen als Histogramme (Säulendiagramme) auch statistische Kennwerte (wie z.B. Mittelwerte) dargestellt. Eine Besonderheit der Abbildungen und Tabellen liegt darin, dass die stichprobenbasierten Unsicherheiten (Standardfehler und hieraus bestimmbare Konfidenzintervalle) mitberücksichtigt werden, um beurteilen zu können, ob beobachtete Unterschiede beziehungsweise Abweichungen zufällig oder statistisch auffällig (signifikant) sind.

Gerade bei der vergleichenden Gegenüberstellung der von den Grundschülerinnen und Grundschülern gezeigten Kompetenzen wird eine Abbildungsvariante genutzt, der sowohl der festgestellte Mittelwert als auch dessen statistische Streuung und die allgemeine Verteilung der Kompetenzen zu entnehmen ist. Bei dieser Abbildung handelt es sich um ein Perzentilband, das die mittleren 90 Prozent einer Merkmalsverteilung in einer Gruppe darstellt. Die jeweils verbleibenden 5 Prozent am unteren und oberen Ende der Verteilung werden in der Regel nicht abgebildet, da diese mit besonders großen statistischen Unsicherheiten behaftet sind. Ein Perzentilband kann, wie in der Beispielabbildung 2.3 dargestellt, wie Höhenlinien einer von oben betrachteten statistischen Verteilung interpretiert werden. Der mittlere (schwarze) Abschnitt kennzeichnet den für die entsprechende Gruppe festgestellten Mittelwert inklusive der stichprobenbasierten Schwankung (Konfidenzintervall). Die hellgrün hinterlegten Balken links und rechts vom Mittelwert kennzeichnen die mittleren 50 Prozent der Beobachtungen bei dem untersuchten Merkmal. Die links und rechts anschließenden Balkenabschnitte (ohne farbliche Füllung) kennzeichnen jeweils die 15 Prozent niedrigeren beziehungsweise 15 Prozent höheren Werte des dargestellten Merkmals.

Werden entsprechende Perzentilbänder für verschiedene Gruppen (z.B. Staaten in Hinblick auf die Kompetenzen der Viertklässlerinnen und Viertklässler oder besondere Schülerinnen- und Schülergruppen innerhalb von Deutschland) dargestellt, so gibt die parallele Darstellung der Konfidenzintervalle der Staaten einen ersten visuellen Eindruck, ob sich diese Gruppen signifikant voneinander unterscheiden oder nicht. Überlappen sich die mittleren Bereiche (Konfidenzintervalle) der Perzentilbänder, sind mögliche Unterschiede in den Mittelwerten nicht signifikant. Wenn die Konfidenzintervalle sich nicht überlappen, liegen entsprechend signifikante Unterschiede zwischen den dargestellten Mittelwerten vor. Neben den visuellen Hinweisen auf signifikante Unterschiede

Abbildung 2.3: Normalverteilung mit Perzentilband, *Benchmarks* und Kompetenzstufen

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

zeigen in den daneben dargestellten Tabellen Symbole explizit statistisch signifikante Unterschiede an.

Die Perzentilbänder geben, insbesondere in Verbindung mit weiteren statistischen Kennwerten wie Mittelwerten und Standardabweichungen, empirisch ermittelte Werte wieder, jedoch sind diese Angaben als Zahlenwerte sehr abstrakt. Um die Skalenwerte der Kompetenztests inhaltlich intuitiver nachvollziehbar zu machen, wurden Bereiche der Kompetenzverteilung zu sogenannten Kompetenzstufen zusammengefasst, die durch international vorgegebene Grenzwerte (*Benchmarks*) getrennt werden. In Abbildung 2.3 sind die *Benchmarks* und Kompetenzstufen unter der Verteilungskurve und dem korrespondierenden Perzentilband dargestellt.

Literatur

- Beauftragte der Bundesregierung für die Belange von Menschen mit Behinderungen. (Hrsg.). (2017). *Die UN-Behindertenrechtskonvention. Übereinkommen über die Rechte von Menschen mit Behinderungen*. Verfügbar unter: https://www.behindertenbeauftragter.de/SharedDocs/Publikationen/UN-Konvention_Schattenuebersetzungpdf;jsessionid=EBFAFBF2759A6F68B60702E3D2AB1255.1_cid509?__blob=publication-File&v=2
- Blömeke, S. (2014). Forschung zur Lehrerbildung im internationalen Vergleich. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 441–467). Münster: Waxmann.
- Bos, W. & Postlethwaite, T. N. (2005). Möglichkeiten, Grenzen und Perspektiven internationaler Schulleistungsforschung. In R. Tippelt (Hrsg.), *Handbuch Bildungsforschung* (S. 241–261). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-99634-3_13
- Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, 43(3), 161–173. <https://doi.org/10.1080/00461520802178466>

- Eickelmann, B., Bos, W. & Labusch, A. (2019). Die Studie ICILS 2018 im Überblick. Zentrale Ergebnisse und mögliche Entwicklungsperspektiven. In B. Eickelmann, W. Bos, J. Gerick, F. Goldhammer, H. Schaumburg, K. Schwippert, M. Senkbeil & J. Vahrenhold (Hrsg.), *ICILS 2018. Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking* (S. 7–31). Münster: Waxmann.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive. *Zeitschrift für Pädagogische Psychologie*, 28(3), 127–137. <https://doi.org/10.1024/1010-0652/a000129>
- Foy, P. & LaRoche, S. (2016). Estimating standard errors in the TIMSS 2015 results. In M. O. Martin, I. V. S. Mullis & M. Hooper (Hrsg.), *Methods and procedures in TIMSS 2015* (S. 4.1–4.69). Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.
- Helmke, A. (2009). Unterrichtsforschung. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (2. Aufl., S. 44–50). Bad Heilbrunn: Klinkhardt.
- Hornberg, S., Bos, W., Buddeberg, I., Potthoff, B. & Stubbe, T. C. (2007). Anlage und Durchführung von IGLU 2006. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 21–45). Münster: Waxmann.
- Husén, T. (Hrsg.). (1967). *International study of achievement in mathematics: A comparison of twelve countries* (Bd. 1). New York: Wiley & Sons.
- Klieme, E., Schümer, G. & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I. ‚Aufgabekultur‘ und Unterrichtsgestaltung im internationalen Vergleich. In E. Klieme & J. Baumert (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 43–57). Bonn: Bundesministerium für Bildung und Forschung.
- Kunter, M. (2016). Bedingungen und Effekte von Unterricht. Aktuelle Forschungsperspektiven aus der pädagogischen Psychologie. In N. McElvany, W. Bos, H. G. Holtappels, M. M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts*. (Dortmunder Symposium der Empirischen Bildungsforschung, Bd. 1), S. 9–32. Münster: Waxmann.
- Lankes, E.-M., Bos, W., Mohr, I., Plaßmeier, N., Schwippert, K., Sibberns, H. & Voss, A. (2003). Anlage und Durchführung der Internationalen Grundschul-Lese-Untersuchung (IGLU) und ihrer Erweiterung um Mathematik und Naturwissenschaften (IGLU-E). In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 7–28). Münster: Waxmann.
- Martin, M. O., Mullis, I. V. S. & Hooper, M. (Hrsg.). (2016). *Methods and procedures in TIMSS 2015*. Verfügbar unter: Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/publications/timss/2015-methods.html>
- Medienpädagogischer Forschungsverbund Südwest (mpfs). (2019). *KIM-Studie 2018. Kindheit, Internet, Medien. Basisuntersuchung zum Medienumgang 6- bis 13-Jähriger*. Stuttgart: mpfs. Verfügbar unter: https://www.mpfs.de/fileadmin/files/Studien/KIM/2018/KIM-Studie_2018_web.pdf
- Medienpädagogischer Forschungsverbund Südwest (mpfs). (2020). *JIM-Studie 2019 Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger in Deutschland*. Stuttgart: mpfs. Verfügbar unter: https://www.mpfs.de/fileadmin/files/Studien/JIM/2019/JIM_2019.pdf
- Mullis, I. V. S. & Martin, M. O. (Hrsg.). (2017). *TIMSS 2019. Assessment frameworks*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College.
- Robitzsch, A., Lüdtke, O., Köller, O., Kröhne, U., Goldhammer, F. & Heine, J.-H. (2017). Herausforderungen bei der Schätzung von Trends in Schulleistungstudien: Eine Skalierung der deutschen PISA-Daten. *Diagnostica*, 63(2), 148–165. <https://doi.org/10.1026/0012-1924/a000177>
- Särndal, C.-E., Swensson, B. & Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer. <https://doi.org/10.1007/978-1-4612-4378-6>
- Schwippert, K., Stubbe, T. C. & Wendt, H. (2015). IGLU/TIMSS: International vergleichende Schulleistungsuntersuchungen im Grundschulbereich von 2001 bis 2011. In H. Wendt, T. C. Stubbe, K. Schwippert & W. Bos (Hrsg.), *IGLU & TIMSS. 10 Jahre international vergleichende Schulleistungsforschung in der Grundschule. Vertiefende Analysen zu IGLU und TIMSS 2001 bis 2011* (S. 19–34). Münster: Waxmann.

- Stigler, J. W. & Hiebert, J. (1999). *The teaching gap. Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.
- Tschannen-Moran, M. & Woolfolk-Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17, 783–805.
[https://doi.org/10.1016/S0742-051X\(01\)00036-1](https://doi.org/10.1016/S0742-051X(01)00036-1)
- UNESCO Institute for Statistics. (2012). *International Standard Classification of Education. ISCED 2011*. Montreal: UNESCO Institute for Statistics.
- Weinert, F. E. (2001). Vergleichende Leistungsmessungen in Schulen. Eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Weinheim: Beltz.
- Wendt, H., Bos, W., Kasper, D., Walzebug, A., Goy, M. & Jusufi, D. (2016). Ziele, Anlage und Durchführung der Trends in International Mathematics and Science Study (TIMSS 2015). In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland* (S. 31–78). Münster: Waxmann.
- Wendt, H., Schwippert, K. & Stubbe, T. C. (2016). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015 Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 315–329). Münster: Waxmann.
- Wendt, H., Tarelli, I., Bos, W., Frey, K. & Vennemann, M. (2012). Ziele, Anlage und Durchführung der Trends in International Mathematics and Science Study (TIMSS 2011). In W. Bos, H. Wendt, O. Köller & C. Selter (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 27–68). Münster: Waxmann.

Kapitel 3

Mathematische Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse

Christoph Selter, Daniel Walter, Aiso Heinze, Johanna Brandt
und Armin Jentsch

3.1 Einleitung

Die internationalen Vergleichsstudien *Third International Mathematics and Science Study* (TIMSS) im Jahr 1995 (Baumert et al., 1997) und *Programme for International Student Assessment* (PISA) in 2000 (Baumert et al., 2001) ergaben vergleichsweise schwache Leistungen von Schülerinnen und Schülern der Sekundarstufen in Deutschland. Infolgedessen entwickelte sich ein zunehmendes gesellschaftliches und bildungspolitisches Interesse an von den Lernenden zu erreichenden Kompetenzen und an Fragen zur Qualitätssicherung im Bildungssystem. Dies betraf auch das fachliche Lernen in der Grundschule.

Mittlerweile liegen hierzu umfangreiche Datensätze aus den vergangenen zwei Dekaden vor, die es ermöglichen, mathematische Kompetenzen von Grundschülerinnen und Grundschülern aus Deutschland im internationalen Vergleich zu beschreiben und zu analysieren. Erstmals wurde dies im Zuge der Beteiligung Deutschlands am Erweiterungsteil der *Internationalen Grundschul-Lese-Untersuchung/Progress in International Reading Literacy Study* (IGLU/PIRLS) IGLU-E im Jahr 2001 realisiert (Bos et al., 2003) und durch die Teilnahme an den *Trends in International Mathematics and Science Studies* (TIMSS) der Jahre 2007, 2011, 2015 sowie der in diesem Band beschriebenen TIMS-Studie 2019 weitergeführt. Wesentliche Befunde der bisherigen drei TIMSS-Zyklen werden nachfolgend im Überblick dargelegt (Bos et al., 2016):

- Im internationalen Vergleich lag die Mathematikleistung von Viertklässlerinnen und Viertklässlern aus Deutschland sowohl 2007 mit 525 Punkten, 2011 mit 528 Punkten als auch 2015 mit 522 Punkten signifikant über dem internationalen Mittelwert von 473, 491 beziehungsweise 509 Punkten. Zudem lagen die Leistungswerte von 2007 und 2011 statistisch signifikant über den Leistungswerten der an der Studie teilnehmenden Staaten der *Europäischen Union* (EU: 514 bzw. 519) sowie der beteiligten Staaten der *Organisation for*

Economic Co-operation and Development (OECD: 513 bzw. 521). Im Jahr 2015 lagen diese hingegen signifikant darunter (EU: 527; OECD: 528).

- Nachdem sich der Leistungswert bei TIMSS 2011 im Vergleich zu TIMSS 2007 leicht verbessert hatte, liegt ein signifikant niedrigerer Wert bei TIMSS 2015 im Vergleich zum vorherigen Zyklus vor. Zudem ist in allen drei Studienzyklen ein erheblicher Abstand zur Leistungsspitze zu verzeichnen. So wurde der Bestwert von 607 Punkten bei TIMSS 2007 von Lernenden aus Hongkong, von 605 Punkten bei TIMSS 2011 von Schülerinnen und Schülern aus der Republik Korea sowie von 618 Punkten bei TIMSS 2015 von Lernenden aus Singapur erzielt.
- In allen drei TIMSS-Zyklen konnte eine im Vergleich zu den anderen Teilnehmerstaaten relative Leistungshomogenität nachgewiesen werden. Gleichwohl ist der Anteil der Viertklässlerinnen und Viertklässler, die lediglich niedrige oder rudimentäre mathematische Fähigkeiten und Fertigkeiten entwickeln konnten (Kompetenzstufe I und II), mit etwa einem Viertel nahezu gleich geblieben.
- Auch hinsichtlich des relativen Anteils der Viertklässlerinnen und Viertklässler auf der höchsten Kompetenzstufe V können nur marginale Veränderungen im Vergleich der drei TIMSS-Zyklen festgestellt werden. Nach wie vor ist der Anteil an Schülerinnen und Schülern aus Deutschland, die der höchsten Kompetenzstufe V zugeordnet werden können, mit rund 5 Prozent im internationalen Vergleich gering.
- In den Inhaltsbereichen *Messen und Geometrie* sowie *Daten* unterschieden sich die Leistungen der Schülerinnen und Schüler im Vergleich von TIMSS 2007 und TIMSS 2015 nicht signifikant. Jedoch lagen die Leistungswerte im Bereich *Arithmetik* in TIMSS 2007 signifikant höher als bei TIMSS 2015. Damit gehört Deutschland zu den wenigen Staaten, bei denen sich in einem Inhaltsbereich im Trend signifikant schwächere Leistungen nachweisen lassen.
- Im kognitiven Anforderungsbereich *Problemlösen* lassen sich über die Studienzyklen hinweg nahezu gleich gute Testleistungen feststellen. Hingegen schnitten Grundschülerinnen und Grundschüler aus Deutschland in *Reproduzieren* in TIMSS 2011 und 2015 signifikant stärker als in TIMSS 2007 ab. Für den kognitiven Anforderungsbereich *Anwenden* zeigten sich nahezu gleich gute Testwerte in TIMSS 2007 und 2011, während in TIMSS 2015 in diesem Bereich signifikant schwächere Testleistungen im Vergleich zu den beiden vorherigen Zyklen erzielt wurden.
- Während in TIMSS 2007 und TIMSS 2011 Jungen signifikant besser als Mädchen abschnitten, ließen sich in TIMSS 2015 erstmals keine bedeutenden Leistungsunterschiede zwischen den Geschlechtern auf der Gesamtskala Mathematik feststellen.
- Darüber hinaus ließen sich für die Studienzyklen Zusammenhänge zwischen der sozialen Herkunft und den mathematischen Kompetenzen am Ende der Grundschulzeit feststellen. Seit TIMSS 2007 ist das Ausmaß der sozial bedingten Ungleichheit in Deutschland konstant geblieben.
- Der überwiegende Teil der Viertklässlerinnen und Viertklässler in Deutschland verfügt über eine positive Einstellung zur Mathematik sowie ein positives Fähigkeitsselbstkonzept. Dieser Befund zeigte sich in allen TIMSS-Zyklen.

Dieses Kapitel befasst sich mit den Ergebnissen Deutschlands in TIMSS 2019. In Abschnitt 3.2 werden zunächst zentrale Aspekte mathematischer Grundbildung in Deutschland beleuchtet, bevor in Abschnitt 3.3 die TIMSS-Rahmenkonzeption

zur Erfassung mathematischer Kompetenzen dargelegt wird. Abschnitt 3.4 beschreibt den TIMSS-Test zur Erfassung der mathematischen Kompetenzen. Im fünften Abschnitt werden die Ergebnisse der gegenwärtigen Studie entlang folgender Fragestellungen beschrieben, wobei in jedem Abschnitt auch untersucht wird, wie sich die Ergebnisse von TIMSS 2019 von denen aus den drei vorangegangenen Studienzyklen 2007, 2011 und 2015 unterscheiden:

- Welche Mathematikleistungen zeigen Schülerinnen und Schüler in Deutschland am Ende der vierten Jahrgangsstufe im internationalen Vergleich? (Abschnitt 3.5.1)
- Wie lassen sich die Leistungskennwerte auf die Kompetenzstufen in Deutschland im internationalen Vergleich einordnen? Wie groß sind die Gruppen der auffällig leistungsstarken und leistungsschwachen Kinder? (Abschnitt 3.5.2)
- Welche Ergebnisse erzielen Schülerinnen und Schüler differenziert nach den mathematischen Teilgebieten *Arithmetik*, *Messen und Geometrie* und *Daten* im internationalen Vergleich? Wie unterscheiden sich die Leistungen der Schülerinnen und Schüler nach mathematischen Teilgebieten? (Abschnitt 3.5.3)
- Welche Ergebnisse erzielen Schülerinnen und Schüler differenziert nach den kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* im internationalen Vergleich? Unterscheiden sich die Leistungen der Schülerinnen und Schüler nach kognitiven Anforderungsbereichen? (Abschnitt 3.5.4)
- Über welche Einstellungen zur Mathematik und über welches mathematische Selbstkonzept verfügen die Schülerinnen und Schüler am Ende der vierten Jahrgangsstufe? (Abschnitt 3.5.5)

Weitere Befunde zu den in TIMSS 2019 erfassten mathematischen Kompetenzen werden in den Kapiteln 5, 8, 9 und 10 in diesem Band dargelegt. Diese Kapitel enthalten Analysen der in TIMSS 2019 untersuchten Kompetenzen vor dem Hintergrund des Moduseffektes beziehungsweise der Variablen *Geschlecht*, *soziale Herkunft* sowie *Migrationshintergrund*, weshalb diese Variablen in diesem Kapitel nicht näher in den Blick genommen werden.

3.2 Mathematische Grundbildung in Deutschland

In diesem Abschnitt werden wesentliche Aspekte mathematischer Grundbildung in Deutschland skizziert, um in den nachfolgenden Kapiteln die Beschreibung der TIMSS-Testkonzeption sowie die daraus resultierenden Testergebnisse einordnen zu können. Hierfür werden festgelegte Kompetenzerwartungen beschrieben sowie deren Umsetzung in den Curricula der Länder der Bundesrepublik Deutschland analysiert.

Im Jahr 2004 wurden die Bildungsstandards Mathematik der *Kultusministerkonferenz* (KMK) für den Primarbereich verabschiedet. Sie legen fest, welche mathematischen Kompetenzen Schülerinnen und Schüler am Ende ihrer Grundschulzeit erworben haben sollten. Neben der Beschreibung und Operationalisierung *inhaltsbezogener* und *allgemeiner mathematischer Kompetenzen* werden die erwarteten Kompetenzen auch in drei *Anforderungsbereichen* auf mathematische Konzepte bezogen.

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

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

Zur Leitidee *Größen und Messen* beispielsweise sind folgende zwei inhaltsbezogene Kompetenzen formuliert:

- Größenvorstellungen besitzen und
- mit Größen in Sachsituationen umgehen.

Diese Kompetenzen wiederum werden in Form von Standards konkretisiert. Die Kompetenz *Größenvorstellungen besitzen* beispielsweise ist wie folgt gegliedert:

- Standardeinheiten aus den Bereichen Geldwerte, Längen, Zeitspannen, Gewichte und Rauminhalte kennen,
- Größen vergleichen, messen und schätzen,
- Repräsentanten für Standardeinheiten kennen, die im Alltag wichtig sind,
- Größenangaben in unterschiedlichen Schreibweisen darstellen (umwandeln) sowie
- im Alltag gebräuchliche einfache Bruchzahlen im Zusammenhang mit Größen kennen und verstehen.

Allgemeine mathematische Kompetenzen: Neben den inhaltsbezogenen Kompetenzen beschreiben die Bildungsstandards auch allgemeine, inhaltsübergreifende mathematische Kompetenzen. Die Berücksichtigung allgemeiner mathematischer Kompetenzen wird als zentral angesehen, um einen auf Verständnis ausgerichteten Mathematikunterricht zu gestalten sowie positive Grundhaltungen zum Fach Mathematik zu entwickeln. Die nachfolgenden fünf allgemeinen mathematischen Kompetenzen sind in den Bildungsstandards festgeschrieben (KMK, 2005):

- Problemlösen,
- Kommunizieren,
- Argumentieren,
- Modellieren,
- Darstellen.

Auch diese Kompetenzen sind in Form von Standards konkretisiert. Für das *Argumentieren* beispielsweise lauten diese:

- mathematische Aussagen hinterfragen und auf Korrektheit prüfen,
- mathematische Zusammenhänge erkennen und Vermutungen entwickeln sowie
- Begründungen suchen und nachvollziehen.

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

- Anforderungsbereich I – *Reproduzieren*: Das Lösen der Aufgabe erfordert Grundwissen und das Ausführen von Routinetätigkeiten.
- Anforderungsbereich II – *Zusammenhänge herstellen*: Das Lösen der Aufgabe erfordert das Erkennen und Nutzen von Zusammenhängen.

- Anforderungsbereich III – *Verallgemeinern und Reflektieren*: Das Lösen der Aufgabe erfordert komplexe Tätigkeiten wie Strukturieren, Entwickeln von Strategien, Beurteilen und Verallgemeinern.

Berücksichtigung der Bildungsstandards in den Lehrplänen der Länder der Bundesrepublik Deutschland: Die Tabelle 3.1 zeigt, inwieweit die Bildungsstandards in den Lehrplänen der Länder (in einigen Ländern auch als *Kerncurricula*, *Fachanforderungen* oder *Rahmenpläne* bezeichnet) umgesetzt wurden. Für die Curriculumanalyse wurden die Lehrpläne herangezogen, die im Mai 2019 zum Zeitpunkt der Datenerhebung für die vierte Jahrgangsstufe relevant waren.

Tabelle 3.1: Veröffentlichungszeitpunkte der Lehrpläne Grundschule Mathematik für die vierte Jahrgangsstufe, Stand Mai 2019

Im Mai 2019 für die 4. Jahrgangsstufe gültige Lehrpläne						
Länder	seit	vor Oktober 2004	zwischen 2004 und 2007	zwischen 2007 und 2011	zwischen 2011 und 2015	zwischen 2015 und 2019
Schleswig-Holstein ¹	1997	✓	–	–	–	–
Bremen	2004	✓	–	–	–	–
Mecklenburg-Vorpommern	2004	✓	–	–	–	–
Sachsen-Anhalt	2007	–	–	✓	–	–
Nordrhein-Westfalen	2008	–	–	✓	–	–
Saarland	2009	–	–	✓	–	–
Sachsen	2009	–	–	✓	–	–
Thüringen	2010	–	–	✓	–	–
Hamburg	2011	–	–	–	✓	–
Hessen	2011	–	–	–	✓	–
Rheinland-Pfalz	2015	–	–	–	–	✓
Baden-Württemberg	2016	–	–	–	–	✓
Bayern	2016	–	–	–	–	✓
Berlin	2017	–	–	–	–	✓
Brandenburg	2017	–	–	–	–	✓
Niedersachsen	2017	–	–	–	–	✓

1 = Für Schleswig-Holstein wurde ein neuer Lehrplan entwickelt, der ab dem Schuljahr 2018/19 für die neu eingeschulten Schülerinnen und Schüler unterrichtswirksam wird. Zum Zeitpunkt der Erhebung von TIMSS 2019 war jedoch noch der bisherige Lehrplan von 1997 für die Viertklässlerinnen und Viertklässler unterrichtswirksam.

Tabelle 3.1 zeigt, dass fünf Länder der Bundesrepublik Deutschland zwischen den TIMSS-Zyklen von 2007 und 2011, zwei Länder nach TIMSS 2011 und vor TIMSS 2015 sowie sechs Länder bis zum aktuellen Studienzyklus neue Lehrpläne verabschiedeten. Ergänzend sei angemerkt, dass Niedersachsen bereits im Jahr 2006 – und damit relativ schnell nach der Veröffentlichung der Bildungsstandards – mit einem neuen Lehrplan reagierte. Im Jahr 2017 folgte eine erneute Überarbeitung.

Insgesamt zeigt sich, dass im Mai 2019 nahezu alle Länder der Bundesrepublik Deutschland neue Lehrpläne eingeführt haben, die in konkretem Bezug auf die Bildungsstandards entstanden sind. Zum Zeitpunkt der Datenerhebung von TIMSS 2015 traf dies hingegen nur für die Hälfte der Länder zu.

Um inhaltliche Zusammenhänge zwischen den Bildungsstandards und den einzelnen Lehrplänen der Länder der Bundesrepublik Deutschland herzustellen, wird nachfolgend geprüft, inwieweit die in den Bildungsstandards formulierten inhalts- und prozessbezogenen Kompetenzen in den Lehrplänen der Länder berücksichtigt werden. Dies wird zunächst für die inhaltsbezogenen Kompetenzen (*Zahlen und Operationen*, *Raum und Form*, *Muster und Strukturen*, *Größen und Messen* sowie *Daten, Häufigkeit und Wahrscheinlichkeit*) anhand von Tabelle 3.2 dargestellt. Dabei gilt es zu beachten, dass eine Kompetenz nur dann in der Tabelle als berücksichtigt vermerkt ist, wenn im jeweiligen Lehrplan ein entsprechender Kompetenzbereich formuliert ist. Die farbig hinterlegten Zeilen deuten zudem darauf hin, dass der jeweilige Lehrplan vor der Veröffentlichung der Bildungsstandards in Kraft getreten ist.

Tabelle 3.2 ist zu entnehmen, dass eine gute Passung zwischen den inhaltsbezogenen Kompetenzformulierungen der Bildungsstandards und denen der

Tabelle 3.2: Inhaltsbezogene Kompetenzen der KMK-Bildungsstandards in den Lehrplänen der Länder der Bundesrepublik Deutschland, Stand Mai 2019

Länder	seit	Zahlen und Operationen	Raum und Form	Muster und Strukturen	Größen und Messen	Daten, Häufigkeiten und Wahrscheinlichkeiten
Schleswig-Holstein	1997	✓	✓	–	✓	– ²
Bremen	2004	✓	✓	–	✓	✓
Mecklenburg-Vorpommern	2004	✓	✓	–	✓	✓
Sachsen-Anhalt	2007	✓	✓	–	✓	✓
Nordrhein-Westfalen	2008	✓	✓	– ¹	✓	✓
Saarland	2009	✓	✓	✓	✓	✓
Sachsen	2009	✓	✓	–	✓	– ²
Thüringen	2010	✓	✓	– ¹	✓	– ²
Hamburg	2011	✓	✓	✓	✓	✓
Hessen	2011	✓	✓	✓	✓	✓
Rheinland-Pfalz	2015	✓	✓	– ¹	✓	✓
Baden-Württemberg	2016	✓	✓	– ¹	✓	✓
Bayern	2016	✓	✓	✓	✓	✓
Berlin	2017	✓	✓	–	✓	✓
Brandenburg	2017	✓	✓	–	✓	✓
Niedersachsen	2017	✓	✓	✓	✓	✓

1 = *Muster und Strukturen* wird als impliziter Bestandteil des Lehrplans betrachtet. Es werden jedoch keine Kompetenzerwartungen zu diesem Bereich beschrieben.

2 = *Daten, Häufigkeiten und Wahrscheinlichkeiten* ist nicht als eigenständiger Kompetenzbereich aufgeführt. Jedoch sind die diesem Bereich zugeordneten Kompetenzerwartungen im Bereich *Arithmetik* benannt.

Grün eingefärbte Zeilen deuten darauf hin, dass die jeweiligen Lehrpläne der Länder der Bundesrepublik Deutschland vor den KMK-Bildungsstandards verabschiedet wurden.

Lehrpläne der Länder der Bundesrepublik Deutschland vorliegt. Auffällig erscheint, dass *Muster und Strukturen* in elf Ländern nicht – wie in den Bildungsstandards – als eigenständiger Kompetenzbereich aufgefasst wird. In den Lehrplänen einiger Länder, wie etwa Rheinland-Pfalz oder Thüringen, wird *Muster und Strukturen* jedoch insofern berücksichtigt, als dass der Kompetenzbereich ausdrücklich als integraler oder übergeordneter Bestandteil aller weiteren Bereiche aufgefasst wird.

Darüber hinaus fehlen lediglich in drei Ländern Kompetenzformulierungen für die inhaltsbezogene Kompetenz *Daten, Häufigkeit und Wahrscheinlichkeit*, die in einigen Ländern auch als *Daten und Zufall* (bspw. in Bayern und Hamburg) geführt ist. Gleichwohl ordnen diese Länder die entsprechenden Kompetenzerwartungen dem Kompetenzbereich *Zahlen und Operationen* unter, sodass eine Berücksichtigung dieses Kompetenzbereiches ebenfalls gegeben ist.

Nachdem die Verankerung inhaltsbezogener Kompetenzen in den Lehrplänen der Länder der Bundesrepublik Deutschland thematisiert wurde, erfolgt nun eine analoge Analyse für die prozessbezogenen Kompetenzen in Tabelle 3.3. Im Weiteren wird der Begriff prozessbezogene Kompetenzen für die allgemeinen mathematischen Kompetenzen verwendet, da die meisten Lehrpläne diese

Tabelle 3.3: Prozessbezogene Kompetenzen der KMK-Bildungsstandards in den Lehrplänen der Länder der Bundesrepublik Deutschland, Stand Mai 2019

Länder	seit	Problemlösen	Kommunizieren	Argumentieren	Modellieren	Darstellen
Schleswig-Holstein	1997	–	–	–	–	–
Bremen ¹	2004	✓	✓	✓	✓	–
Mecklenburg-Vorpommern ¹	2004	✓	✓	✓	✓	–
Sachsen-Anhalt ³	2007	✓	✓	✓	✓	–
Nordrhein-Westfalen ³	2008	✓	✓	✓	✓	✓
Saarland ²	2009	✓	✓	✓	✓	✓
Sachsen	2009	✓	✓	–	–	–
Thüringen ²	2010	✓	✓	✓	✓	✓
Hamburg ^{2,3}	2011	✓	✓	✓	✓	✓
Hessen ²	2011	✓	✓	✓	✓	✓
Rheinland-Pfalz	2015	✓	✓	✓	✓	✓
Baden-Württemberg	2016	✓	✓	✓	✓	✓
Bayern	2016	✓	✓	✓	✓	✓
Berlin ²	2017	✓	✓	✓	✓	✓
Brandenburg ²	2017	✓	✓	✓	✓	✓
Niedersachsen	2017	✓	✓	✓	✓	✓

1 = Die Lehrpläne der Länder der Bundesrepublik Deutschland benennen keine eigenständigen prozessbezogenen Kompetenzbereiche, formulieren jedoch Kompetenzerwartungen, die den markierten Kompetenzen entsprechen.

2 = Neben den fünf prozessbezogenen Kompetenzen der KMK-Bildungsstandards benennen die Lehrpläne der Länder der Bundesrepublik Deutschland noch weitere Kompetenzbereiche.

3 = Verschiedene Kompetenzbereiche werden in den Lehrplänen der Länder der Bundesrepublik Deutschland zusammengefasst.

Grün eingefärbte Zeilen deuten darauf hin, dass die jeweiligen Lehrpläne der Länder der Bundesrepublik Deutschland vor den KMK-Bildungsstandards verabschiedet wurden.

Bezeichnung verwenden. Gemeint sind damit die in den Bildungsstandards angeführten Kompetenzen *Problemlösen*, *Kommunizieren*, *Argumentieren*, *Modellieren* sowie *Darstellen*.

Tabelle 3.3 zeigt, dass auch hinsichtlich der Formulierung der prozessbezogenen Kompetenzen eine gute Passung zwischen den Bildungsstandards und den Lehrplänen der Länder der Bundesrepublik Deutschland erkennbar ist. Ein Blick auf die vor Oktober 2004 in Kraft getretenen Lehrpläne in Bremen und Mecklenburg-Vorpommern zeigt, dass diese weniger explizit Bezug auf die prozessbezogenen Kompetenzen nehmen und keine eigenständigen Kompetenzbereiche mit entsprechenden Kompetenzerwartungen anführen. Jedoch sei darauf verwiesen, dass wesentliche Charakteristika prozessbezogener Kompetenzen auch in den Lehrplänen dieser Länder beschrieben werden, wenngleich nicht durch die explizite Formulierung von Kompetenzerwartungen.

Im bisherigen Verlauf dieses Abschnitts wurde gezeigt, dass zum Erhebungszeitpunkt im Mai 2019 die in den KMK-Bildungsstandards festgelegten inhalts- und prozessbezogenen Kompetenzen in den Lehrplänen der Länder für das Fach Mathematik der Grundschule (4. Jahrgangsstufe) nahezu flächendeckend explizit benannt und hierzu Kompetenzerwartungen beschrieben werden. Insgesamt ist die Entwicklung zu verzeichnen, dass sich die Lehrpläne der 16 Länder zunehmend einander angenähert haben.

Natürlich ist die Verankerung entsprechender Kompetenzen in den Lehrplänen kein Garant für die adäquate Implementierung in Schulen. Hier zeigt sich die typische Übergangsproblematik von intendiertem zu implementiertem Curriculum. Außerdem beschränkt sich die vorgenommene Analyse nur auf die Ebene der Kategorien, nicht auf darunterliegende Formulierungen von Teilbereichen oder einzelner Standards.

3.3 TIMSS-Rahmenkonzeption zur Erfassung mathematischer Kompetenzen

Ein wesentliches Charakteristikum der TIMS-Studien besteht darin, die Testinstrumente auf internationaler Ebene so zu gestalten, dass die mathematischen Kompetenzen der Teilnehmenden in den verschiedenen Staaten und Regionen möglichst präzise erfasst werden können. Die Analyse der erfassten Kompetenzen erlaubt dann auf Systemebene differenzierte Aussagen über spezifische Stärken und Schwächen der Schülerinnen und Schüler am Ende ihrer Grundschulzeit im internationalen Vergleich. Um anhand der in TIMSS gewonnenen Daten Erkenntnisse über die mathematischen Kompetenzen von Schülerinnen und Schülern aus Deutschland gewinnen zu können, sollten mit den eingesetzten Testaufgaben daher wesentliche Aspekte mathematischer Grundbildung in Deutschland (Abschnitt 3.2) abgebildet werden. Um dies zu überprüfen, wird in diesem Abschnitt die TIMSS-Rahmenkonzeption zur Erfassung mathematischer Kompetenzen von Viertklässlerinnen und Viertklässlern im Zyklus 2019 (Mullis & Martin, 2017) beschrieben. Dabei wird dargelegt, inwiefern eine Passung zu den für Grundschulkinder aus Deutschland festgeschriebenen Kompetenzerwartungen vorliegt.

Die TIMSS-Rahmenkonzeption unterscheidet einerseits zwischen drei *mathematischen Inhaltsbereichen* (Abschnitt 3.3.1) und andererseits zwischen drei *kognitiven Anforderungsbereichen* (Abschnitt 3.3.2). Bei den kognitiven Anforderungsbereichen sind im Vergleich zu den bisherigen TIMSS-Zyklen le-

diglich minimale Veränderungen vorgenommen worden. Da TIMSS 2019 jedoch den Übergang zu eTIMSS (siehe Kapitel 2 und 5 in diesem Band) darstellt, ist die Rahmenkonzeption vor allem bei den mathematischen Inhaltsbereichen aktualisiert worden und nun sowohl für digitale als auch für papierbasierte Erhebungsformate geeignet. So können perspektivisch die Vorteile computerbasierter Testformate zur Weiterentwicklung des TIMSS-Tests genutzt werden. Die Änderungen beschränken sich jedoch hauptsächlich auf Umsortierungen und Neustrukturierungen. Insgesamt ist also auch bei den Ausführungen zu den mathematischen Inhaltsbereichen eine Vergleichbarkeit zwischen TIMSS 2019 und den bisherigen TIMSS-Zyklen gegeben.

3.3.1 Mathematische Inhaltsbereiche

Die Testaufgaben der TIMS-Studie sind so gewählt, dass eine möglichst große Übereinstimmung mit der Gesamtheit der Curricula der teilnehmenden Staaten und Regionen vorhanden ist. Die für die Grundschule relevanten mathematischen Inhalte werden in der TIMSS-Rahmenkonzeption dazu in drei Inhaltsbereiche differenziert:

- Arithmetik (*Number*),
- Messen und Geometrie (*Measurement and Geometry*) sowie
- Daten (*Data*).

Jeder der drei Inhaltsbereiche ist wiederum in verschiedene Teilgebiete untergliedert. Beispielsweise enthält der Inhaltsbereich *Arithmetik* die Teilgebiete *Natürliche Zahlen*, *Terme*, *einfache Gleichungen und Beziehungen* sowie *Brüche und Dezimalzahlen*. Für diese Teilgebiete sind auf einer weiteren Ebene inhaltsbezogene Kompetenzerwartungen formuliert.

Tabelle 3.4 zeigt die drei Inhaltsbereiche mit ihren Teilgebieten und den darauf bezogenen Kompetenzerwartungen. Hier ist keine durchgängige Entsprechung zur Struktur der KMK-Bildungsstandards von 2004 gegeben. Beispielsweise sind Aufgaben, die in der Perspektive der Bildungsstandards zur Leitidee *Größen und Messen* zählen – wie Aufgaben zu Geldwerten, Massen (Gewichten), Längen, Flächen, Volumina –, bei TIMSS 2019 entweder den Inhaltsbereichen *Arithmetik* oder *Messen und Geometrie* zuzuordnen.

Darüber hinaus ist zu bedenken, dass es aufgrund der unterschiedlichen Curricula der an TIMSS 2019 für die vierte Jahrgangsstufe teilnehmenden 58 Staaten und der sechs sogenannten *Benchmark*-Regionen (Mullis, Martin, Foy, Kelly & Fishbein, 2020, Anhang) nicht möglich ist, einen Test zu entwickeln, der zu allen Curricula am Ende der vierten Jahrgangsstufe passt. Ein Minimalcurriculum aufzustellen, würde den Zielen von TIMSS nicht gerecht werden. Insofern wird es immer auch Teile der TIMSS-Konzeption und demzufolge der Testaufgaben geben, die im Curriculum eines teilnehmenden Staates nicht oder nur ansatzweise enthalten sind. Besonders augenfällig ist das im Kontext Deutschlands beim Teilgebiet *Brüche und Dezimalzahlen*. Die in den Testaufgaben abgebildeten Kompetenzen dieses Bereiches sind überwiegend nicht in den Grundschulcurricula in Deutschland verankert.

Tabelle 3.4: Inhaltsbereiche und darauf bezogene Kompetenzerwartungen

Inhaltsbereiche	Teilgebiete
Arithmetik	<p>1. Natürliche Zahlen</p> <ul style="list-style-type: none"> - Wissen über Stellenwerte abrufen (zweistellige bis sechsstellige Zahlen); natürliche Zahlen mithilfe von Worten, Zeichnungen (Diagramme, Zahlenstrahl) und Symbolen repräsentieren; Zahlen ordnen - Addieren und subtrahieren (max. vierstellige Zahlen), auch in einfachen Kontextaufgaben - Multiplizieren (max. dreistellig mal einstellig oder zweistellig mal zweistellig) und dividieren (max. dreistellig durch einstellig), auch in einfachen Kontextaufgaben - Aufgaben mit geraden und ungeraden Zahlen oder mit Vielfachen und Teilern lösen; Zahlen runden; Überschlagsrechnungen durchführen - Zwei oder mehr Eigenschaften von Zahlen oder Operationen nutzen, um Kontextaufgaben zu lösen <p>2. Terme, einfache Gleichungen und Beziehungen</p> <ul style="list-style-type: none"> - Fehlende Zahlen oder Operationen in einem Zahlensatz finden (z.B. $17 + __ = 29$) - Terme oder Gleichungen zur Darstellung von Problemsituationen mit Unbekannten identifizieren oder aufschreiben - Beziehungen in klar definierten Mustern erkennen und nutzen (z.B. die Beziehung zwischen benachbarten Zahlen beschreiben; Zahlenpaare erzeugen, die einer vorgegebenen Regel folgen) <p>3. Brüche und Dezimalzahlen</p> <ul style="list-style-type: none"> - Brüche als Teile eines Ganzen, als Teil mehrerer Ganzer erkennen; Brüche durch Worte, Symbole oder geometrisch darstellen; einfache Brüche vergleichen und ordnen; einfache Brüche addieren und subtrahieren, auch in Kontextaufgaben (Nenner: 2, 3, 4, 5, 6, 8, 10, 12, 100) - Dezimalzahlverständnis nachweisen, einschließlich des Darstellens von Dezimalzahlen durch Worte, Symbole oder geometrisch; Dezimalzahlen vergleichen, ordnen und runden; Dezimalzahlen addieren und subtrahieren, auch in Kontextaufgaben (mit einer oder zwei Nachkommastellen beim Rechnen mit Geld)
Messen und Geometrie	<p>1. Messen</p> <ul style="list-style-type: none"> - Längen messen und schätzen (mm, cm, m, km); Aufgaben zum Größenbereich Längen lösen - Aufgaben zu den Größenbereichen Masse (g, kg), Volumen (ml, l) und Zeit (min, h) lösen; geeignete Einheiten identifizieren und Skalen lesen - Aufgaben zum Umfang von Vielecken, zum Flächeninhalt von Rechtecken, zum Flächeninhalt von Figuren, die (von Teilen) von Einheitsquadraten abgedeckt sind sowie zum Volumen von Figuren, die mit Einheitswürfeln ausgelegt sind, lösen <p>2. Geometrie</p> <ul style="list-style-type: none"> - Zueinander parallele und zueinander senkrechte Geraden identifizieren und zeichnen; rechte Winkel und Winkel, die kleiner oder größer sind als der rechte Winkel identifizieren und zeichnen; Winkel gemäß ihrer Größe vergleichen - Elementare Eigenschaften nutzen (einschließlich Achsen- und Drehsymmetrie), um zweidimensionale geometrische Figuren zu beschreiben, zu vergleichen und zu erzeugen (Kreise, Dreiecke, Vierecke und andere Vielecke) - Elementare Eigenschaften nutzen, um dreidimensionale geometrische Figuren zu beschreiben und zu vergleichen (Würfel, Quader, Kegel, Pyramide und Kugeln) und sie zu ihren zweidimensionalen Repräsentationen in Beziehung zu setzen
Daten	<p>1. Daten lesen, interpretieren und darstellen</p> <ul style="list-style-type: none"> - Daten aus Tabellen, Piktogrammen, Balken-, Linien- und Kreisdiagrammen entnehmen und sie interpretieren - Daten organisieren und darstellen, um dadurch Aufgaben lösen zu können <p>2. Daten zum Lösen von Aufgaben nutzen</p> <ul style="list-style-type: none"> - Daten nutzen, um Fragen zu beantworten, die über das bloße Entnehmen von Daten hinausgehen (z.B. Probleme lösen und datengestützt rechnen, Daten aus zwei oder mehr Quellen vergleichen, datengestützte Schlussfolgerungen ziehen)

3.3.2 Kognitive Anforderungsbereiche

Die Rahmenkonzeption von TIMSS 2019 beinhaltet neben den inhaltsbezogenen mathematischen Kompetenzen auch Formulierungen für die Denkprozesse, welche für die Aufgabenbearbeitung erforderlich sind. TIMSS unterscheidet drei kognitive Anforderungsbereiche:

- *Reproduzieren*: Reproduzieren von Wissen, Fertigkeiten und Grundvorstellungen (*Knowing*),
- *Anwenden*: Anwenden von Wissen, Fertigkeiten und Grundvorstellungen beim Bearbeiten von Standardaufgaben (*Applying*) sowie
- *Problemlösen*: Lösen von komplexeren, anspruchsvolleren Aufgabenstellungen (*Reasoning*).

Durch diese Unterteilung der Anforderungsbereiche soll ein möglichst umfassendes Bild der kognitiven Leistungsfähigkeit der Lernenden gewonnen werden. Dies kann nur gelingen, wenn sich Testitems nicht nur auf einen kognitiven Anforderungsbereich wie etwa *Reproduzieren* beschränken (Mullis & Martin, 2017).

Die kognitiven Aktivitäten der drei Anforderungsbereiche werden im Folgenden anhand beispielhafter Beschreibungen konkretisiert. Tabelle 3.5 zeigt zunächst die kognitiven Aktivitäten im Anforderungsbereich *Reproduzieren*.

Tabelle 3.5: Kognitive Aktivitäten im Anforderungsbereich *Reproduzieren*

Kognitive Aktivität <i>Reproduzieren</i>	Beschreibung
Standardwissen abrufen	Definitionen, Begriffe, Zahleigenschaften, Maßeinheiten, geometrische Eigenschaften und Notationen abrufen
Einfache mathematische Objekte und Beziehungen erkennen	Zahlen, Terme, Mengen und Figuren erkennen; mathematisch gleichwertige Einheiten erkennen (z.B. gleichwertige Brüche, Dezimalzahlen und Prozentangaben; verschieden angeordnete einfache geometrische Figuren)
Klassifizieren/Ordnen	Zahlen, Terme, Mengen und Figuren nach gemeinsamen Eigenschaften klassifizieren
Rechnen	Algorithmische Prozeduren für die vier Grundrechenarten mit natürlichen Zahlen, Brüchen, Dezimalzahlen und ganzen Zahlen ausführen; einfache algebraische Prozeduren ausführen
Informationen ablesen	Graphen, Tabellen, Texten oder anderen Quellen Informationen entnehmen
Messen	Messinstrumente nutzen und geeignete Maßeinheiten wählen

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Tabelle 3.6 stellt die kognitiven Aktivitäten des Bereiches *Anwenden* dar, in dem über das Abrufen von Wissen hinaus auch der situative Gebrauch in vielfältigen Anforderungssituationen gefordert ist. Tabelle 3.7 zeigt schließlich die kognitiven Aktivitäten des Bereiches *Problemlösen*, die vor allem das systematische und logische Denken sowie schlussfolgernd-begründende Aktivitäten betonen.

Die kognitiven Aktivitäten der drei dargestellten Anforderungsbereiche stimmen bis auf geringfügige Unterschiede mit den Anforderungsbereichen der KMK-Bildungsstandards (KMK, 2005) überein. Nennenswerte Unterschiede bestehen beispielsweise hinsichtlich der Aktivität *Einfache mathematische Objekte und Beziehungen erkennen* des TIMSS-Anforderungsbereiches *Reproduzieren*, die in

Tabelle 3.6: Kognitive Aktivitäten im Anforderungsbereich *Anwenden*

Kognitive Aktivität <i>Anwenden</i>	Beschreibung
Effiziente und geeignete Operationen und Strategien wählen	Effiziente bzw. geeignete Operationen, Strategien und Werkzeuge für das Lösen von Standardaufgaben bestimmen
Informationen darstellen und modellieren	Daten in Tabellen oder Graphen darstellen; Gleichungen, Ungleichungen, geometrische Figuren oder Diagramme erstellen, die Problemsituationen modellieren; einander entsprechende Repräsentationen eines gegebenen mathematischen Objektes oder einer Beziehung erzeugen
Durchführen	Mithilfe bekannter mathematischer Konzepte und Prozeduren Strategien und Operationen für das Lösen von Aufgaben nutzen

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Tabelle 3.7: Kognitive Aktivitäten im Anforderungsbereich *Problemlösen*

Kognitive Aktivität <i>Problemlösen</i>	Beschreibung
Analysieren	Beziehungen zwischen Zahlen, Termen, Mengen und Figuren bestimmen, beschreiben oder nutzen
Integrieren und Kombinieren	Verschiedene Wissens Elemente, zueinander zugehörige Darstellungen und Prozeduren für das Lösen von Problemen miteinander verknüpfen
Evaluiieren	Alternative Problemlösestrategien und Lösungen evaluieren
Schlussfolgern	Korrekte Schlussfolgerungen auf der Grundlage von Informationen und Evidenzen ziehen
Verallgemeinern	Verallgemeinernde Aussagen treffen, die Beziehungen in breit anwendbarer Weise darstellen
Begründen	Mathematische Begründungen für eine Strategie oder Lösung geben

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

den KMK-Bildungsstandards eher Anforderungsbereich II (siehe Abschnitt 3.2) zuzuordnen wäre. Ferner fällt die dem TIMSS-Anforderungsbereich *Anwenden* zugeordnete Aktivität *Effiziente und geeignete Operationen und Strategien wählen* eher in Anforderungsbereich I der Bildungsstandards.

Die allgemeinen mathematischen beziehungsweise prozessbezogenen Kompetenzerwartungen der Bildungsstandards und der Lehrpläne werden in der TIMSS-Rahmenkonzeption nicht explizit gelistet, sind aber insbesondere in den kognitiven Anforderungsbereichen *Anwenden* und vor allem *Problemlösen* partiell enthalten.

3.3.3 Aufgabenbeispiele

Die TIMSS-Rahmenkonzeption gibt drei mathematische Inhaltsbereiche und drei kognitive Anforderungsbereiche vor. Die Gesamtheit der Testaufgaben von TIMSS 2019 deckt die daraus resultierenden neun Kombinationen durch entsprechende Aufgaben ab. Abbildung 3.1 zeigt zu jeder dieser Kombinationen eine Beispielaufgabe.

Abbildung 3.1: Beispielaufgaben zu Inhaltsbereichen und kognitiven Anforderungsbereichen

Reproduzieren

Anwenden

Problemlösen

Arithmetik

Bei welcher Zahl steht eine 7 an der Hunderterstelle und eine 6 an der Einerstelle?

☐ A 167
☐ B 176
☒ C 716
☐ D 761

(int. .79/nat. .81)

Bei dieser Aufgabe wird Grundwissen zum Stellenwertverständnis erhoben, indem nach der Position zweier unterschiedlicher Ziffern einer dreistelligen Zahl gefragt wird (Standardwissen abrufen).

Merle verschenkt 48 Aufkleber. Sie schenkt 4 Freunden eine gleich große Anzahl.

Welche Rechnung zeigt, wie viele Aufkleber Merle jedem Freund schenkt?

☐ A $48 + 4$
☐ B $48 - 4$
☐ C $48 \cdot 4$
☒ D $48 : 4$

(int. .70/nat. .66)

Bei dieser Aufgabe soll einer textlich dargebotenen Aufgabe der entsprechende Term zugeordnet werden (geeignete Operationen wählen).

Ein Lehrer will 30 Schüler so in Gruppen einteilen:

- In jeder Gruppe soll dieselbe Anzahl Schüler sein **und**
- in jeder Gruppe soll eine ungerade Anzahl Schüler sein.

Zeige zwei Möglichkeiten, wie der Lehrer die Gruppen bilden kann.

1. Möglichkeit

Anzahl der Gruppen:

Anzahl der Schüler in jeder Gruppe:

2. Möglichkeit

Anzahl der Gruppen:

Anzahl der Schüler in jeder Gruppe:

(int. .24/nat. .25)

Hier soll im Rahmen einer beschriebenen Realsituation die Aufgabe gelöst werden, die Zahl 30 auf zwei unterschiedliche Weisen als Produkt mit einem ungeraden Faktor darzustellen. Dazu müssen Beziehungen zwischen Zahlen und Termen genutzt werden (analysieren).

Messen/Geometrie

Welche Maßeinheiten gehören zu diesen Messwerten?

Ziehe eine Maßeinheit zu jedem Messwert.

Das Auto wiegt 1400

In den Eimer passen 10

Der Bleistift wiegt 5

(int. .67/nat. .79)

Bei dieser Aufgabe sollen drei der vier Maßeinheiten g, kg, l und ml drei Messwerten zugeordnet werden. Hierzu ist Grundwissen über die Verwendung und die Größenordnung von Maßeinheiten erforderlich (messen).

Zeichne die Figur fertig, sodass die gestrichelte Linie eine Spiegelachse ist.

Klicke dafür Quadrate in dem Gitter an.

(int. .73/nat. .85)

Hier soll eine vorliegende Figur durch eine Spiegelung an einer horizontalen Achse komplettiert werden. Das Wissen zur Erstellung achsensymmetrischer Figuren ist hierbei anzuwenden (durchführen).

Die 2 Birnen wiegen gleich viel wie 4 Clementinen. Wie viel wiegt 1 Birne?

☐ A 480 g
☐ B 240 g
☒ C 120 g
☐ D 60 g

(int. .57/nat. .52)

Hier soll das Gewicht einer Birne aus dem Gewicht zweier Birnen bzw. von vier Clementinen ermittelt werden, deren Einzelgewicht angegeben ist. Dazu müssen Schlussfolgerungen auf Grundlage der vorliegenden Einzelnformationen angestellt werden (schlussfolgern).

Daten

Wasserhöhe vom Stausee

Wie war die Wasserhöhe in Woche 8?

Antwort: m

(int. .77/nat. .71)

Hier soll ein bestimmter Wert aus einem Liniendiagramm abgelesen werden, wozu Grundwissen im Umgang mit Diagrammen notwendig ist (Informationen ablesen).

Sylvie hat die Anzahl der Autos aufgeschrieben, die jeden Morgen durch ihre Straße gefahren sind.

Tag	Anzahl der Autos
Montag	8
Dienstag	5
Mittwoch	7
Donnerstag	10
Freitag	12

Sie hat angefangen, aus ihren Daten ein Diagramm zu zeichnen.

Welche Zahlen sollte Sylvie an die waagerechten Linien von ihrem Diagramm schreiben?

Schreibe die Zahlen in die Kästchen in Sylvies Diagramm.

Anzahl der Autos

(int. .40/nat. .35)

Bei dieser Aufgabe sind durch eine Tabelle vorgegebene Informationen in ein Säulendiagramm insofern zu übertragen, als dass die noch fehlende Beschriftung auf der y-Achse einzutragen ist (Informationen darstellen).

Gewicht der Tiere

Tier	Gewicht (kg)
Gepard	50
Löwe	100
Leopard	75

Erstelle ein Piktogramm vom Gewicht von jedem Tier.

Ziehe Symbole zum Piktogramm, um es fertigzustellen. Das Gewicht des Gepards wurde schon eingetragen.

Tier	Gewicht (kg)
Gepard	
Löwe	
Leopard	

Legende: = 50 kg

(int. .67/nat. .68)

Bei dieser Aufgabe sind die Daten aus einer Tabelle und vorgegebene ikonische Darstellungen miteinander zu verknüpfen; dabei ist die Darstellung für 25 kg als solche noch zu interpretieren. Verschiedene Darstellungen sind miteinander zu verknüpfen (integrieren und kombinieren).

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

Abdruck und Nutzung der Aufgaben nur mit ausdrücklicher Genehmigung der IEA: <https://www.iea.nl/publications/form/iea-permission-request-form>

3.4 Der TIMSS-Test zur Erfassung der mathematischen Kompetenzen

Im Folgenden werden zentrale Aspekte des TIMSS-Tests zur Erfassung mathematischer Kompetenzen von Viertklässlerinnen und Viertklässlern beschrieben. Dazu wird zunächst der Testaufbau erläutert (Abschnitt 3.4.1) und danach der Aspekt der curricularen Validität von Testaufgaben betrachtet (Abschnitt 3.4.2). In Abschnitt 3.4.3 wird das in dieser Studie verwendete Kompetenzstufenmodell zur Beschreibung mathematischer Kompetenzen skizziert.

3.4.1 Testentwicklung und Testaufbau

Die Testaufgaben wurden in einem kooperativen Prozess entwickelt, an dem neben dem *TIMSS & PIRLS International Study Center* am Boston College sowie den nationalen Studienleitungen auch Mathematikdidaktikerinnen und Mathematikdidaktiker beteiligt waren (Mullis & Martin, 2017). Der Entwicklungsprozess begann mit Aufgaben- und Auswertungsvorschlägen durch die nationalen Studienleitungen und die beteiligten Fachdidaktikerinnen und Fachdidaktiker. Diese wurden von weiteren Expertinnen und Experten aus dem *Item Review Committee* (Mullis & Martin, 2017) begutachtet und gegebenenfalls überarbeitet. Anschließend wurde insbesondere auf eine bildungssprachlich adäquate Übersetzung der Testaufgaben geachtet. Nachdem ein vorläufiger Aufgabenpool ausgewählt wurde, erfolgte ein deutschlandweiter Feldtest, aus dem die gewonnenen Erkenntnisse in die zentral erfolgte Überarbeitung der Testaufgaben geflossen sind.

Der TIMSS-Test 2019 zur Erfassung mathematischer Kompetenzen von Viertklässlerinnen und Viertklässlern bestand aus 175 Items. Aufgrund nicht geeigneter statistischer Kennwerte mussten vier Items aus der Skalenbildung ausgeschlossen werden. Somit resultieren die in diesem Kapitel dargestellten Ergebnisse aus den Bearbeitungen von 171 Mathematikaufgaben. Zur besseren Vergleichbarkeit der TIMSS-Erhebungsdaten 2019 von den Daten bisheriger TIMSS-Zyklen wurden 92 von 171 Items dem TIMSS-Test von 2015 entnommen. Da diese Untersuchung von 2015 wiederum in ähnlichem Umfang Testaufgaben des vorangegangenen Zyklus beinhaltete, kann von einer weitgehend gewährleisteten Vergleichbarkeit der Daten über die Erhebungszeitpunkte hinweg ausgegangen werden (siehe Kapitel 5 in diesem Band zur Adjustierung der Moduseffekte durch den Wechsel von der papierbasierten auf die computerbasierte Testung).

Hinsichtlich der Antwortformate der 171 Items ist festzustellen, dass jeweils etwa die Hälfte der Aufgaben in einem *Multiple-Choice*-Format beziehungsweise in einem offenen Format zu beantworten waren.

Tabelle 3.8 zeigt für die vier Studienzyklen von 2007 bis 2019 die Verteilung der Aufgaben auf die jeweiligen mathematischen Inhaltsbereiche. Der Tabelle ist zu entnehmen, dass in 2019 *Arithmetik* mit 49 Prozent der Aufgaben den größten Teilbereich darstellt. Daneben entstammen 30 Prozent der Testaufgaben dem Bereich *Messen und Geometrie*, während 21 Prozent dem Inhaltsbereich *Daten* zuzuordnen sind. Die Verteilung der Aufgaben über die drei Inhaltsbereiche stimmt mit denen in den drei vergangenen TIMSS-Zyklen in der Tendenz überein, wenngleich die Anteile der Items aus den Bereichen *Arithmetik* 4 Prozentpunkte und *Messen und Geometrie* 3 Prozentpunkte niedriger und aus dem Bereich *Daten* etwa 7 Prozentpunkte höher waren als in den Studienzyklen

zuvor. Ein wesentlicher Grund für diese Verschiebung ist eine Änderung der Vorgaben der TIMSS-Rahmenkonzeption, die 2015 die Verteilung 50 – 35 – 15 für die drei Inhaltsbereiche vorgab (Mullis & Martin, 2013), während es 2019 eine Vorgabe von 50 – 30 – 20 gab.

Tabelle 3.8: Verteilung der Testaufgaben auf die mathematischen Inhaltsbereiche

Mathematische Inhaltsbereiche	2007		2011		2015		2019	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Arithmetik	93	52	90	51	89	53	83	49
Messen/Geometrie	60	34	61	34	56	33	52	30
Daten	26	15	26	15	24	14	36	21
Gesamt	179	100	177	100	169	100	171	100

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Wie bereits erwähnt, sind in den Testaufgaben von TIMSS 2019 die drei kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen* repräsentiert. Dabei entstammt mit 43 Prozent der Aufgaben der überwiegende Teil der Testaufgaben dem Bereich *Anwenden*. *Reproduzieren* ist hingegen bei 35 Prozent der Aufgaben und *Problemlösen* bei 22 Prozent der Aufgaben gefordert. Tabelle 3.9 illustriert für die vier TIMSS-Zyklen die Verteilung der Testaufgaben auf die drei kognitiven Anforderungsbereiche.

Tabelle 3.9: Verteilung der Testaufgaben auf die kognitiven Anforderungsbereiche

Kognitive Anforderungsbereiche	2007		2011		2015		2019	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Reproduzieren	69	39	71	40	64	38	59	35
Anwenden	70	39	71	40	72	43	74	43
Problemlösen	40	22	35	20	33	20	38	22
Gesamt	179	100	177	100	169	100	171	100

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Auch die Verteilung der Testaufgaben auf die drei kognitiven Anforderungsbereiche unterscheidet sich beim Vergleich der vier Studienzyklen nur minimal. Die Werte liegen jeweils in der Nähe der vorgegebenen Relation 40 – 40 – 20 für *Reproduzieren*, *Anwenden* und *Problemlösen*.

Die TIMS-Studie 2019 beinhaltetete zusätzlich eine Reihe neuer, komplexer Erkundungsaufgaben. Diese wurden so zusammengestellt, dass sie lebensweltliche oder Laborsituationen simulierten, in denen die Schülerinnen und Schüler höhere mathematische Fähigkeiten sowie Kenntnisse und Fertigkeiten integriert zur Lösung von Problemstellungen heranziehen konnten. Die insgesamt 29 Items würden zu einer leichten Erhöhung der Anteile der Prozentsätze in den Bereichen *Anwenden* und *Problemlösen* führen. Da diese Items im vorliegenden Zyklus erstmals eingesetzt wurden, wurde darauf verzichtet, sie bereits in diesem Studienzyklus bei der Berechnung der Kompetenzwerte zu berücksichtigen. Gleichwohl konnten Berechnungen die Konstruktvalidität der Aufgaben nachweisen. Daher ist eine stärkere Berücksichtigung der Aufgaben im nachfolgenden Zyklus von TIMSS im Jahr 2023 angedacht.

3.4.2 Curriculare Validität

Die Ausführungen der Abschnitte 3.3.1 und 3.3.2 haben eine zufriedenstellende Passung der TIMSS-Rahmenkonzeption 2019 mit den curricularen Vorgaben der KMK-Bildungsstandards von 2004 gezeigt – wie es auch bei den bisherigen TIMSS-Zyklen (Selter, Walter, Walther & Wendt, 2016) der Fall war.¹

Der Aspekt der curricularen Validität der entwickelten Testitems ist – wie in Abschnitt 3.3.1 bereits angedeutet – mit einer für internationale Schulleistungsstudien typischen Schwierigkeit verbunden: Die Testaufgaben sollen den Curricula möglichst vieler teilnehmender Staaten und Regionen in angemessenem Umfang genügen. Bei der Verschiedenheit der Curricula bleibt es nicht aus, dass sich in den an TIMSS beteiligten Staaten jeweils nur ein bestimmter Teil der Testaufgaben auf das jeweilige Curriculum bezieht. Diese Aufgaben werden als curricular valide Aufgaben bezeichnet. Der Rest der Testaufgaben enthält inhaltliche Aspekte, die im Curriculum nicht vorkommen. Die Überprüfung der curricularen Validität wird für jeden teilnehmenden Staat beziehungsweise jede teilnehmende Region separat von einem aus nationalen Fachdidaktikerinnen und Fachdidaktikern zusammengestellten Expertinnen- und Expertenteam vorgenommen. Dieser Arbeitsschritt wird in den TIMSS-Zyklen als *Test-Curriculum-Matching-Analysis* (TCMA) bezeichnet.

Für die Bundesrepublik Deutschland wurde jede Testaufgabe – wie in den vergangenen Zyklen – hinsichtlich ihrer Vereinbarkeit mit den KMK-Bildungsstandards von 2004 (KMK, 2005), dem Lehrplan Mathematik für die Grundschule des bevölkerungsstärksten Landes Nordrhein-Westfalen (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008) und einer Auswahl vielfach eingesetzter Schulbücher geprüft.

Die TCMA ergab für TIMSS 2019 einen Anteil von 23 Prozent curricular nicht valider Aufgaben. Vergleicht man die Anteile curricular nicht valider Testaufgaben über die vier Studienzyklen hinweg, so liegt deren Anteil von 25 Prozent bei TIMSS 2007 (Walther, Selter, Bosen & Bos, 2008), über 21 Prozent bei TIMSS 2011 (Selter, Walther, Wessel & Wendt, 2012) und 19 Prozent bei TIMSS 2015 (Selter et al., 2016) bei der gegenwärtigen Studie etwas höher als bei den

¹ Wesentliche Teile dieses Abschnittes sind bereits im Kapitel ‚Ziele, Anlage und Durchführung in der *Trends in International Mathematics and Science Study* (TIMSS)‘ im Berichtsband von TIMSS 2015 (Selter et al., 2016) bearbeitet worden. Die Ausführungen im vorliegenden Berichtsband sind daher an den Ausführungen zum Studienzyklus von 2015 orientiert.

letzten beiden Zyklen, aber niedriger als 2007. Bei der TCMA der Aufgaben für TIMSS 2019 wurde jede einzelne Aufgabe hinsichtlich der folgenden drei Aufgabenmerkmale auf ihre curriculare Validität geprüft:

- Grafische Darstellungsform,
- Mathematische Begriffe, Verfahren und Sachverhalte sowie
- Bezeichnungen.

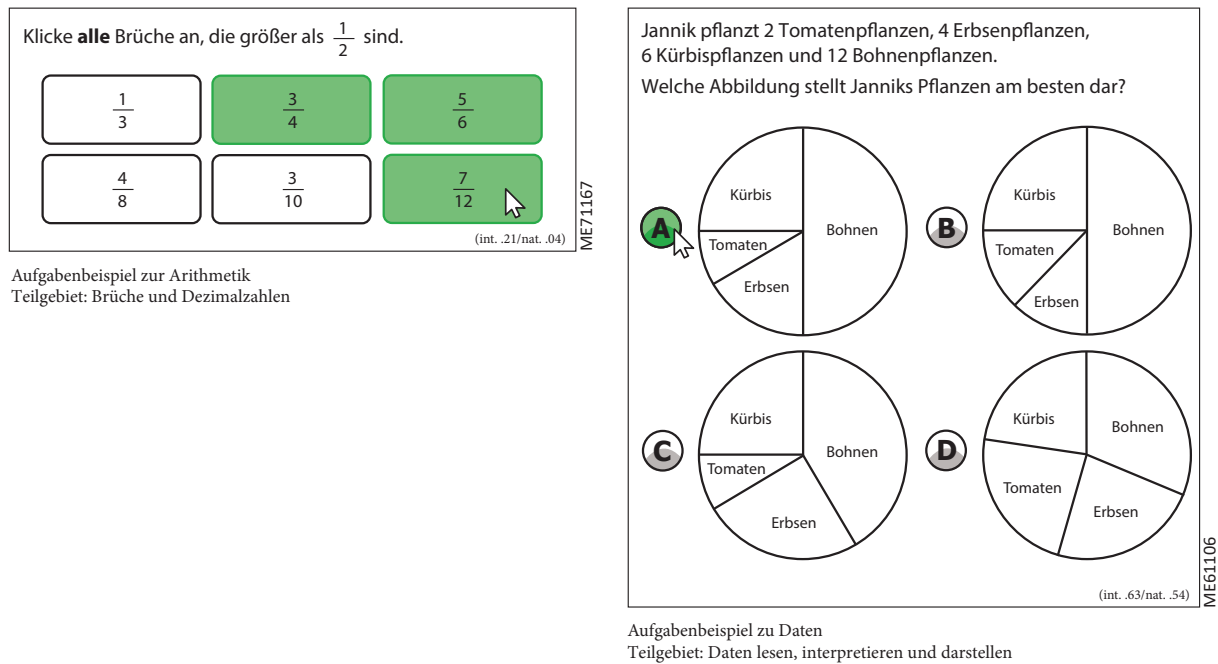
Tabelle 3.10: Merkmale von curricular nicht validen Testaufgaben in Mathematik

Aufgabenmerkmale	curricular nicht valide Inhalte
Grafische Darstellungsform	Kreisdiagramm, Liniendiagramm
Mathematische Begriffe, Verfahren und Sachverhalte	<p><i>Brüche:</i> Bruchteile, Bruchvergleich (kleiner, gleich, größer), Addition oder Subtraktion von Brüchen</p> <p><i>Dezimalbrüche:</i> Vergleich von Dezimalbrüchen (kleiner, gleich, größer), Addition oder Subtraktion von Dezimalbrüchen</p> <p><i>Gleichungen:</i> der Form $a : b = b$ mit b als Unbekannter, Rechnen mit Klammern</p> <p><i>Winkel:</i> Maßzahlen (90°, 180°), Vergleich von Winkelgrößen (kleiner, größer)</p>
Bezeichnungen	drehsymmetrisch, Quadratmeter, Quadratzentimeter und Seite A

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Da sich im Vergleich zum Zyklus von TIMSS 2015 bei der TCMA zu TIMSS 2019 keine nennenswerten Unterschiede bei den in der Tabelle gelisteten ‚kritischen‘ Aufgabenmerkmalen curricular nicht valider Aufgaben ergeben haben, wird auf deren detaillierte Erläuterung im TIMSS-Berichtsband 2011 verwiesen (Selter et al., 2012). Abbildung 3.2 zeigt zwei konkrete Beispiele curricular nicht valider Testaufgaben.

Abbildung 3.2: Beispiele curricular nicht valider Testaufgaben in Mathematik

Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an.
Abdruck und Nutzung der Aufgaben nur mit ausdrücklicher Genehmigung der IEA: <https://www.iea.nl/publications/form/iea-permission-request-form>

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Tabelle 3.11 zeigt die Verteilung curricular nicht valider Testaufgaben auf die drei mathematischen Inhaltsbereiche.² Aus den Daten geht hervor, dass erneut mehr als ein Viertel der Aufgaben aus *Arithmetik* (27%) nicht mit den curricularen Vorgaben in Einklang zu bringen sind. Der vergleichsweise hohe Anteil curricular nicht valider Testaufgaben in der *Arithmetik* erklärt sich dadurch, dass das zu *Arithmetik* zählende Themenfeld *Brüche und Dezimalzahlen* Aufgaben enthält, die nicht zum Standard des Grundschulcurriculums der Länder der Bundesrepublik Deutschland gehören. Es sind vor allem Aufgaben mit ‚reinen‘, also nicht auf Größen bezogenen Brüchen und Dezimalzahlen. Bei *Daten* liegt der Anteil

Tabelle 3.11: Anteile curricular nicht valider Testaufgaben nach mathematischen Inhaltsbereichen

Mathematische Inhaltsbereiche	2007			2011			2015			2019		
	Gesamt <i>n</i>	nicht valide		Gesamt <i>n</i>	nicht valide		Gesamt <i>n</i>	nicht valide		Gesamt <i>n</i>	nicht valide	
		<i>n</i>	%		<i>n</i>	%		<i>n</i>	%		<i>n</i>	%
Arithmetik	93	20	22	90	22	24	89	23	26	83	22	27
Messen/Geometrie	60	20	33	61	11	18	56	5	9	52	7	13
Daten	26	5	19	26	4	15	24	4	17	36	10	28
Gesamt	179	45	25	177	37	21	169	32	19	171	39	23

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

² Aufgrund nachträglicher Ausschlüsse einzelner Items haben sich im Vergleich zur Berücksichtigung von TIMSS 2007 (Walther et al., 2008) leichte Verschiebungen der Prozentwerte zur curricularen Validität der Testaufgaben ergeben.

sogar noch höher, nämlich bei 28 Prozent. Dieser Umstand ist auf zehn Aufgaben zurückzuführen, die Kreis- beziehungsweise Liniendiagramme thematisierten. Im Inhaltsbereich *Messen und Geometrie* hingegen liegt der Anteil curricular nicht valider Aufgaben bei lediglich 13 Prozent.

In Tabelle 3.12 ist die Verteilung curricular nicht valider Testaufgaben auf die drei kognitiven Anforderungsbereiche dargestellt. Die Verteilung der 23 Prozent curricular nicht validen Aufgaben über die drei Anforderungsbereiche sieht folgendermaßen aus: Aufgaben des Anforderungsbereiches *Reproduzieren* wurden mit 25 Prozent als curricular nicht valide eingeschätzt, im Bereich *Anwenden* sind es 26 Prozent und im Bereich *Problemlösen* entsprechen 13 Prozent der Aufgaben nicht den curricularen Vorgaben.

Tabelle 3.12: Anteile curricular nicht valider Testaufgaben nach kognitiven Anforderungsbereichen

Kognitive Anforderungsbereiche	2007			2011			2015			2019		
	Gesamt		nicht valide	Gesamt		nicht valide	Gesamt		nicht valide	Gesamt		nicht valide
	<i>n</i>	<i>n</i>	%	<i>n</i>	<i>n</i>	%	<i>n</i>	<i>n</i>	%	<i>n</i>	<i>n</i>	%
Reproduzieren	69	25	36	71	21	30	64	14	22	59	15	25
Anwenden	70	13	19	71	13	18	72	13	18	74	19	26
Problemlösen	40	7	18	35	3	9	33	5	15	38	5	13
Gesamt	179	45	25	177	37	21	169	32	19	171	39	23

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

In den vorangegangenen Studien wurden in gesonderten Berechnungen die Leistungsmittelwerte ermittelt, nachdem die für die Schülerinnen und Schüler in Deutschland curricular nicht validen Aufgaben aus den Berechnungen eliminiert wurden (Bonsen, Lintorf, Bos & Frey, 2008; Selter et al., 2012; Selter et al., 2016). Hierbei ergaben sich für Deutschland jeweils nominelle Verbesserungen um wenige Rangplätze. Allerdings ließ sich insgesamt zur Frage der curricularen Validität festhalten, dass das Abschneiden Deutschlands auf der TIMSS-Leistungsskala nicht auf das Testdesign zurückgeführt werden kann. „Vergleicht man die Rangreihenfolge der Staaten bei den individuellen Subtests mit der Rangreihe beim Gesamtest, so zeigen sich keine bedeutsamen Unterschiede“ (Bonsen et al., 2008, S. 32). Somit wurde hier auf die Darstellung für TIMSS 2019 verzichtet.

3.4.3 Kompetenzstufen in Mathematik

Wie in den vorangegangenen TIMS-Studien werden die Leistungsstände der teilnehmenden Viertklässlerinnen und Viertklässler zunächst über bestimmte statistische Kennwerte beschrieben. Diese erlauben zwar Aussagen über relative Unterschiede der Leistungsstände, geben aber keine inhaltlichen Einblicke in das mathematische Wissen und Können der teilnehmenden Lernenden. Genau dies wird durch das in TIMSS eingesetzte *Benchmark*-Verfahren ermöglicht. Damit können die Personenfähigkeitswerte an konkreten Testaufgaben verdeutlicht werden.

Durch ein bei Wendt et al. (2016) näher dargestelltes Verfahren wird diese Skala bei TIMSS in fünf Intervalle untergliedert. Die Teilungspunkte, durch die die Intervalle erzeugt werden, heißen *Benchmarks*. Jedes der Intervalle kann durch spezifische mathematische Kompetenzen beschrieben werden, die zur Lösung entsprechender Aufgaben auf dieser Stufe erforderlich sind. Die fünf Intervalle werden als aufeinander aufbauende *Kompetenzstufen* (KS) bezeichnet.

Die inhaltliche Charakterisierung einer solchen Kompetenzstufe entsteht wie folgt: Ein Testitem wird beispielsweise der Kompetenzstufe III zugerechnet, wenn mindestens 65 Prozent der Schülerinnen und Schüler, die den unteren *Benchmark*-Wert dieser Stufe erreichen, das Item korrekt lösen, während dies für höchstens 50 Prozent der Kinder auf Kompetenzstufe II der Fall ist. Aus den so gewonnenen Items werden in einem weiteren Analyseschritt die charakterisierenden Kompetenzanforderungen ermittelt.

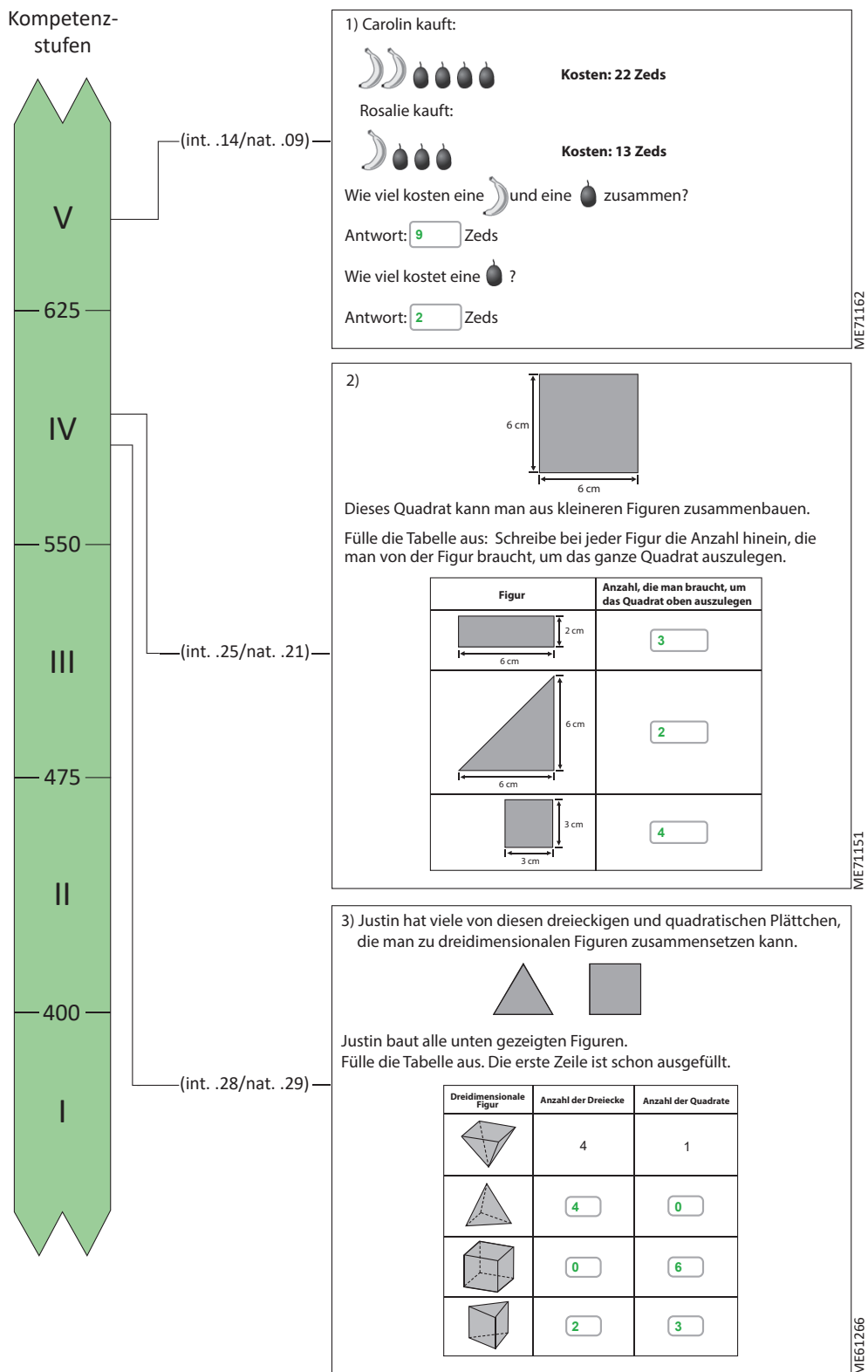
In Tabelle 3.13 sind wesentliche Charakteristika der mathematischen Fähigkeiten und Fertigkeiten beschrieben, die für jede der fünf Kompetenzstufen als typisch gelten können. Eine ausführliche Beschreibung des Kompetenzstufenmodells findet sich im internationalen Berichtsband (Mullis et al., 2020, Exhibits 1.7.1–1.13.4). Bis auf kleine Veränderungen stimmen die angegebenen mathematischen Fähigkeiten und Fertigkeiten mit denen in den bisherigen TIMSS-Zyklen überein.

Tabelle 3.13: Beschreibung der fünf Kompetenzstufen für die Gesamtskala Mathematik

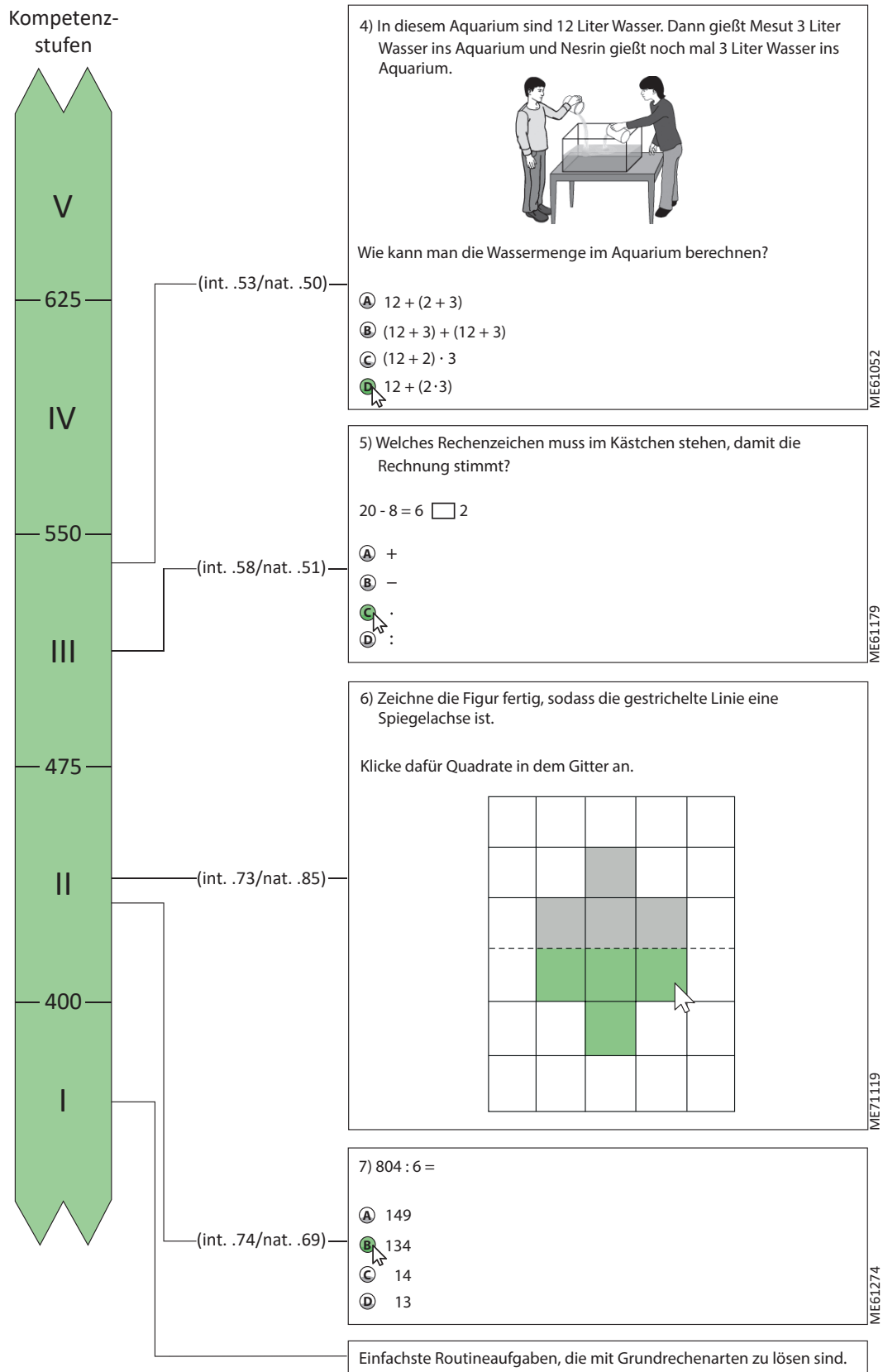
Kompetenzstufe V (ab 625): fortgeschritten
<i>Die Schülerinnen und Schüler können ihre mathematischen Fertigkeiten und Fähigkeiten für das Lösen von relativ komplexen Problemen anwenden und ihre Begründungen erklären.</i> Sie können vielfältige, auch mehrschrittige Textaufgaben im Bereich der natürlichen Zahlen lösen. Auf dieser Kompetenzstufe zeigen Schülerinnen und Schüler ein zunehmendes Verständnis von Brüchen und Dezimalzahlen. Sie können ihr Wissen über zwei- und dreidimensionale geometrische Figuren in vielfältigen Situationen anwenden. Sie können Daten interpretieren und darstellen, um mehrschrittige Aufgaben zu lösen.
Kompetenzstufe IV (550–624): hoch
<i>Die Schülerinnen und Schüler können ihre mathematischen Fertigkeiten und Fähigkeiten für das Lösen von Problemen anwenden.</i> Sie können Textaufgaben lösen, die natürliche Zahlen, einfache Brüche und Dezimalzahlen mit zwei Nachkommastellen umfassen. Schülerinnen und Schüler zeigen Verständnis von geometrischen Eigenschaften von Figuren sowie von Winkeln, die kleiner oder größer als ein rechter Winkel sind. Sie können Daten aus Tabellen und verschiedenen Diagrammen interpretieren und nutzen, um Probleme zu lösen.
Kompetenzstufe III (475–549): durchschnittlich
<i>Die Schülerinnen und Schüler können elementares mathematisches Wissen sowie elementare mathematische Fertigkeiten und Fähigkeiten in einfachen Situationen anwenden.</i> Sie zeigen Grundwissen über natürliche Zahlen und anfängliches Wissen über Brüche und Dezimalzahlen. Die Schülerinnen und Schüler können einfache zwei- und dreidimensionale Figuren aufeinander beziehen sowie einfache Figuren erkennen und zeichnen. Die Schülerinnen und Schüler können Säulendiagramme und Tabellen lesen und interpretieren.
Kompetenzstufe II (400–474): niedrig
<i>Die Schülerinnen und Schüler verfügen über elementares mathematisches Wissen sowie elementare mathematische Fertigkeiten und Fähigkeiten.</i> Sie können einfache Additions- und Subtraktionsaufgaben mit natürlichen Zahlen lösen, verfügen über anfängliches Verständnis der Multiplikation mit einstelligen Zahlen und können einfache Textaufgaben lösen. Sie haben Kenntnisse über einfache Brüche, geometrische Grundformen und Maßeinheiten. Die Schülerinnen und Schüler können einfache Balkendiagramme und Tabellen lesen und vervollständigen.
Kompetenzstufe I (unter 400): rudimentär
<i>Die Schülerinnen und Schüler verfügen über rudimentäres schulisches Anfangswissen.</i> Auf diesem Niveau können viele Kinder einfache Routineaufgaben mit Grundrechenarten ausführen, jedoch selbst einfache Aufgaben des Niveaus II nur gelegentlich oder ansatzweise lösen.

Die Abbildungen 3.3 und 3.4 enthalten Beispielaufgaben zur Illustration der Kompetenzstufen.

Abbildung 3.3: Kompetenzstufen und Beispielaufgaben (Gesamtskala Mathematik) I



Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an.
Abdruck und Nutzung der Aufgaben nur mit ausdrücklicher Genehmigung der IEA: <https://www.iea.nl/publications/form/iea-permission-request-form>

Abbildung 3.4: Kompetenzstufen und Beispielaufgaben (Gesamtskala Mathematik) II

In den beiden Abbildungen 3.3 und 3.4 werden für jede Aufgabe zwei in Klammern gesetzte Werte angegeben, die die internationalen und die nationalen Lösungshäufigkeiten zeigen. Beispielsweise wurde die erste Aufgabe von 14 Prozent der Schülerinnen und Schüler im internationalen Durchschnitt und von 9 Prozent der Lernenden in Deutschland gelöst. Die Zuordnung einer Aufgabe zu einer der fünf Kompetenzstufen erfolgt anhand der internationalen Leistungsdaten. Es kann also auch der Fall eintreten, dass in einzelnen Teilnehmerstaaten und Regionen eine Aufgabe, die für eine höhere Kompetenzstufe charakteristisch ist, eine höhere Lösungshäufigkeit aufweist als eine Aufgabe, die auf Grundlage der internationalen Lösungshäufigkeiten eigentlich einer niedrigeren Kompetenzstufe zugeordnet wurde. Im Folgenden wird für die einzelnen Aufgaben beschrieben, über welche mathematischen Kompetenzen Schülerinnen und Schüler verfügen müssen, um die jeweilige Aufgabe erfolgreich lösen zu können.

Kompetenzstufe V: Bei Aufgabe 1 wird den Schülerinnen und Schülern die Aufgabe gestellt, die Kosten zweier Obstsorten auf der Grundlage gegebener Informationen zu bestimmen. Die Aufgabe erfordert mehrere Lösungsschritte. Die gesuchten Größen können dabei nicht durch basale Routinetätigkeiten errechnet, sondern müssen geschickt abgeleitet werden.

Kompetenzstufe IV: Bei Aufgabe 2 wird den Kindern ein Quadrat inklusive der Seitenlängen dargelegt. Daraufhin sollen sie anhand dreier Figuren entscheiden, wie viele der jeweiligen Figuren nötig sind, um das gegebene Quadrat vollständig auszulegen. Um diese Aufgabe passend bearbeiten zu können, müssen die Kinder zum mentalen Operieren mit ebenen Figuren fähig sein und geometrische Eigenschaften der Figuren nutzen.

Bei Aufgabe 3 sollen die Schülerinnen und Schüler die Anzahl quadratischer und dreieckiger Plättchen bestimmen, die für die Erstellung vorgegebener dreidimensionaler Figuren nötig sind. Hierzu müssen die Schülerinnen und Schüler die geometrischen Eigenschaften der Figuren nutzen und über Kompetenzen zum räumlichen Vorstellungsvermögen verfügen.

Kompetenzstufe III: Bei Aufgabe 4 müssen die Kinder entscheiden, welcher Term zu einer gegebenen einfachen Sachsituation passt. Hierzu ist insbesondere die Fähigkeit zur Übersetzung einer schriftlich-symbolischen und ikonischen Darstellung in eine mathematisch-symbolische Darstellung notwendig.

Bei Aufgabe 5 sollen die Schülerinnen und Schüler ein fehlendes Rechenzeichen in einer Gleichung so bestimmen, dass eine Parität der beiden zu vergleichenden Terme vorliegt. Dazu ist ein Operationsverständnis über die vier Grundrechenarten erforderlich.

Kompetenzstufe II: Bei Aufgabe 6 sollen die Schülerinnen und Schüler eine achsensymmetrische Ergänzung vornehmen. Hierfür müssen die Kinder über ein grundlegendes Verständnis zur Achsensymmetrie verfügen, das auf eine einfache geometrische Form angewendet wird.

Bei Aufgabe 7 sollen die Kinder das Ergebnis einer Divisionsaufgabe mit dreistelligem Dividenten und einstelligem Divisor ermitteln. Hierzu ist es erforderlich, eine passende Strategie zur Berechnung dieser Aufgabe gedanklich abzurufen und durchführen zu können. Eine Möglichkeit ist die Anwendung der

multiplikativen Umkehraufgabe und ein Überschlagen der Ergebnisse für die vier Antwortalternativen.

Kompetenzstufe I (ohne Beispiel): Die Schülerinnen und Schüler, die der Kompetenzstufe I zuzuordnen sind, lassen im TIMSS-Test allenfalls basale Kompetenzen erkennen. Dementsprechend ist davon auszugehen, dass von diesen Lernenden lediglich einfache Routineaufgaben erfolgreich bearbeitet werden können.

3.5 Ergebnisse

In diesem Abschnitt erfolgt die Darstellung zentraler Resultate von TIMSS 2019 entlang der in Abschnitt 3.1 formulierten Forschungsfragen. Diese werden zudem mit den Ergebnissen der drei vorangegangenen Studienzyklen verglichen, an denen Deutschland teilgenommen hat. Hierzu werden in Abschnitt 3.5.1 die Ergebnisse der Viertklässlerinnen und Viertklässler im internationalen Vergleich beschrieben. Abschnitt 3.5.2 ist Ausführungen über die Verteilung der Schülerinnen und Schüler der teilnehmenden Staaten und Regionen auf die fünf TIMSS-Kompetenzstufen gewidmet. Hierbei wird auch dargestellt, welche Anteile besonders leistungsstarker beziehungsweise besonders leistungsschwacher Schülerinnen und Schüler ermittelt wurden. In den Abschnitten 3.5.3 und 3.5.4 werden die Testergebnisse im Hinblick auf die drei Inhaltsbereiche beziehungsweise die drei kognitiven Anforderungsbereiche dargestellt. Schließlich werden in Abschnitt 3.5.5 die nationalen Ergebnisse zum mathematischen Selbstkonzept von Viertklässlerinnen und Viertklässlern berichtet. Weitere Ergebnisse zu Zusammenhängen von mathematischer Kompetenz und den Variablen *Geschlecht*, *soziale Disparitäten* sowie *Migrationshintergrund* werden in den Kapiteln 8, 9 und 10 in diesem Band behandelt.

Insgesamt 58 Staaten sowie sechs sogenannte *Benchmark*-Teilnehmer, also einzelne Regionen von Teilnehmerstaaten, wie Québec (Kanada) oder Abu Dhabi (VAE), nahmen an TIMSS 2019 teil. In diesem Abschnitt werden somit die Ergebnisse aus 64 Staaten und Regionen berichtet, darunter 24 Staaten aus der EU beziehungsweise 29 OECD-Staaten (siehe Kapitel 2 in diesem Band). Im Vergleich zu TIMSS 2015 (53 Teilnehmende), TIMSS 2011 (59 Teilnehmende) und TIMSS 2007 (43 Teilnehmende) liegt eine höhere Beteiligung vor.

Zur Einordnung der Testergebnisse der Schülerinnen und Schüler aus Deutschland werden im weiteren Verlauf dieses Abschnitts Bezüge zu sogenannten *Vergleichsgruppen* (VG) hergestellt. Die Ergebnisse für Deutschland werden verglichen mit den Mittelwerten der

- Teilnehmerstaaten, die Mitglieder der EU sind (VG_{EU}),
- Teilnehmerstaaten, die der OECD angehören (VG_{OECD}),
- Teilnehmerstaaten, die auf der Gesamtskala Mathematik signifikant bessere Ergebnisse erzielt haben,
- Teilnehmerstaaten, deren Ergebnisse sich auf der Gesamtskala Mathematik nicht statistisch signifikant von Deutschland unterscheiden.

Eine Gesamtschau aller Resultate von TIMSS 2019 ist im Ergebnisbericht der internationalen Studienleitung einzusehen (Mullis et al., 2020).

Im vorliegenden Bericht werden die Ergebnisse des Jahres 2019 nicht nur berichtet, sondern auch zu denen der Jahre 2007, 2011 und 2015 in Beziehung

gesetzt, also der Studienzyklen, an denen Deutschland bisher beteiligt war. Abweichend zu den Abbildungen, in denen die Testergebnisse der teilnehmenden Staaten und Regionen aus dem Jahr 2019 dargestellt sind, wird bei allen Abbildungen, die von Unterschieden zwischen den verschiedenen TIMSS-Zyklen berichten, auf die Angabe des internationalen Mittelwertes sowie des Mittelwertes der VG_{EU} und der VG_{OECD} verzichtet. Diese Entscheidung wurde getroffen, weil sich die Zusammensetzung der Teilnehmenden der verschiedenen Vergleichsgruppen zwischen den TIMSS-Zyklen – wenn auch nur geringfügig – unterscheiden, wodurch Interpretationen zu Trends nur schwer zu formulieren wären.

3.5.1 Kompetenzen im internationalen Vergleich

In diesem Abschnitt werden die Testergebnisse der in Deutschland teilnehmenden Viertklässlerinnen und Viertklässler im internationalen Vergleich dargestellt. Dabei werden die Mittelwerte, die Streuung der Ergebnisse sowie die Positionierung des Ergebnisses im internationalen Vergleich berichtet. Die gestufte internationale Rangreihe und die Einordnung der Ergebnisse für Deutschland wird zunächst numerisch dargestellt, ohne auf inhaltliche Aspekte der Ergebnisse einzugehen, also den Bezug zu den mathematischen Inhaltsbereichen oder zu den kognitiven Anforderungsbereichen herzustellen (siehe Abschnitt 3.5.3 und 3.5.4).

Leistungsmittelwerte der Staaten: Für die an TIMSS 2019 teilnehmenden Staaten und Regionen werden in Abbildung 3.5 die Leistungsmittelwerte (M) mit Standardabweichungen (SD) und die zugehörigen Standardfehler (SE) angegeben. Außerdem werden die entsprechenden Werte für alle Teilnehmerstaaten (internationaler Mittelwert) sowie die Mittelwerte der VG_{OECD} und VG_{EU} vermerkt. Zudem erfolgt eine gesonderte Darstellung der *Benchmark*-Teilnehmer.

Zur Vergleichbarkeit der Testergebnisse sind strenge methodische Standards vorgegeben. Wenn diese Standards nicht eingehalten werden, sind die für den jeweiligen Teilnehmenden erzielten Ergebnisse nur bedingt vergleichbar. In TIMSS 2019 ist dieser Umstand für 22 der 58 Teilnehmerstaaten gegeben. In den folgenden Abbildungen sind diese Staaten jeweils kursiv gesetzt und vorhandene Fußnoten weisen auf spezielle Merkmale der jeweiligen Untersuchungspopulationen und Stichproben hin (siehe Anhang). Die übrigen 36 Teilnehmerstaaten, darunter auch Deutschland, erfüllen die international vorgegebenen Kriterien, sodass die internationale Vergleichbarkeit dieser Testergebnisse umfänglich gewährleistet ist.

Die Viertklässlerinnen und Viertklässler in Deutschland erzielen einen mittleren Leistungswert von 521 Punkten, der sich im Mittelfeld der internationalen Rangreihe aller Teilnehmerstaaten befindet. Damit liegen die Testergebnisse signifikant über dem Skalenmittelwert von 500 sowie dem internationalen Mittelwert aller teilnehmenden Staaten und Regionen, der bei 501 liegt. Die Teilnehmerstaaten mit den besten Testergebnissen sind Singapur (625), Hongkong (602) sowie die Republik Korea (600). Die schwächsten Testergebnisse sind für die Teilnehmerstaaten Philippinen (297), Pakistan (328) und Südafrika (374) zu verzeichnen.

Signifikante Leistungsunterschiede: Eine an Rangplätzen orientierte Darstellung liefert oft unzureichende Aufschlüsse über die zu beobachtenden Leistungs-

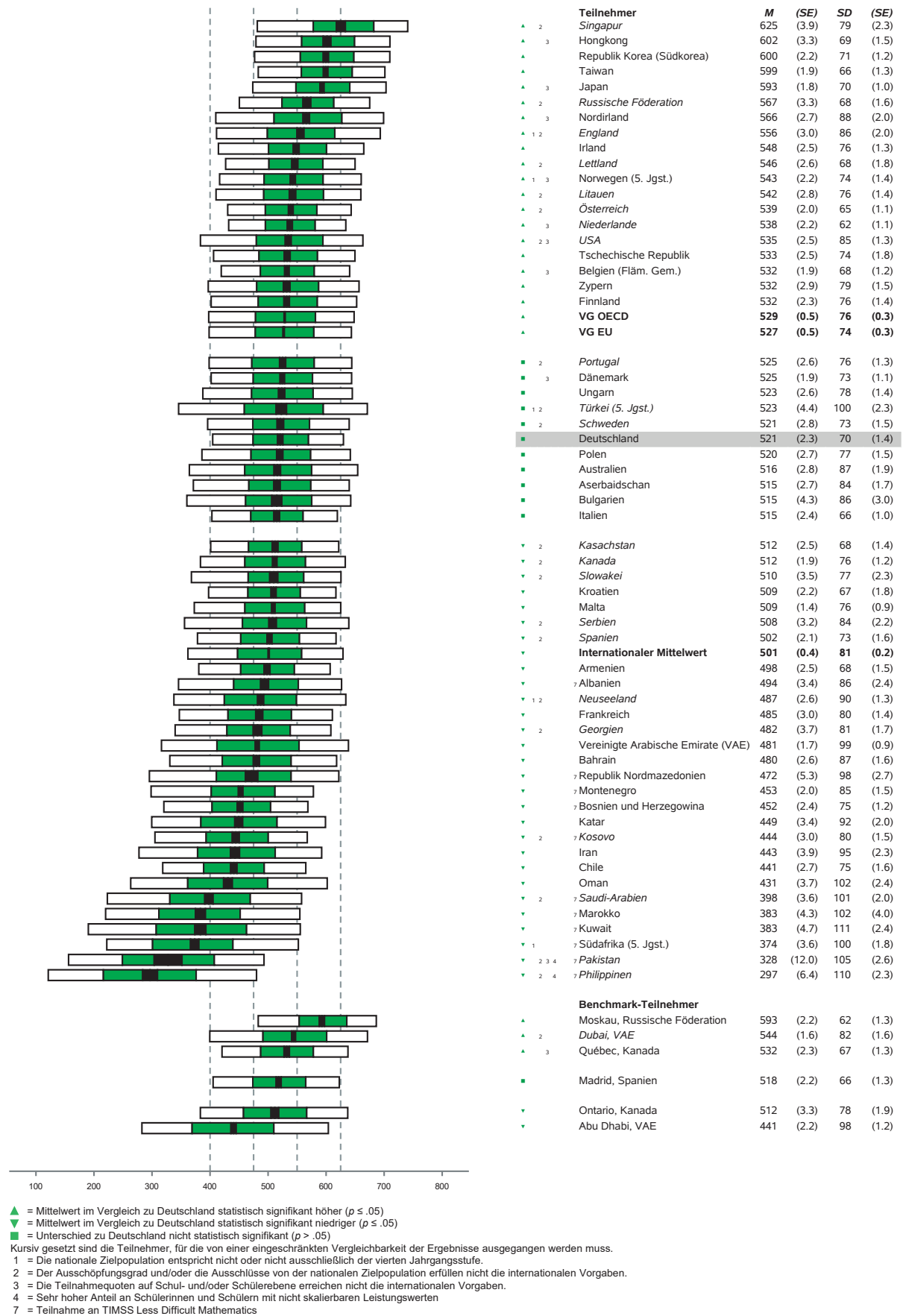
unterschiede. Die numerisch angegebenen Testergebnisse können in Einzelfällen – insbesondere bei nur geringen Abweichungen – Unterschiede in Bezug auf die mathematischen Kompetenzen suggerieren, obwohl sie keine statistische Bedeutsamkeit haben.

Dementsprechend sind in Abbildung 3.5 die Teilnehmerstaaten mit einem grünen Quadrat versehen, deren Mittelwerte sich zwar geringfügig, jedoch – auf Signifikanzniveau von 5 Prozent – nicht signifikant vom Leistungsmittelwert für Deutschland (521) unterscheiden. Diese Staatengruppe, die im Folgenden auch als ‚Leistungsgruppe Deutschland‘ beschrieben wird, umfasst Portugal (525), Dänemark (525), Ungarn (523), die Türkei (5. Jahrgangsstufe, 523), Schweden (521), Polen (520), Australien (516), Aserbaidschan (515), Bulgarien (515) und Italien (515). Signifikant bessere Leistungsmittelwerte als in Deutschland wurden von Schülerinnen und Schülern aus 19 Teilnehmerstaaten erzielt. Neben den bereits genannten Spitzenreitern liegen auch für Taiwan (599), Japan (593), die Russische Föderation (567), Nordirland (566), England (556), Irland (548), Lettland (546), Norwegen (5. Jahrgangsstufe, 543), Litauen (542), Österreich (539), die Niederlande (538), die USA (535), die Tschechische Republik (533), die Flämische Gemeinschaft in Belgien (532), Zypern (532) und Finnland (532) signifikant bessere Testergebnisse als für Deutschland vor. Demgegenüber wurden signifikant schlechtere Leistungsmittelwerte bei 28 Teilnehmerstaaten festgestellt. Zu diesen Staaten zählen unter anderem Kasachstan (512), Kanada (512), die Slowakei (510), Kroatien (509), Serbien (508), Spanien (502), Albanien (494), Neuseeland (487), Frankreich (485), Bosnien und Herzegowina (452), Katar (449), Kosovo (444), der Iran (443), Chile (441), Saudi-Arabien (398), Marokko (383) und Südafrika (5. Jahrgangsstufe, 374). Ferner liegt der von Grundschulkindern aus Deutschland erzielte Leistungsmittelwert signifikant unter dem Mittelwert der teilnehmenden EU-Staaten (VG_{EU} 527) und OECD-Staaten (VG_{OECD} 529).

Leistungsstreuung: Bislang wurden die Leistungsmittelwerte der teilnehmenden Staaten und Regionen sowie die Unterschiede zwischen diesen Werten beschrieben. Von Interesse ist jedoch auch, inwieweit sich die mathematischen Leistungswerte jeweils verteilen. Hierzu wird die Leistungsstreuung der Teilnehmerstaaten in Abbildung 3.5 durch Perzentilbänder dargestellt. Dabei werden jeweils das 5., das 25., das 75. und das 95. Perzentil angegeben (siehe Kapitel 2 in diesem Band). So sagt beispielsweise der für das 5. Perzentil angegebene Leistungswert aus, dass 5 Prozent der teilnehmenden Schülerinnen und Schüler einen gleich großen oder kleineren und dass 95 Prozent einen größeren Leistungswert erreicht haben.

Die in Abbildung 3.5 dargestellten Perzentilbänder zeigen für die Testergebnisse Deutschlands eine im Vergleich zu vielen anderen Teilnehmerstaaten eher geringe Leistungsstreuung. Im internationalen Vergleich liegt also eine eher homogene Verteilung der Mathematikleistung vor. Dies ist an der Breite der Perzentilbänder zu erkennen, in denen sich der Punkteunterschied zwischen den Perzentilen 95 und 5 ausdrückt.

Die Leistungsstreuung wird auch in der Standardabweichung (SD) der teilnehmenden Staaten und Regionen vom jeweiligen Leistungsmittelwert deutlich. In Abbildung 3.5 ist für Deutschland ein Wert von 70 Punkten angegeben. Dieser Wert besagt, dass etwa zwei Drittel der getesteten Schülerinnen und Schüler Leistungen erzielten, die zwischen 451 ($521 - 70$) und 591 ($521 + 70$) liegen.

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


Sowohl im europäischen als auch im internationalen Vergleich ist Deutschland ein Staat mit einer eher kleinen Standardabweichung.

Konkret bedeutet dies beispielsweise für einen Vergleich der Testergebnisse Deutschlands mit denen der USA, dass die USA mit 535 zwar einen signifikant höheren Mittelwert als Deutschland (521) erzielt hat, sich die Leistungswerte der Viertklässlerinnen und Viertklässler in den USA jedoch weitaus stärker voneinander unterscheiden und damit eine stärkere Heterogenität mathematischer Leistung vorliegt. Die schwächsten Lernenden in den USA erzielen niedrigere Punktwerte als die schwächsten Lernenden in Deutschland. Ein höherer Mittelwert kann also auch mit einer Benachteiligung leistungsschwacher Kinder einhergehen.

Ein Staat (Japan) weist mit 70 eine gleich große, 45 Teilnehmerstaaten weisen eine größere und elf eine kleinere Standardabweichung auf. Das ist insofern eine interessante Beobachtung, als dass sich in den vergangenen TIMSS-Durchgängen die Anzahl der Staaten mit einer geringeren Leistungsstreuung als Deutschland immer zwischen eins und drei bewegte. Das liegt zum einen daran, dass in Deutschland die Standardabweichung einen höheren Wert als in den vergangenen Durchgängen aufweist (2015: 65; 2011: 62; allerdings 2007: auch schon 68), und zum anderen daran, dass dieser Wert bei einigen anderen Teilnehmerstaaten – wie etwa der Russischen Föderation – nennenswert gesunken ist (2007: 83; 2011: 74; 2015: 73; 2019: 68).

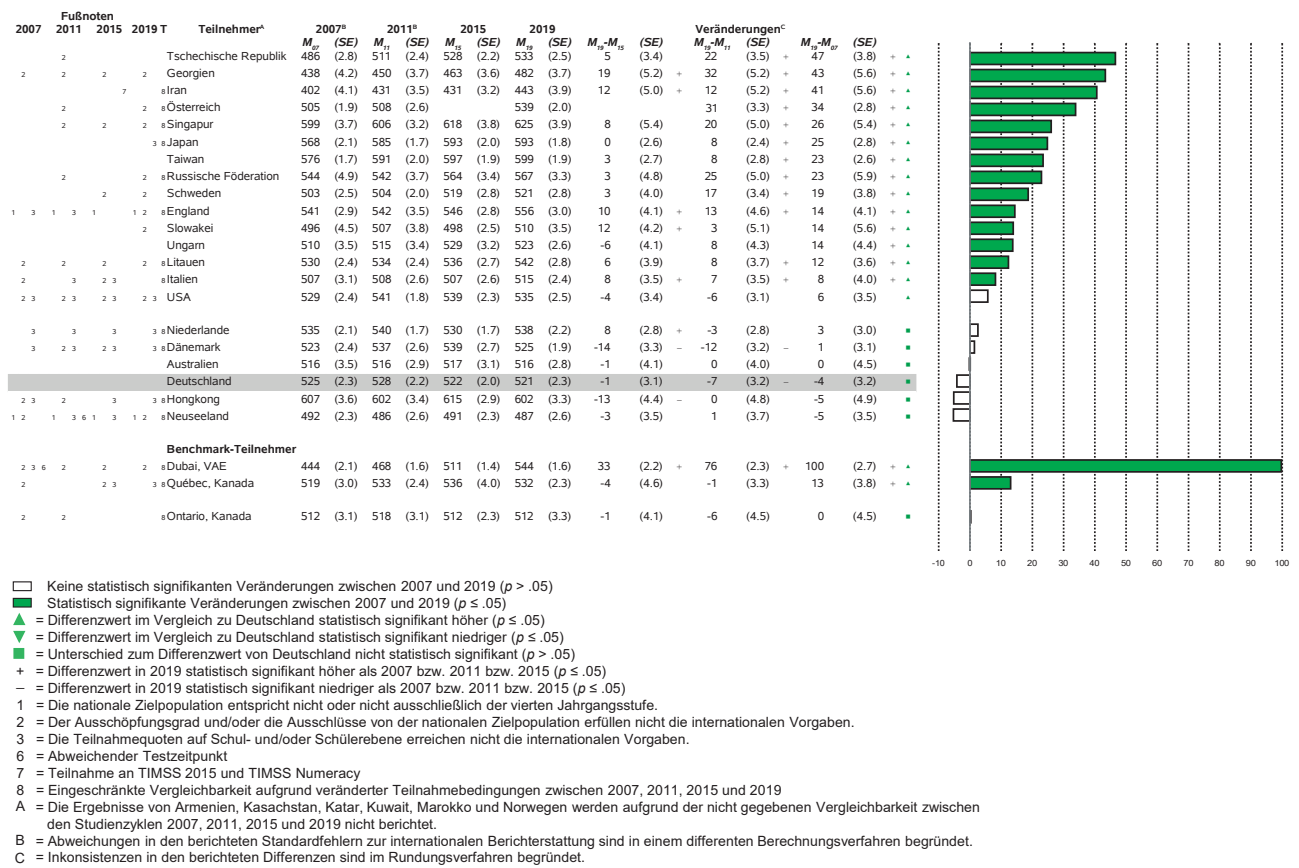
Vergleich der Testleistungen 2019 mit 2015, 2011 und 2007: Im Folgenden sollen die Ergebnisse der TIMS-Studie 2019 zu den Resultaten der drei vorangegangenen Zyklen in Beziehung gesetzt werden. Hierzu werden in Abbildung 3.6 die nationalen Mittelwerte und Standardfehler derjenigen Teilnehmerstaaten verglichen, die ebenfalls an den vier genannten Untersuchungen teilgenommen haben. Zudem wird Österreich angeführt, obwohl es 2015 nicht teilgenommen hat (siehe Kapitel 2 in diesem Band). Darüber hinaus werden auch die Differenzen der Testergebnisse aus 2019 zu denen der Zyklen 2007, 2011 und 2015 angegeben.

Auf der rechten Seite der Abbildung ist zu erkennen, dass 14 Teilnehmerstaaten signifikant höhere Leistungswerte für TIMSS 2019 im Vergleich zu TIMSS 2007 aufweisen. Die entsprechenden Differenzbalken sind in der Abbildung grün eingefärbt. Bei sieben Staaten, zu denen auch Deutschland zählt, sind hingegen keine signifikanten Unterschiede festzustellen. Diese Differenzbalken sind in weißer Farbe dargestellt.

Der Vergleich der Zyklen von 2019 und 2011 zeigt signifikant höhere Leistungsmittelwerte in 2019 bei zwölf Teilnehmerstaaten, wohingegen Dänemark und auch Deutschland signifikant schwächer als 2011 abschnitten. Bei den übrigen sieben Staaten unterscheiden sich die Werte nicht signifikant.

Die festgestellten Leistungsmittelwerte der jüngsten Zyklen aus 2019 und 2015 unterscheiden sich weniger häufig signifikant. Für sechs Teilnehmerstaaten liegen bedeutend höhere und für zwei bedeutend geringere Leistungsmittelwerte in 2019 vor. Für die übrigen zwölf Staaten, darunter auch Deutschland, sind keine signifikanten Unterschiede festzustellen.

Extremgruppen: Den Tabellen 3.14.1 und 3.14.2 können die nach Perzentilen aufgeführten Mathematikleistungen entnommen werden. Um auch Veränderungen der Leistungswerte zwischen den einzelnen Studienzyklen beschreiben zu können, werden in die Darstellung diejenigen Teilnehmerstaaten aufgenommen, die an allen vier Untersuchungen teilgenommen haben. Die Daten sind entsprechend

Abbildung 3.6: Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 – Gesamtskala Mathematik

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

der in Abbildung 3.5 beschriebenen Mittelwerte in der Gesamtskala Mathematik geordnet.

Tabelle 3.14.2 zeigt, dass in 2019 die 5 Prozent leistungsschwächsten Kinder in Deutschland Leistungsmittelwerte von höchstens 405 Punkten erzielten. Die in der Tabelle für TIMSS 2019 durch ein grünes Quadrat markierten Werte der Teilnehmer weisen jeweils darauf hin, dass keine statistisch signifikanten Unterschiede zu den Punktwerten für Deutschland vorliegen, ein nach oben beziehungsweise nach unten weisendes Dreieck weist auf signifikante Unterschiede hin. Beispielsweise unterscheiden sich die zu Perzentil 5 gehörenden Werte von England (411), Litauen (410), der Tschechischen Republik (406), Dänemark (402), Schweden (396) und Italien (403) nicht signifikant von Deutschland. Hingegen konnten die leistungsschwächsten Viertklässlerinnen und Viertklässler aus insgesamt sieben Staaten – Singapur (481), Hongkong (479), Taiwan (483), Japan (474), der Russischen Föderation (451), Österreich (430) und den Niederlanden (432) – signifikant höhere Leistungswerte in TIMSS 2019 erreichen. Andererseits sind in sieben Staaten, nämlich den USA (383), Ungarn (388), Australien (364), der Slowakei (368), Neuseeland (338), Georgien (340) und dem Iran (278) im Vergleich zu Deutschland signifikant niedrigere Leistungswerte in Perzentil 5 erkennbar.

Die leistungstärksten 5 Prozent der Schülerinnen und Schüler in Deutschland erreichten Leistungswerte von mindestens 630 Punkten. Während sich in dieser Hinsicht die Leistungswerte aus Deutschland von denjenigen aus den Niederlanden (634), Schweden (640), Italien (619), der Slowakei (626) und

Tabelle 3.14.1: Mathematikleistungen nach Perzentilen im Vergleich: TIMSS 2007, 2011, 2015 und 2019 – Gesamtskala Mathematik

Fußnoten														
2007	2011	2015	2019 T	Teilnehmer ^a	TIMSS 2007 Perzentile					TIMSS 2011 Perzentile				
					5 (SE)	25 (SE)	75 (SE)	95 (SE)		5 (SE)	25 (SE)	75 (SE)	95 (SE)	
	2	2	2	8 Singapur	447 (6.5) +	548 (5.1) +	659 (4.0) +	725 (4.1) +	464 (7.4)	559 (5.3) +	661 (4.2) +	723 (3.2) +		
2 3	2	3	3	8 Hongkong	493 (9.1)	564 (4.2)	653 (4.0)	712 (5.3)	488 (11.0)	563 (3.3)	645 (3.4)	702 (2.7)		
				Taiwan	457 (4.1) +	532 (2.1) +	623 (2.2) +	686 (2.1) +	459 (6.4) +	546 (3.6) +	642 (2.3)	704 (2.5)		
				3 8 Japan	438 (2.6) +	520 (2.1) +	620 (2.1) +	688 (3.8) +	460 (6.9)	540 (2.1) +	635 (2.4)	700 (4.7)		
	2		2	8 Russische Föderation	400 (4.1) +	492 (5.4) +	599 (5.1) +	677 (9.8)	417 (6.7) +	493 (4.0) +	593 (4.5) +	660 (6.5)		
1 3	1 3	1	1 2	8 England	392 (4.2) +	487 (3.5) +	600 (3.6) +	676 (4.3) +	385 (5.9) +	483 (6.4) +	605 (2.5) +	677 (5.7)		
2	2	2	2	8 Litauen	396 (3.7) +	482 (3.5) +	583 (2.9) +	645 (5.6) +	405 (4.5)	486 (2.9)	585 (2.7) +	650 (2.1)		
	2		2	8 Österreich	386 (3.1) +	462 (3.6) +	552 (3.0) +	612 (1.5) +	401 (4.3) +	466 (3.7) +	552 (2.9) +	606 (4.4) +		
3	3	3	3	8 Niederlande	429 (7.0)	495 (2.9)	577 (2.8)	632 (2.3)	449 (3.2) –	505 (2.6) –	577 (1.9)	623 (2.7) +		
2 3	2 3	2 3	2 3	USA	401 (3.8) –	479 (2.3)	581 (3.0) +	650 (5.2) +	410 (3.3) –	492 (2.2) –	593 (2.3)	660 (1.5)		
	2			Tschechische Republik	361 (6.6) +	440 (4.9) +	536 (2.9) +	597 (2.9) +	387 (6.0) +	467 (2.8) +	560 (2.3) +	621 (3.7) +		
3	2 3	2 3	3	8 Dänemark	403 (9.9)	478 (3.9)	571 (2.4)	634 (4.8)	413 (5.3)	493 (2.8) –	585 (3.0) –	646 (4.5)		
				Ungarn	347 (12.4) +	452 (6.6) +	574 (3.7)	647 (4.6)	352 (8.3) +	462 (3.8)	577 (2.8)	650 (2.8)		
	2	2	2	Schweden	388 (4.5)	459 (3.3) +	548 (3.0) +	608 (2.7) +	388 (6.9)	462 (2.8) +	549 (2.4) +	610 (4.8) +		
				Deutschland	409 (10.3)	483 (2.6) –	572 (2.2)	629 (2.6)	420 (7.8)	488 (2.9) –	570 (2.3)	626 (2.1)		
				Australien	373 (8.2)	463 (4.1)	573 (4.2)	647 (3.9)	366 (6.2)	462 (3.1)	574 (4.0)	652 (6.0)		
2	3	2 3		8 Italien	374 (6.1) +	457 (3.7) +	558 (3.3)	629 (12.2)	386 (6.2) +	461 (4.6)	557 (3.3)	622 (3.6)		
			2	Slowakei	350 (9.8)	446 (4.2) +	553 (3.8)	623 (5.2)	365 (8.4)	460 (3.6)	561 (2.2)	626 (5.2)		
1 2	1 3 6 1	3	1 2	8 Neuseeland	341 (7.3)	436 (2.6) –	553 (2.5)	626 (3.6)	339 (6.7)	432 (2.3)	545 (2.7)	614 (3.5) +		
2	2	2	2	Georgien	289 (6.3) +	378 (6.8) +	501 (5.8) +	582 (6.4) +	293 (8.4) +	392 (6.0) +	512 (2.7) +	589 (6.0) +		
			7	8 Iran	260 (5.6)	346 (4.9) +	461 (4.2) +	534 (6.4) +	271 (6.4)	370 (3.3)	496 (3.9) +	575 (2.9) +		
				Benchmark-Teilnehmer										
2 3 6	2	2	2	8 Dubai, VAE	406 (2.8)	473 (4.1) +	566 (4.5) +	627 (4.5) +	432 (4.1) –	492 (2.9)	574 (2.8) +	631 (4.5) +		
2		2 3	3	8 Québec, Kanada	293 (9.5) +	384 (3.6) +	506 (2.3) +	589 (6.2) +	292 (3.5) +	400 (3.8) +	541 (2.6) +	623 (4.4) +		
2	2			8 Ontario, Kanada	395 (4.5)	468 (6.2)	558 (5.2)	621 (3.6) +	393 (4.3)	470 (4.1) –	568 (5.8)	634 (6.1)		

+ = Mittelwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)– = Mittelwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

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

6 = Abweichender Testzeitpunkt

7 = Teilnahme an TIMSS 2015 und TIMSS Numeracy

8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

Neuseeland (634) nicht signifikant unterscheiden, erzielten die leistungsstärksten 5 Prozent der Schülerinnen und Schüler aus 13 Staaten, darunter neben den Spitzenreitern der Gesamtskala aus den asiatischen Staaten auch England (693), die USA (663), Litauen (660), die Tschechische Republik (650) oder Ungarn (645), signifikant höhere Leistungsmittelwerte. Lediglich in zwei Staaten, Georgien (608) und dem Iran (592), liegen signifikant niedrigere Punktwerte vor. Diese Verteilungen deuten darauf hin, dass die Förderung leistungsstarker Grundschülerinnen und Grundschüler in Deutschland im internationalen Vergleich noch weniger gut zu gelingen scheint als die Förderung von schwächeren Lernenden.

In den Tabellen 3.14.1 und 3.14.2 sind neben den bereits beschriebenen Mathematikleistungen des 5. und 95. Perzentils in TIMSS 2019 auch die des 25. und 75. Perzentils angegeben. Darüber hinaus sind auch die entsprechenden Werte der Zyklen 2007, 2011 und 2015 vermerkt. So können Veränderungen der Mathematikleistungen im Kontext der verschiedenen Perzentile nachvollzogen werden. Statistisch signifikante Veränderungen werden dabei durch ein Pluszeichen für signifikant höhere und durch ein Minuszeichen für signifikant niedrigere Leistungsmittelwerte markiert.

Tabelle 3.14.2: Mathematikleistungen nach Perzentilen im Vergleich: 2007, 2011, 2015 und 2019 – Gesamtskala Mathematik

Fußnoten																						
2007	2011	2015	2019	T	Teilnehmer ^A	TIMSS 2015 Perzentile								TIMSS 2019 Perzentile								
						5	(SE)	25	(SE)	75	(SE)	95	(SE)	5	(SE)	25	(SE)	75	(SE)	95	(SE)	
	2	2	2	2	8 Singapur	458	(9.0) +	566	(5.2)	677	(3.8)	746	(4.6)	481	(8.0) ▲	578	(4.9) ▲	682	(3.8) ▲	741	(3.8) ▲	
2	3	2	3	3	8 Hongkong	505	(5.5) −	573	(3.8) −	659	(3.1) −	721	(5.4)	479	(5.2) ▲	558	(4.1) ▲	649	(3.7) ▲	710	(4.3) ▲	
					Taiwan	474	(4.0)	552	(2.9)	645	(2.8)	709	(4.0)	483	(5.6) ▲	557	(2.7) ▲	645	(3.2) ▲	701	(2.6) ▲	
					3 8 Japan	476	(3.9)	549	(2.0)	639	(2.6)	703	(4.3)	474	(3.7) ▲	548	(2.5) ▲	641	(2.3) ▲	703	(3.1) ▲	
	2			2	8 Russische Föderation	441	(6.0)	517	(3.7)	613	(4.2)	683	(7.7)	451	(7.7) ▲	524	(4.1) ▲	613	(4.0) ▲	675	(5.2) ▲	
1	3	1	3	1	2 8 England	407	(5.9)	490	(3.9)	602	(3.7) +	682	(7.8)	411	(5.5) ■	499	(3.5) ▲	615	(4.1) ▲	693	(6.5) ▲	
	2	2	2	2	8 Litauen	411	(5.9)	489	(3.8)	584	(2.9) +	646	(4.7) +	410	(5.7) ■	493	(3.4) ▲	596	(2.8) ▲	660	(4.9) ▲	
	2			2	8 Österreich									430	(4.8) ▲	495	(2.9) ▲	584	(2.7) ▲	643	(3.5) ▲	
3	3		3	3	3 8 Niederlande	436	(3.7)	492	(2.4)	569	(2.0) +	619	(3.6) +	432	(4.3) ▲	495	(2.7) ▲	581	(2.5) ▲	634	(3.7) ■	
2	3	2	3	2	3 8 USA	398	(4.5)	485	(2.4)	596	(2.3)	667	(3.5)	383	(5.9) ▼	480	(3.2) ■	594	(3.0) ▲	663	(4.0) ▲	
	2				Tschechische Republik	409	(5.6)	484	(3.3)	576	(2.4) +	640	(4.3)	406	(5.8) ■	484	(3.5) ▲	585	(2.8) ▲	650	(3.9) ▲	
3	2	3	2	3	3 8 Dänemark	408	(5.6)	490	(3.7) −	591	(3.4) −	656	(3.1) −	402	(5.5) ■	474	(2.7) ■	577	(2.4) ■	644	(4.2) ▲	
					Ungarn	372	(7.5)	475	(5.4)	591	(2.9) −	660	(4.1) −	388	(5.2) ▼	472	(4.1) ■	578	(2.8) ▲	645	(3.3) ▲	
			2	2	Schweden	398	(7.0)	475	(3.9)	567	(3.2)	626	(3.6) +	396	(5.4) ■	474	(4.0) ■	571	(3.4) ■	640	(4.4) ■	
					Deutschland	410	(5.4)	479	(2.7)	566	(2.2)	626	(3.0)	405	(4.5) ■	474	(2.9) ■	570	(2.8) ■	630	(3.2) ■	
					Australien	374	(6.3)	462	(4.0)	575	(3.5)	649	(3.8)	364	(6.0) ▼	460	(3.6) ▼	575	(3.4) ■	654	(5.8) ▲	
2		3	2	3	8 Italien	383	(5.5) +	461	(4.3)	556	(2.9)	619	(4.8)	403	(4.8) ■	470	(3.2) ■	560	(2.8) ▼	619	(4.4) ■	
				2	Slowakei	354	(6.9)	450	(3.7) +	552	(2.3) +	618	(2.8)	368	(10.9) ▼	466	(4.6) ■	561	(3.2) ■	626	(5.0) ■	
1	2	1	3	6	1	3 8 Neuseeland	335	(5.2)	432	(4.1)	553	(2.0)	632	(4.7)	338	(4.7) ▼	425	(3.4) ▼	549	(3.3) ▼	634	(4.2) ■
	2	2	2	2	Georgien	310	(7.3) +	408	(6.2) +	524	(4.1) +	597	(6.5)	340	(6.2) ▼	429	(5.3) ▼	538	(3.9) ▼	608	(4.4) ▼	
				7	8 Iran	248	(10.0) +	367	(5.1)	504	(3.3)	583	(4.0)	278	(8.6) ▼	379	(5.7) ▼	512	(3.9) ▼	592	(4.7) ▼	
					Benchmark-Teilnehmer																	
2	3	6	2	2	2 8 Dubai, VAE	344	(3.2) +	450	(2.4) +	576	(2.6) +	658	(4.5) +	400	(4.5) ■	491	(2.7) ▲	601	(2.6) ▲	671	(4.2) ▲	
	2		2	3	3 8 Québec, Kanada	428	(7.9)	492	(5.1)	580	(4.2)	644	(7.4)	421	(5.3) ▲	487	(3.8) ▲	578	(3.2) ▲	638	(3.6) ■	
	2	2			8 Ontario, Kanada	388	(4.5)	464	(3.1)	562	(2.9)	630	(4.2)	383	(5.7) ▼	458	(4.5) ▼	567	(4.2) ■	637	(6.1) ■	

▲ = Punktwerte in den Perzentilen im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)▼ = Punktwerte in den Perzentilen im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)■ = Unterschied zu Punktwerten in den Perzentilen im Vergleich zu Deutschland nicht statistisch signifikant ($p > .05$)+ = Mittelwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)– = Mittelwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

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

6 = Abweichender Testzeitpunkt

7 = Teilnahme an TIMSS 2015 und TIMSS Numeracy

8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Für Deutschland zeigt sich im Vergleich der Daten aus 2007 und 2011 mit den Daten aus 2019, dass sich beim Leistungswert (474) für Perzentil 25 in 2019 signifikant niedrigere Werte als in 2007 (483) und 2011 (488) feststellen lassen, weshalb der Perzentilwert bei 2007 und bei 2011 mit einem Minuszeichen versehen ist. Im Vergleich zu 2015 sind keine signifikanten Differenzen zu berichten.

Damit zählt Deutschland zu den drei Teilnehmerstaaten, in denen in einem Perzentil ein signifikant niedrigerer Wert in 2019 als in 2007 vorliegt. Während in Neuseeland ebenfalls in Perzentil 25 ein signifikant niedrigerer Wert zu beobachten ist, ist dieses in den USA bei Perzentil 5 der Fall. Hier sind allerdings auch in den Perzentilen 75 und 95 signifikant bessere Ergebnisse zu konstatieren. Bei vier Teilnehmerstaaten sind keine signifikanten Veränderungen festzustellen, wohingegen die übrigen 15 Teilnehmer in mindestens einem Perzentil signifikant bessere Leistungen erzielen, davon in acht Staaten bei allen vier Werten.

Ähnliche Befunde lassen sich auch für den Vergleich der Testergebnisse von 2011 und 2019 festhalten. Hier liegen für vier Teilnehmerstaaten in keinem Perzentil signifikante Unterschiede vor. Signifikant schwächere Werte in einem Perzentil lassen sich nur für Deutschland erkennen. In zwei Perzentilen liegen

für Dänemark, die USA und die Niederlande signifikant niedrigere Werte vor, wobei letztere im Perzentil 95 signifikant besser abschneiden. Mit insgesamt 15 Teilnehmern hat sich hingegen der überwiegende Anteil der Staaten in mindestens einem Perzentil signifikant verbessert. Für sieben Teilnehmerstaaten liegen dabei signifikante Verbesserungen in mindestens drei der vier betrachteten Perzentile vor.

Der Vergleich der Werte von 2019 mit 2015 schließlich zeigt für Deutschland wie für acht weitere Teilnehmerstaaten keine statistisch bedeutsamen Unterschiede. In drei Staaten ist mindestens ein signifikant niedrigerer (ohne einen signifikant höheren in einem anderen Perzentil) und in zehn Staaten mindestens ein signifikant höherer (ohne einen signifikant niedrigeren in einem anderen Perzentil) Wert festzustellen.

Insgesamt zeigt der Vergleich der Anzahl von Pluszeichen und der Minuszeichen die klare Tendenz, dass sich bei den Werten, die den einzelnen Perzentilen zuzuordnen sind, in vielen Staaten positive Entwicklungen zeigen. Im Vergleich von 2019 mit 2007 beispielsweise stehen 48 Pluszeichen 3 Minuszeichen gegenüber, im Vergleich zu 2011 ist die Relation 34 zu 7 und im Vergleich zu 2015 15 zu 8. Keine signifikant besseren Ergebnisse lassen sich in Deutschland im Vergleich zu den vorangegangenen Studien feststellen, sondern eher Stagnation beziehungsweise im Vergleich zu 2007 und 2011 sogar signifikant schlechtere Resultate.

3.5.2 Kompetenzstufen

Nachdem im Abschnitt 3.4.3 das bei TIMSS eingesetzte Kompetenzstufenmodell und die Charakteristika der Kompetenzstufen I bis V erläutert wurden, soll nun die Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen beschrieben werden. Dabei werden die Ergebnisse im internationalen Vergleich gedeutet und mit den Resultaten der bisherigen Zyklen verglichen.

Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen: Abbildung 3.7 zeigt die Verteilung der teilnehmenden Viertklässlerinnen und Viertklässler auf die fünf Kompetenzstufen. Auf der rechten Seite ist der prozentuale Anteil der Lernenden vermerkt, die die jeweilige Kompetenzstufe erreicht haben. Zudem werden links die kumulierten Anteile notiert, die angeben, wie viel Prozent der teilnehmenden Schülerinnen und Schüler mindestens eine bestimmte Kompetenzstufe erreichen konnten. Für Deutschland zeigt sich in diesem Zusammenhang, dass sich 96 Prozent der Kinder mindestens auf Kompetenzstufe II, 75 Prozent mindestens auf Kompetenzstufe III, 36 Prozent mindestens auf Kompetenzstufe IV und 6 Prozent auf der höchsten Kompetenzstufe V befinden.

Der Vergleich mit den Staaten der ‚Leistungsgruppe Deutschland‘ zeigt, dass anteilig weniger Schülerinnen und Schüler aus Deutschland die höchste Kompetenzstufe V erreichen. Während anteilig mehr Schülerinnen und Schüler aus der Türkei (5. Jahrgangsstufe, 15%), Australien (10%), Portugal, Ungarn (jeweils 9%), Dänemark, Schweden, Polen, Aserbaidschan und Bulgarien (jeweils 8%) im Vergleich zu Deutschland (6%) die höchste Kompetenzstufe erreichen konnten, gilt dies lediglich für Italien (4%) für anteilig weniger Kinder. In den Vergleichsgruppen VG_{OECD} und VG_{EU} sind die entsprechenden Anteile mit 12 Prozent beziehungsweise 9 Prozent höher als in Deutschland. Die Teilnehmerstaaten mit den größten Anteilen von Schülerinnen und Schülern in

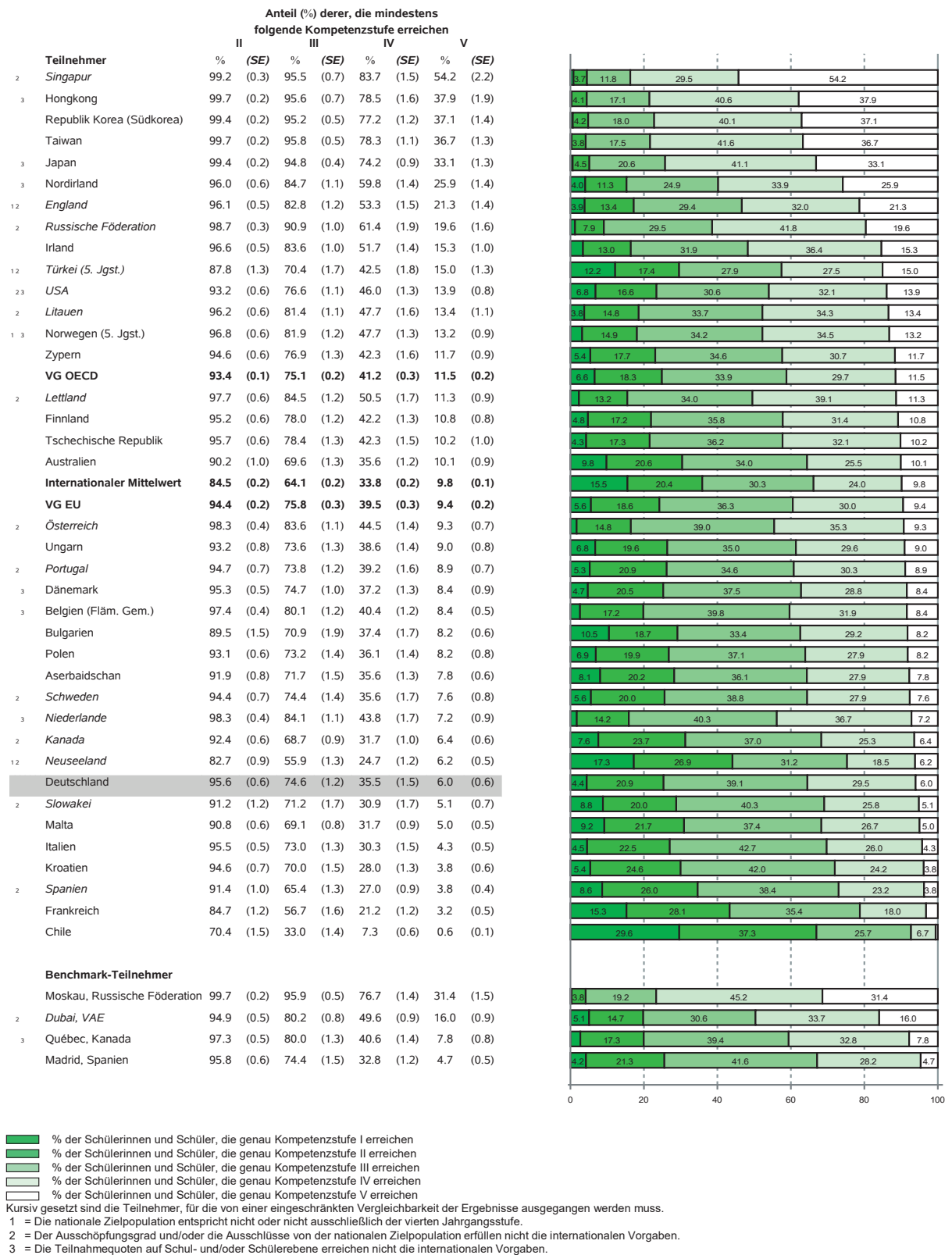
Abbildung 3.7: Prozentuale Verteilung der Schülerinnen und Schüler auf die fünf Kompetenzstufen (Mathematik) im internationalen Vergleich

Tabelle 3.15.1: Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007, 2011, 2015 und 2019 (Angaben in Prozent)

Fußnoten					TIMSS 2007 Kompetenzstufen										TIMSS 2011 Kompetenzstufen									
2007	2011	2015	2019 T	Teilnehmer ^A	I	II	III	IV	V	I	II	III	IV	V										
					% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)	% (SE)										
2 3	2	2	2	8 Singapur	1.8 (0.3)	6.4 (0.8)	17.3 (1.1)	33.7 (1.3)	40.7 (2.1)	1.1 (0.2)	5.0 (0.6)	15.9 (1.1)	35.1 (1.4)	42.9 (2.0)										
	2	3	3 8 Hongkong	0.4 (0.1)	2.8 (0.4)	16.1 (1.3)	40.3 (1.3)	40.4 (2.2)	0.8 (0.5)	3.1 (0.7)	15.8 (1.1)	42.9 (1.2)	37.5 (1.8)											
			Taiwan	1.1 (0.2)	6.5 (0.5)	26.4 (1.2)	41.9 (1.4)	24.1 (1.2)	1.0 (0.2)	5.8 (0.5)	19.7 (1.0)	40.0 (1.1)	33.6 (1.2)											
			3 8 Japan	2.2 (0.4)	8.7 (0.6)	27.7 (0.9)	38.7 (1.3)	22.8 (1.2)	1.0 (0.2)	5.8 (0.5)	22.7 (0.9)	40.8 (1.0)	29.7 (1.0)											
1 3	2	2	2 8 Russische Föderation	5.0 (0.7)	14.5 (1.3)	32.3 (1.4)	32.3 (1.2)	15.9 (1.8)	3.1 (0.6)	15.0 (1.0)	35.1 (1.1)	33.9 (1.1)	12.9 (1.4)											
	1 3	1 3	1 2 8 England	5.8 (0.7)	15.3 (0.9)	31.0 (1.1)	31.7 (0.9)	16.2 (1.2)	6.7 (0.7)	15.7 (1.1)	28.2 (1.1)	31.4 (1.4)	18.0 (1.3)											
	2	2	2 8 Litauen	5.6 (0.7)	17.1 (1.3)	35.4 (1.1)	32.2 (1.2)	9.7 (0.7)	4.3 (0.6)	16.6 (1.1)	36.3 (1.2)	32.5 (1.2)	10.3 (0.8)											
			2 8 Österreich	7.1 (0.8)	24.0 (1.1)	42.8 (1.3)	23.0 (0.9)	3.1 (0.3)	4.7 (0.8)	24.9 (1.5)	44.1 (1.3)	24.0 (1.4)	2.4 (0.3)											
3	3	3	3 8 Niederlande	2.0 (0.4)	14.2 (1.2)	42.1 (1.2)	35.1 (1.5)	6.6 (0.7)	0.6 (0.2)	10.9 (0.8)	44.0 (1.3)	39.9 (1.2)	4.6 (0.6)											
2 3	2 3	2 3	2 3 8 USA	4.8 (0.5)	18.4 (0.9)	36.9 (0.9)	29.8 (0.9)	10.0 (0.8)	3.9 (0.3)	15.1 (0.6)	34.1 (0.9)	34.1 (0.8)	12.8 (0.8)											
			2 8 Tschechische Republik	11.8 (1.1)	29.6 (1.2)	39.9 (1.3)	16.9 (1.5)	1.9 (0.4)	6.8 (0.8)	21.4 (1.1)	42.0 (1.4)	25.5 (1.4)	4.2 (0.5)											
3	2 3	2 3	3 8 Dänemark	4.7 (0.8)	19.0 (0.9)	40.3 (1.4)	29.3 (1.3)	6.6 (0.7)	3.4 (0.6)	14.7 (0.8)	37.7 (1.2)	34.6 (1.5)	9.7 (1.0)											
			Ungarn	11.8 (1.2)	21.1 (1.1)	31.9 (1.2)	26.2 (1.2)	9.0 (0.8)	10.5 (1.0)	19.1 (1.0)	33.2 (0.9)	27.7 (1.0)	9.5 (0.8)											
			2 8 Schweden	6.8 (0.7)	25.6 (1.0)	43.7 (1.0)	21.2 (1.3)	2.7 (0.3)	6.8 (0.7)	24.7 (1.1)	43.9 (1.4)	21.7 (1.1)	2.9 (0.4)											
			Deutschland	4.1 (0.5)	17.5 (0.9)	41.3 (0.9)	31.6 (1.1)	5.6 (0.5)	2.7 (0.6)	16.5 (1.1)	43.3 (1.3)	32.2 (1.1)	5.2 (0.5)											
2	3	2 3	8 Australien	8.5 (1.0)	21.0 (1.2)	35.6 (1.2)	26.1 (1.7)	8.8 (0.8)	9.7 (1.0)	20.2 (0.9)	35.0 (1.0)	25.3 (1.0)	9.8 (0.9)											
			8 Italien	8.9 (1.0)	23.9 (1.1)	38.2 (1.6)	23.5 (1.3)	5.5 (0.7)	7.2 (0.8)	24.0 (1.0)	40.5 (1.2)	23.6 (1.2)	4.6 (0.6)											
			2 8 Slowakei	11.9 (1.5)	24.9 (1.1)	36.8 (1.0)	21.6 (1.2)	4.8 (0.7)	9.6 (1.2)	21.6 (1.0)	38.3 (1.1)	25.2 (1.3)	5.2 (0.7)											
			1 2 8 Neuseeland	14.7 (1.0)	24.8 (0.9)	34.1 (0.9)	21.3 (1.0)	5.1 (0.5)	15.3 (0.8)	26.7 (1.0)	35.1 (0.8)	19.3 (0.9)	3.6 (0.5)											
2	2	2	2 8 Georgien	33.2 (2.0)	31.6 (1.2)	25.4 (1.2)	8.4 (0.9)	1.3 (0.4)	27.6 (1.7)	31.2 (1.3)	28.9 (1.5)	10.6 (1.0)	1.8 (0.5)											
			7 8 Iran	47.3 (2.0)	32.7 (1.4)	16.9 (1.2)	2.9 (0.5)	0.2 (0.1)	35.6 (1.5)	31.2 (0.9)	23.8 (0.9)	8.3 (0.7)	1.1 (0.2)											
Benchmark-Teilnehmer																								
2 3 6 2	2	2	2	8 Dubai, VAE	30.8 (1.3)	31.8 (1.7)	25.6 (1.1)	10.0 (0.6)	1.8 (0.3)	25.0 (0.9)	25.2 (1.0)	28.2 (0.8)	16.8 (0.7)	4.8 (0.5)										
2		2 3	3	8 Québec, Kanada	4.2 (0.6)	21.6 (1.3)	40.6 (1.8)	28.3 (1.8)	5.4 (0.7)	1.4 (0.2)	15.6 (1.1)	43.2 (1.3)	33.6 (1.3)	6.2 (0.8)										
2	2			8 Ontario, Kanada	5.9 (1.1)	22.9 (1.3)	42.1 (1.5)	24.8 (1.5)	4.3 (0.6)	5.9 (0.7)	21.3 (1.2)	38.7 (1.3)	27.5 (1.3)	6.6 (0.8)										

+ = Prozentwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)- = Prozentwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

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

6 = Abweichender Testzeitpunkt

7 = Teilnahme an TIMSS 2015 und TIMSS Numeracy

8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

Kompetenzstufe V sind Singapur (54%), Hongkong (38%), die Republik Korea (37%) und Taiwan (37%). Im europäischen Vergleich erreichen hier Nordirland (26%) und England (21%) die höchsten Werte. Der internationale Vergleich zeigt somit, dass anteilig deutlich weniger Viertklässlerinnen und Viertklässler aus Deutschland die höchste Kompetenzstufe erreichen als in den Staaten an der Spitze der Leistungsskala. Bis auf Italien sind es zudem weniger Lernende als in den anderen Staaten der „Leistungsgruppe Deutschland“.

Am anderen Ende des Spektrums beträgt der Anteil der Schülerinnen und Schüler aus Deutschland auf der untersten Kompetenzstufe I 4 Prozent. In den 19 Teilnehmerstaaten, für die signifikant bessere Testleistungen auf der Gesamtskala vermerkt sind, liegt – mit Ausnahme der USA (7%), Finnland und Zypern (jeweils 5%) sowie Nordirland, England, Litauen und der Tschechischen Republik (jeweils 4%) – der Anteil der Viertklässlerinnen und Viertklässler auf Kompetenzstufe I bei unter 4 Prozent. Bei den VG_{EU} (6%) und VG_{OECD} (7%) liegen die entsprechenden Werte leicht oberhalb des Wertes für Deutschland. In der „Leistungsgruppe Deutschland“ weist Deutschland den niedrigsten Wert auf. Für alle anderen Staaten sind höhere Werte zu beobachten: Portugal, Italien und Dänemark (jeweils 5%), Schweden (6%), Ungarn und Polen (jeweils 7%), Aserbaidschan (8%) Australien (10%), Bulgarien (11%) sowie Türkei (5. Jahrgangsstufe, 12%).

Tabelle 3.15.2: Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007, 2011, 2015 und 2019 (Angaben in Prozent)

Fußnoten					TIMSS 2015 Kompetenzstufen										TIMSS 2019 Kompetenzstufen																	
2007	2011	2015	2019	T	Teilnehmer ^A	I		II		III		IV		V		I		II		III		IV		V								
						%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)							
2	3	2	2	2	8	Singapur	1.4	(0.3)	5.2	(0.7)	13.4	(1.1)	29.9	(1.7)	50.1	(2.1)	0.8	(0.3)	3.7	(0.5)	11.8	(1.1)	29.5	(1.5)	54.2	(2.2)						
		2	3	3	3	Hongkong	0.2	(0.1)	2.0	(0.4)	+ 13.5	(1.1)	+ 39.5	(1.5)	44.8	(2.0)	–	0.3	(0.2)	4.1	(0.6)	17.1	(1.3)	40.6	(1.1)	37.9	(1.9)					
						Taiwan	0.5	(0.2)	4.7	(0.5)	19.1	(0.9)	40.4	(1.3)	35.3	(1.5)	0.3	(0.2)	3.8	(0.5)	17.5	(0.9)	41.6	(1.0)	36.7	(1.3)						
					3	Japan	0.5	(0.2)	4.4	(0.4)	20.7	(0.9)	42.2	(0.8)	32.2	(1.1)	0.6	(0.2)	4.5	(0.4)	20.6	(0.9)	41.1	(1.2)	33.1	(1.3)						
1	3	1	3	1	2	8	Russische Föderation	1.8	(0.4)	9.3	(1.0)	30.1	(1.3)	39.0	(1.6)	19.8	(1.8)	1.3	(0.3)	7.9	(0.8)	29.5	(1.3)	41.8	(1.3)	19.6	(1.6)					
					2	8	England	4.3	(0.7)	15.7	(1.1)	31.4	(1.2)	31.8	(1.2)	16.8	(1.2)	+	3.9	(0.5)	13.4	(1.0)	29.4	(1.5)	32.0	(1.1)	21.3	(1.4)				
					2	2	8	Litauen	3.7	(0.5)	15.7	(0.9)	36.7	(1.2)	34.3	(1.1)	9.7	(1.0)	+	3.8	(0.6)	14.8	(1.0)	33.7	(1.2)	34.3	(1.2)	13.4	(1.1)			
					2	2	2	Österreich												1.7	(0.4)	14.8	(1.0)	39.0	(1.1)	35.3	(1.2)	9.3	(0.7)			
2	3	2	3	2	3	8	Niederlande	1.2	(0.3)	15.8	(1.0)	45.7	(1.3)	– 33.5	(1.3)	3.8	(0.6)	+	1.7	(0.4)	14.2	(0.9)	40.3	(1.6)	36.7	(1.5)	7.2	(0.9)				
					2	3	8	USA	5.3	(0.5)	+ 16.1	(0.6)	32.0	(0.7)	32.5	(0.8)	14.2	(0.8)	6.8	(0.6)	16.6	(0.8)	30.6	(0.7)	32.1	(1.1)	13.9	(0.8)				
					2	2	2	Tschechische Republik	3.9	(0.5)	17.7	(0.9)	40.1	(1.1)	– 30.5	(1.2)	7.8	(0.7)	+	4.3	(0.6)	17.3	(1.0)	36.2	(1.1)	32.1	(1.1)	10.2	(1.0)			
					3	2	3	8	Dänemark	4.1	(0.6)	15.6	(0.9)	+ 34.6	(1.3)	33.7	(1.3)	– 12.0	(1.0)	–	4.7	(0.5)	20.5	(0.8)	37.5	(1.4)	28.8	(0.9)	8.4	(0.9)		
2	3	2	2	2			Ungarn	8.3	(0.9)	16.8	(0.9)	+ 31.1	(1.3)	+ 31.2	(1.3)	12.6	(0.9)	–	6.8	(0.8)	19.6	(1.0)	35.0	(1.1)	29.6	(1.2)	9.0	(0.8)				
					2	2	8	Schweden	5.3	(0.8)	19.8	(1.2)	41.2	(1.3)	28.5	(1.3)	5.2	(0.5)	+	5.6	(0.7)	20.0	(1.0)	38.8	(1.2)	27.9	(1.4)	7.6	(0.8)			
							8	Deutschland	3.7	(0.6)	19.6	(0.9)	42.7	(1.1)	– 28.7	(1.3)	5.3	(0.6)	4.4	(0.6)	20.9	(1.1)	39.1	(1.0)	29.5	(1.4)	6.0	(0.6)				
					2	3	2	3	Australien	8.6	(0.9)	21.1	(0.9)	34.1	(1.2)	27.0	(1.2)	9.2	(0.9)	9.8	(1.0)	20.6	(1.0)	34.0	(1.1)	25.5	(1.0)	10.1	(0.9)			
1	2	2	2	2			Italien	7.5	(0.8)	– 23.8	(1.1)	40.7	(1.1)	23.8	(1.3)	4.2	(0.6)	4.5	(0.5)	22.5	(1.2)	42.7	(1.0)	26.0	(1.3)	4.3	(0.5)					
					2	2	2	Slowakei	11.7	(0.9)	23.3	(1.0)	– 38.9	(1.2)	22.2	(1.0)	+	4.0	(0.4)	8.8	(1.2)	20.0	(1.2)	40.3	(1.3)	25.8	(1.5)	5.1	(0.7)			
					2	2	2	8	Neuseeländ	16.0	(0.9)	25.3	(1.0)	32.3	(0.9)	20.5	(0.7)	5.9	(0.5)	17.3	(0.9)	26.9	(0.9)	31.2	(0.9)	18.5	(1.0)	6.2	(0.5)			
					2	2	2	2	Georgien	22.4	(1.7)	– 30.3	(1.3)	31.9	(1.4)	+ 13.4	(1.1)	+ 2.0	(0.6)	16.2	(1.4)	28.0	(1.2)	35.7	(1.3)	17.3	(1.3)	2.8	(0.4)			
2	2	2	2	2	7	8	Iran	34.8	(1.4)	29.1	(0.9)	25.1	(0.9)	9.6	(0.6)	1.5	(0.3)	31.9	(1.5)	29.2	(0.9)	25.7	(1.1)	11.3	(0.8)	1.9	(0.3)					
							8	Benchmark-Teilnehmer																								
					2	3	6	2	2	2	Dubai, VAE	12.6	(0.5)	– 21.0	(0.7)	– 31.0	(1.0)	24.8	(0.8)	+ 10.6	(0.5)	+	5.1	(0.5)	14.7	(0.7)	30.6	(1.0)	33.7	(1.2)	16.0	(0.9)
					2	2	2	3	8	Québec, Kanada	2.2	(0.6)	15.6	(1.6)	40.6	(1.6)	32.9	(1.9)	8.7	(1.3)	2.7	(0.5)	17.3	(1.1)	39.4	(1.1)	32.8	(1.4)	7.8	(0.8)		
2	2	2	2	2			Ontario, Kanada	6.7	(0.6)	23.2	(1.1)	39.4	(1.2)	– 25.2	(1.1)	5.7	(0.6)	7.7	(0.9)	24.6	(1.3)	35.4	(1.4)	25.0	(1.6)	7.2	(1.0)					

Ein Vergleich der Daten von 2007 und 2019 zeigt, dass für die USA für Kompetenzstufe I sowie für Deutschland für Kompetenzstufe II als jeweils einziger Teilnehmer eine signifikante Steigerung der Anteile der Lernenden auf der Kompetenzstufe zu verzeichnen ist. Im Kontrast dazu sind bei zwölf beziehungsweise zehn Staaten signifikante Reduzierungen der Anteile auf diesen Kompetenzstufen festzustellen. Bei acht beziehungsweise elf Teilnehmerstaaten können signifikante Steigerungen der Anteile auf der Kompetenzstufe IV beziehungsweise V verzeichnet werden, Deutschland gehört nicht dazu.

Vergleicht man die Daten von 2011 und 2019, so ist festzustellen, dass für drei Teilnehmer, unter anderem Deutschland, eine signifikante Steigerung der Anteile der Lernenden auf der Kompetenzstufen I beziehungsweise II zu verzeichnen ist. Hingegen können bei acht beziehungsweise sechs Staaten signifikante Reduzierungen der Anteile der Lernenden auf diesen Kompetenzstufen beobachtet werden. Bei sechs beziehungsweise neun Teilnehmerstaaten sind signifikante Steigerungen der Anteile auf der Kompetenzstufe IV beziehungsweise V zu verzeichnen, für Deutschland nicht. Allerdings zeigt sich ein signifikanter Rückgang der Anteile der Lernenden auf der Kompetenzstufe III.

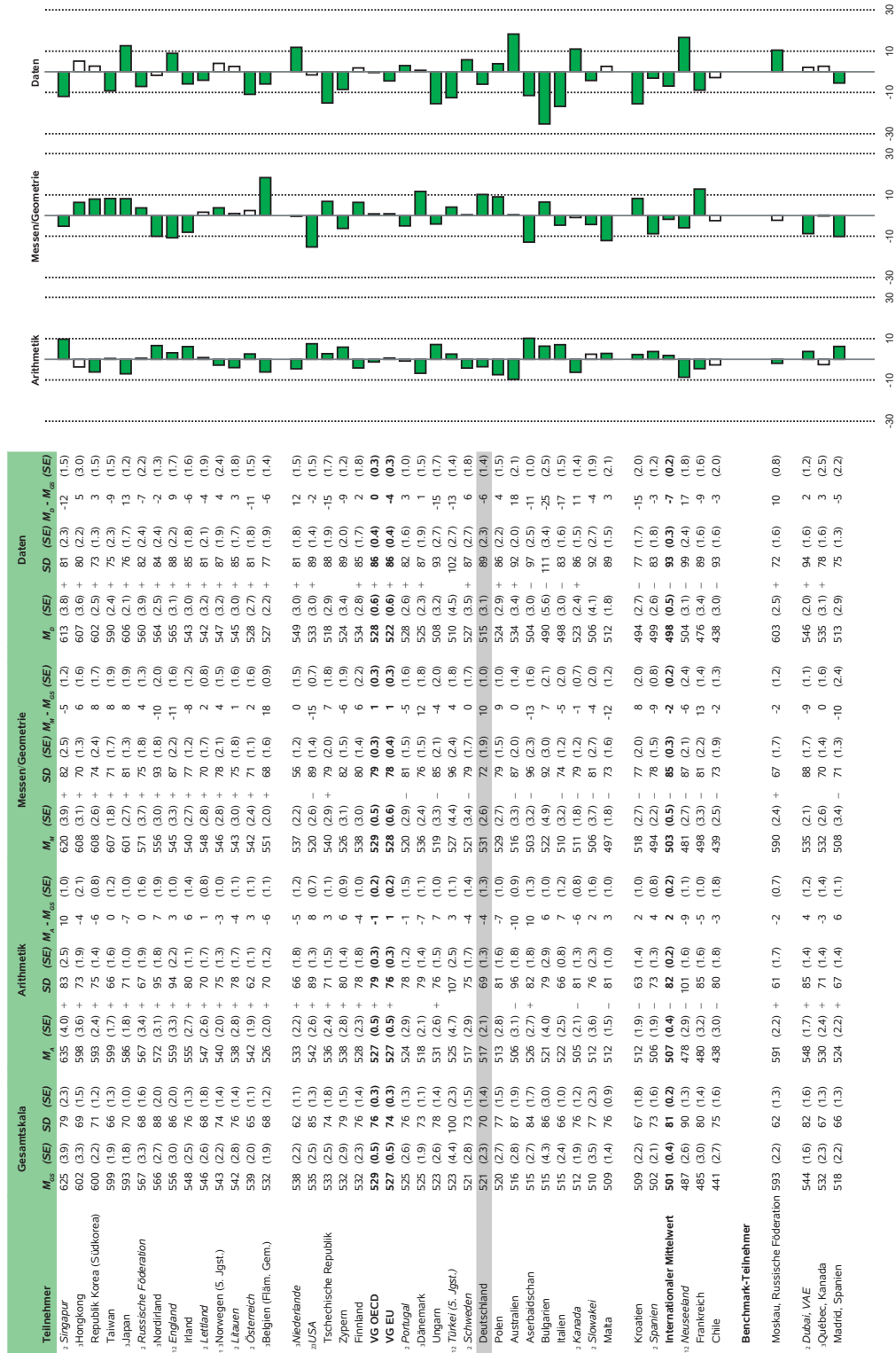
Schließlich ergeben die Analysen für den Vergleich der Untersuchungen TIMSS 2015 und 2019 nicht so klare Unterschiede wie im Vergleich von TIMSS 2019 mit den anderen TIMSS-Untersuchungen. In einem beziehungsweise drei Staaten ist eine signifikante Steigerung und in zwei beziehungsweise einem Teilnehmerstaat eine signifikante Reduzierung der Anteile der Lernenden auf der Kompetenzstufe I beziehungsweise II zu verzeichnen. Bei zwei beziehungsweise fünf Teilnehmerstaaten können signifikante Steigerungen und bei einem beziehungsweise drei Staaten signifikante Abnahmen der Anteile auf den Kompetenzstufen IV beziehungsweise V verzeichnet werden. Für Deutschland lässt sich allein ein signifikanter Rückgang der Anteile der Lernenden auf der Kompetenzstufe III feststellen.

3.5.3 Inhaltsbereiche

Anhand der Ergebnisse von TIMSS können auch relative Stärken und Schwächen in Bezug auf die verschiedenen Inhaltsbereiche identifiziert werden. Daher soll nun ein differenzierter Blick auf die Ergebnisse in den Inhaltsbereichen *Arithmetik*, *Messen und Geometrie* sowie *Daten* geworfen werden.

Kompetenzen in den Inhaltsbereichen im internationalen Vergleich: Abbildung 3.8 zeigt die Leistungsmittelwerte in den drei Inhaltsbereichen für die EU- und OECD-Staaten sowie für diejenigen Teilnehmer, die auf der Gesamtskala Mathematik signifikant bessere oder zu Deutschland keine signifikant unterschiedlichen Ergebnisse erzielten. Dabei handelt es sich insgesamt um 36 Teilnehmerstaaten sowie drei *Benchmark*-Teilnehmer. Darüber hinaus sind auch die Mittelwerte aller TIMSS-Teilnehmerstaaten (internationaler Mittelwert) sowie diejenigen der VG_{OECD} und VG_{EU} vermerkt. Im linken Teil der Abbildung sind für jedes Teilnehmerland die erzielten Leistungsmittelwerte (M) in den drei Inhaltsbereichen und die Standardabweichungen (SD) sowie die Standardfehler (SE) dieser beiden Werte aufgeführt. Darüber hinaus sind im rechten Teil der Abbildung die Abweichungen der Mittelwerte der Inhaltsbereiche zur Gesamtskala visualisiert. Erneut werden signifikant bessere Leistungswerte anderer Teilnehmerstaaten

Abbildung 3.8: Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den Inhaltsbereichen *Arithmetik, Messen und Geometrie* sowie *Daten* im Vergleich zur Gesamtskala



1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten.
2 = Der Ausschöfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation
3 = Die Teilnehmenden auf Schul- und/oder Schülerebene erreichen nicht die intern

Keine statistisch signifikanten Unterschiede zur Gesamtskala ($p > 0,05$)

Statistisch signifikante Unterschiede zur Gesamtskala ($p \leq 0,05$)

+ = Mittelwert statistisch signifikant höher als der Mittelwert von Deutschland ($p \leq 0,05$)

- = Mittelwert statistisch signifikant niedriger als der Mittelwert von Deutschland ($p \leq 0,05$)

Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss (für die Besonderheiten von Chile siehe Anhang).

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

durch ein Pluszeichen markiert, während signifikant schlechtere Testergebnisse mit einem Minuszeichen gekennzeichnet sind.

Zu den Ergebnissen: Die Viertklässlerinnen und Viertklässler in Deutschland erzielten im Bereich *Arithmetik* 517 Punkte, im Bereich *Messen und Geometrie* 531 Punkte und im Bereich *Daten* 515 Punkte. Die Differenz zwischen den Leistungswerten in *Messen und Geometrie* einerseits und *Daten* andererseits beträgt 16 Punkte. Die Verteilung der Schülerinnen- und Schülerleistungen über die drei Inhaltsbereiche ist somit heterogen. Die Mittelwerte in den Inhaltsbereichen *Arithmetik* und *Daten* fallen signifikant niedriger und im Inhaltsbereich *Messen und Geometrie* signifikant höher aus als der Gesamtmittelwert Deutschlands (521).

Der Vergleich mit den Staaten der ‚Leistungsgruppe Deutschland‘ zeigt, dass lediglich für Portugal (8) und Schweden (10) eine homogenere Leistungsverteilung über die drei Inhaltsbereiche zu verzeichnen ist. Für Polen (16) liegt derselbe Wert wie für Deutschland vor. Für alle anderen Staaten, die auf der Gesamtskala keine zu Deutschland signifikanten Unterschiede der Testergebnisse aufweisen, ist hingegen eine größere Heterogenität zu erkennen. So liegen für die Türkei (17), Dänemark (18), Ungarn (23), Italien (24), Australien (28), und Bulgarien (32) größere Differenzen zwischen den Leistungswerten der Inhaltsbereiche vor.

Im internationalen Vergleich zeigt sich darüber hinaus, dass 14 Teilnehmerstaaten in allen drei Inhaltsbereichen signifikant bessere Leistungsmittelwerte als Deutschland erreichen, während fünf Teilnehmer in allen drei Inhaltsbereichen signifikant schwächer abschneiden.

Hinsichtlich des Vergleichs der Leistungsmittelwerte Deutschlands in den drei Inhaltsbereichen mit denen der VG_{EU} und VG_{OECD} ergeben sich folgende Erkenntnisse: Keine signifikanten Unterschiede liegen für den Bereich *Messen und Geometrie* vor. Demgegenüber sind die Leistungswerte in Deutschland in *Arithmetik* sowie *Daten* signifikant niedriger als die für die Vergleichsgruppen der EU und OECD-Staaten.

Vergleich der Kompetenzen in den Inhaltsbereichen 2007, 2011, 2015 und 2019: Tabelle 3.16 und Abbildung 3.9 zeigen die nach Inhaltsbereichen getrennten mathematischen Kompetenzen für die vier TIMS-Studienzyklen. Aus Tabelle 3.16 sind die Leistungsmittelwerte der Teilnehmerstaaten ersichtlich.

Abbildung 3.9 zeigt die Veränderungen der Leistungsmittelwerte über die Studienzyklen hinweg. Im oberen Teil dieser Abbildung sind die Differenzwerte zwischen den Zyklen numerisch dargestellt. Liegen positive Werte vor, wurden höhere Punktwerte im Vergleich zum jeweils früheren Zyklus erzielt. Negative Werte deuten entsprechend auf das Gegenteil hin. Im unteren Teil der Abbildung folgt eine grafische Darstellung der Differenzen der Ergebnisse von 2007 und 2019 mittels Differenzbalken. Weiße Balken deuten auf statistisch nicht signifikante und grüne auf signifikante Veränderungen zwischen den Erhebungszyklen hin. Wie in Abbildung 3.6 sowie in den Tabellen 3.12 und 3.13 zur Beschreibung von Trendvergleichen werden auch in Tabelle 3.16 und Abbildung 3.9 die teilnehmenden Staaten und Regionen gelistet, die an den vier oben angegebenen TIMSS-Zyklen teilgenommen haben. Außerdem wurde Österreich wieder in die Aufstellung aufgenommen, obwohl es an TIMSS 2015 nicht teilgenommen hat. Die Reihung der Staaten entspricht den Veränderungen der Leistungsmittelwerte auf der Gesamtskala (siehe Abbildung 3.6).

Tabelle 3.16: Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 in den Inhaltsbereichen *Arithmetik, Messen und Geometrie* und *Daten*¹

Fußnoten			Arithmetik				Messen/Geometrie				Daten					
2007	2011	2015	2019 T	Teilnehmer ⁴	M_{gr} (SE)	M_{T1} (SE)	M_{T5} (SE)	M_{T9} (SE)	2007 ^B	2011 ^B	2015	2019	M_{gr} (SE)	M_{T1} (SE)	2015	2019
2	2	2	2	Tschechische Republik	486 (2.9)	509 (2.5)	528 (2.4)	536 (2.4)	487 (3.3)	513 (3.0)	531 (2.5)	540 (2.9)	482 (4.2)	519 (3.1)	525 (3.0)	518 (2.9)
				Georgien	470 (3.7)	473 (3.1)	483 (3.5)	501 (3.6)	395 (5.9)	411 (4.3)	429 (4.6)	470 (4.1)	390 (5.4)	433 (4.0)	435 (4.4)	444 (4.6)
				Iran	407 (3.5)	440 (3.3)	435 (3.2)	446 (4.0)	408 (3.9)	435 (3.9)	428 (3.5)	445 (3.6)	374 (5.0)	397 (4.3)	416 (3.2)	424 (3.8)
2	2	2	Österreich	506 (2.1)	506 (2.4)		542 (1.9)	506 (2.6)	512 (3.4)		542 (2.4)		515 (3.2)		528 (2.7)	
			Singapur	611 (4.1)	619 (3.4)	630 (4.2)	635 (4.0)	584 (4.4)	589 (3.6)	607 (4.2)	620 (3.9)	597 (3.9)	588 (3.4)	600 (4.1)	613 (3.8)	
			Japan	564 (2.1)	584 (1.6)	592 (1.9)	586 (1.8)	575 (2.6)	589 (2.0)	601 (2.5)	601 (2.7)	588 (3.5)	590 (2.9)	593 (2.6)	606 (2.1)	
2	2	2	Taiwan	583 (1.8)	599 (2.0)	599 (1.8)	599 (1.7)	566 (2.7)	573 (2.1)	597 (3.0)	607 (1.8)	576 (2.3)	600 (2.6)	591 (2.2)	590 (2.4)	
			Russische Föderation	549 (4.4)	545 (3.3)	567 (3.3)	567 (3.4)	543 (6.2)	542 (4.3)	557 (4.4)	571 (3.7)	529 (6.2)	533 (4.1)	573 (3.6)	560 (3.9)	
			Schweden	495 (2.5)	500 (2.2)	514 (2.7)	517 (2.9)	503 (2.9)	500 (2.4)	523 (3.3)	521 (3.4)	527 (3.2)	523 (3.0)	529 (3.9)	527 (3.5)	
1	3	1	England	535 (3.1)	539 (3.7)	547 (3.2)	559 (3.3)	552 (3.3)	545 (3.9)	542 (3.3)	545 (3.3)	551 (3.1)	549 (4.6)	552 (3.2)	565 (3.1)	
			Slowakei	500 (3.9)	511 (3.7)	502 (2.4)	512 (3.6)	494 (5.3)	500 (4.3)	491 (2.6)	506 (3.7)	482 (5.4)	504 (4.6)	496 (3.8)	506 (4.1)	
			Ungarn	515 (3.5)	515 (3.2)	531 (3.0)	531 (2.6)	507 (3.9)	520 (3.6)	536 (3.6)	519 (3.3)	497 (4.2)	510 (4.2)	513 (3.6)	508 (3.2)	
2	2	2	Litauen	536 (2.2)	537 (2.4)	539 (2.8)	538 (2.8)	518 (3.0)	531 (3.0)	527 (3.2)	543 (3.0)	529 (3.6)	526 (3.0)	542 (4.0)	545 (3.0)	
			Italien	510 (3.0)	510 (2.7)	510 (2.4)	522 (2.5)	507 (3.6)	513 (3.1)	503 (2.8)	510 (3.2)	499 (4.1)	495 (3.1)	498 (2.9)	498 (3.0)	
			USA	529 (2.6)	543 (2.0)	546 (2.2)	542 (2.6)	522 (3.0)	535 (2.2)	525 (2.6)	520 (2.6)	546 (2.9)	545 (1.8)	540 (2.8)	533 (3.0)	
3	3	3	Niederlande	539 (2.2)	543 (1.7)	531 (2.2)	533 (2.2)	522 (2.7)	524 (2.9)	522 (1.9)	537 (2.2)	545 (2.8)	559 (2.9)	539 (3.4)	549 (3.0)	
			Dänemark	513 (2.9)	534 (2.4)	535 (2.7)	518 (2.1)	546 (3.2)	548 (3.0)	555 (3.2)	536 (2.4)	527 (4.2)	532 (3.0)	526 (3.5)	525 (2.3)	
			Australien	503 (3.5)	508 (3.2)	509 (3.1)	506 (3.1)	536 (3.7)	534 (3.0)	527 (3.3)	516 (3.3)	532 (4.1)	515 (3.1)	533 (3.6)	534 (3.4)	
Deutschland				524 (2.2)	520 (2.3)	515 (2.1)	517 (2.1)	527 (2.4)	536 (2.6)	531 (2.5)	531 (2.6)	532 (3.7)	546 (2.8)	535 (2.6)	515 (3.1)	
2	3	3	Hongkong	608 (3.7)	604 (3.3)	616 (3.1)	598 (3.6)	613 (3.7)	605 (3.4)	617 (3.4)	608 (3.1)	600 (3.4)	593 (3.6)	611 (3.8)	607 (3.6)	
			Neuseeland	485 (2.6)	483 (2.5)	485 (2.7)	478 (2.9)	495 (2.6)	483 (2.5)	489 (2.8)	481 (2.7)	506 (3.0)	491 (2.7)	506 (2.9)	504 (3.1)	
			Benchmark-Teilnehmer													
2	3	2	Dubai, VAE	452 (2.0)	474 (1.7)	514 (1.5)	548 (1.7)	424 (3.3)	449 (2.3)	503 (1.9)	535 (2.1)	444 (3.1)	471 (3.1)	517 (1.7)	546 (2.0)	
			Québec, Kanada	515 (3.0)	531 (2.6)	533 (4.2)	530 (2.4)	524 (3.9)	536 (3.2)	542 (4.6)	532 (2.6)	523 (4.4)	538 (3.7)	541 (5.0)	535 (3.1)	
			Ontario, Kanada	495 (3.5)	504 (3.4)	500 (2.6)	501 (3.6)	530 (3.6)	535 (3.4)	526 (2.9)	516 (3.2)	545 (4.0)	536 (3.5)	536 (2.6)	527 (4.0)	

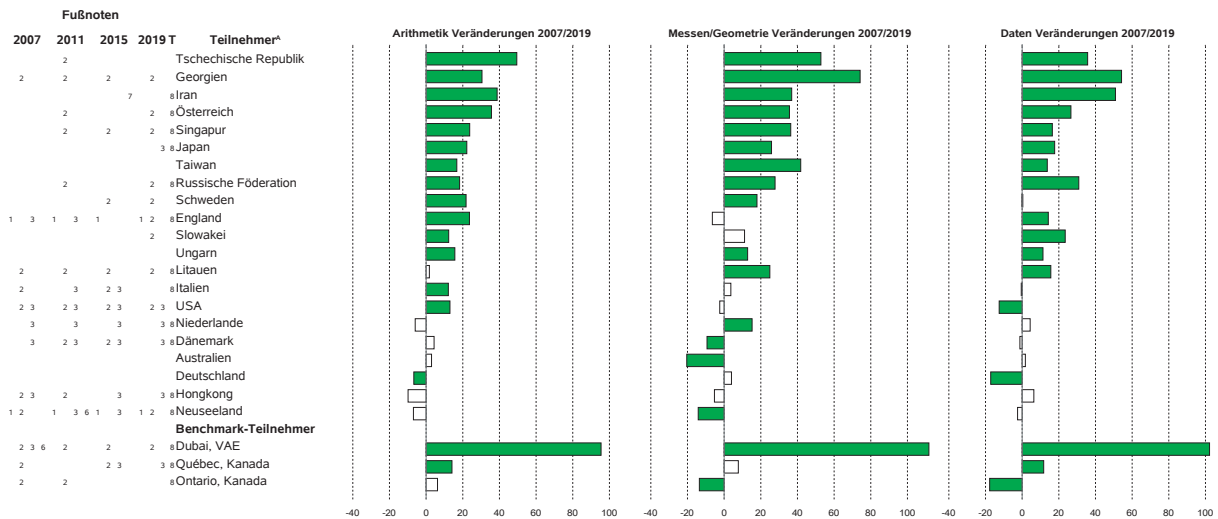
1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
6 = Abweichender Testzeitpunkt
7 = Teilnahme an TIMSS 2015 und TIMSS Numeracy
8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019
A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.
B = Abweichungen in den berichteten Standardfehlern zur internationalen Berichterstattung sind in einem differierten Berechnungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Abbildung 3.9: Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 in den Inhaltsbereichen *Arithmetik, Messen und Geometrie* und *Daten II*

Fußnoten					Teilnehmer ^a	Arithmetik Veränderungen ^b			Messen/Geometrie Veränderungen ^b			Daten Veränderungen ^a			
2007	2011	2015	2019	T		$M_{19}-M_{07}$ (SE)	$M_{19}-M_{11}$ (SE)	$M_{19}-M_{15}$ (SE)	$M_{19}-M_{07}$ (SE)	$M_{19}-M_{11}$ (SE)	$M_{19}-M_{15}$ (SE)	$M_{19}-M_{07}$ (SE)	$M_{19}-M_{11}$ (SE)	$M_{19}-M_{15}$ (SE)	
					Tschechische Republik	49 (3.8) +	27 (3.5) +	8 (3.4) +	53 (4.4) +	27 (4.2) +	9 (3.8) +	36 (5.1) +	-1 (4.2)	-7 (4.1)	
2	2	2	2	2	Georgien	30 (5.1) +	28 (4.7) +	18 (5.0) +	74 (7.2) +	58 (5.9) +	41 (6.1) +	54 (7.1) +	12 (6.1)	10 (6.4)	
			7		Iran	39 (5.3) +	6 (5.2)	11 (5.1) +	37 (5.3) +	11 (5.3) +	17 (5.0) +	51 (6.3) +	27 (5.8) +	9 (5.0)	
	2			2	Österreich	36 (2.9) +	36 (3.1) +		36 (3.5) +	30 (4.1) +		27 (4.3) +	13 (4.2) +		
	2	2	2	2	Singapur	24 (5.8) +	16 (5.3) +	5 (5.8)	36 (5.8) +	31 (5.3) +	13 (5.7) +	16 (5.5) +	26 (5.1) +	14 (5.6) +	
				3	Japan	22 (2.8) +	2 (2.4)	-6 (2.7) -	26 (3.7) +	12 (3.3) +	0 (3.6)	18 (4.1) +	16 (3.6) +	12 (3.3) +	
					Taiwan	17 (2.5) +	0 (2.6)	0 (2.4)	42 (3.2) +	35 (2.8) +	11 (3.5) +	14 (3.3) +	-10 (3.5) -	-1 (3.2)	
	2			2	Russische Föderation	18 (5.5) +	23 (4.7) +	1 (4.7)	28 (7.2) +	29 (5.6) +	14 (5.7) +	31 (7.3) +	27 (5.6) +	-13 (5.3) -	
		2		2	Schweden	22 (3.8) +	17 (3.6) +	3 (3.9)	18 (4.5) +	22 (4.1) +	-1 (4.7)	0 (4.7)	4 (4.6)	-2 (5.2)	
1	3	1	3	1	England	24 (4.6) +	20 (5.0) +	12 (4.6) +	-6 (4.7)	0 (5.1)	3 (4.6)	14 (4.4) +	15 (5.6) +	13 (4.5) +	
				2	Slowakei	12 (5.3) +	1 (5.1)	10 (4.3) +	11 (6.5)	6 (5.6)	15 (4.5) +	23 (6.8) +	2 (6.2)	10 (5.6)	
					Ungarn	16 (4.4) +	16 (4.1) +	-1 (4.0)	13 (5.1) +	-1 (4.9)	-17 (4.9) -	11 (5.3) +	-2 (5.3)	-5 (4.8)	
2	2	2		2	Litauen	2 (3.6)	1 (3.7)	-1 (4.0)	25 (4.2) +	12 (4.2) +	16 (4.4) +	16 (4.7) +	18 (4.2) +	3 (5.0)	
	2	3	2	3	Italien	12 (3.9) +	12 (3.6) +	12 (3.4) +	4 (4.8)	-2 (4.4)	7 (4.2)	0 (5.1)	4 (4.4)	0 (4.2)	
2	3	2	3	2	USA	13 (3.7) +	-1 (3.3)	-3 (3.4)	-2 (3.9)	-15 (3.4) -	-6 (3.7)	-13 (4.2) -	-12 (3.5) -	-7 (4.1)	
3	3	3		3	Niederlande	-6 (3.1)	-10 (2.7) -	2 (3.1)	15 (3.5) +	13 (3.6) +	15 (2.9) +	4 (4.1)	-10 (4.1) -	-11 (4.5)	
	3	2	3	3	Dänemark	4 (3.5)	-16 (3.2) -	-17 (3.4) -	-9 (4.0) -	-12 (3.8) -	-19 (4.0) -	-1 (4.8)	-6 (3.8)	-1 (4.2)	
					Australien	3 (4.7)	-2 (4.5)	-3 (4.4)	-20 (4.9) -	-18 (4.4) -	-11 (4.7) -	2 (5.4)	19 (4.6) +	2 (5.0)	
					Deutschland	-7 (3.1) -	-3 (3.1)	2 (3.0)	4 (3.5)	-5 (3.7)	0 (3.6)	-17 (4.8) -	-31 (4.2) -	-20 (4.0) -	
2	3	2		3	Hongkong	-10 (5.1)	-6 (4.9)	-18 (4.7) -	-5 (4.8)	3 (4.6)	-9 (4.6)	6 (5.0)	13 (5.1) +	-4 (5.3)	
1	2	1	3	6	1	Neuseeland	-7 (3.9)	-4 (3.8)	-7 (3.9)	-14 (3.7) -	-2 (3.7)	-7 (3.9)	-3 (4.3)	13 (4.1) +	-2 (4.2)
Benchmark-Teilnehmer															
2	3	6	2	2	Dubai, VAE	95 (2.6) +	74 (2.4) +	34 (2.3) +	112 (4.0) +	86 (3.2) +	33 (2.9) +	102 (3.7) +	75 (3.7) +	30 (2.6) +	
2			2	3	Quebec, Kanada	14 (3.8) +	-2 (3.6)	-3 (4.8)	8 (4.7)	-4 (4.1)	-11 (5.3) -	12 (5.3) +	-3 (4.8)	-6 (5.9)	
2	2				Ontario, Kanada	6 (5.0)	-2 (4.9)	2 (4.4)	-14 (4.8) -	-19 (4.7) -	-10 (4.4) -	-18 (5.6) -	-9 (5.3)	-9 (4.7)	



□ Keine statistisch signifikanten Veränderungen zwischen 2007 und 2019 ($p > .05$)

■ Statistisch signifikante Veränderungen zwischen 2007 und 2019 ($p \leq .05$)

+ = Mittelwert in 2019 statistisch signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

- = Mittelwert in 2019 statistisch signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

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

6 = Abweichender Testzeitpunkt

7 = Teilnahme an TIMSS 2015 und TIMSS Numeracy

8 = Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

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

Aus Abbildung 3.9 geht für TIMSS 2019 hervor, dass in Deutschland im Bereich *Arithmetik* signifikant schwächere Testleistungen als in 2007 erzielt wurden, was im Vergleich von 2019 mit 2015 beziehungsweise 2011 nicht der Fall ist. Im Bereich *Messen und Geometrie* sind im Vergleich der Ergebnisse von TIMSS 2019 mit denen der vorangegangenen Zyklen keine statistisch bedeutsamen Veränderungen zu verzeichnen. Hingegen sind im Bereich *Daten* im Vergleich von TIMSS 2019 zu allen anderen Studienzyklen signifikant schlechtere Punktwerte zu verzeichnen.

Im internationalen Vergleich der Daten von 2019 und 2007 zeigt sich folgendes Bild: In *Arithmetik* ist Deutschland der einzige Teilnehmerstaat, für den in

2019 ein signifikant niedrigerer Leistungswert als 2007 vorliegt. Hingegen sind für 14 Staaten signifikant höhere Werte festzustellen. In *Messen und Geometrie* sind für zwölf Staaten signifikant höhere Leistungswerte festzustellen, während für zwei Staaten signifikant niedrigere Werte vorliegen. Für sechs Staaten – so auch für Deutschland – lassen sich keine Unterschiede feststellen. Im Inhaltsbereich *Daten* sind in 2019 für zwölf Staaten signifikant höhere, für zwei Staaten signifikant niedrigere und für sechs Teilnehmerstaaten keine Unterschiede bei den Leistungswerten festzustellen. Letztlich ist festzuhalten, dass für neun Staaten in allen Inhaltsbereichen, für fünf Staaten in zwei Inhaltsbereichen und für drei Staaten in einem Inhaltsbereich signifikant höhere Leistungswerte in 2019 als noch in 2007 vorliegen. Für vier Staaten sind in einem Inhaltsbereich signifikant schwächere Testleistungen in 2019 im Vergleich zu 2007 festzustellen. Deutschland ist der einzige Staat, der in zwei Inhaltsbereichen signifikant schwächer abschneidet.

Der Vergleich der Daten der Zyklen von 2019 und 2011 zeigt, dass neun Teilnehmerstaaten in 2019 signifikant besser und zwei signifikant schlechter in *Arithmetik* abschneiden. Für *Messen und Geometrie* gilt dies für elf beziehungsweise drei Teilnehmerstaaten und für *Daten* für zehn beziehungsweise vier Staaten. Drei Staaten schneiden in allen drei Inhaltsbereichen in 2019 besser als in 2011 ab. Für sechs Teilnehmerstaaten – so auch Deutschland – ist in mindestens einem Inhaltsbereich ein signifikant niedrigerer Leistungswert in 2019 im Vergleich zu 2011 festzustellen.

Die Daten der TIMS-Studien aus 2019 und 2015 zeigen zudem, dass sechs Teilnehmerstaaten in 2019 signifikant höhere und drei signifikant niedrigere Leistungswerte in *Arithmetik* erzielen. Für *Messen und Geometrie* gilt dies für neun beziehungsweise drei Staaten und für *Daten* für vier beziehungsweise zwei Staaten. Anders als bei den Vergleichen des Zyklus 2019 zu den Zyklen von 2007 und 2011 ist für keinen Teilnehmerstaat ein signifikant höherer Leistungswert in allen drei Inhaltsbereichen festzustellen. Für sieben Staaten, zu denen auch Deutschland gehört, liegt in mindestens einem Inhaltsbereich für 2019 ein signifikant niedrigerer Leistungswert im Vergleich zu 2015 vor. Gegenteiliges gilt für 14 Teilnehmerstaaten.

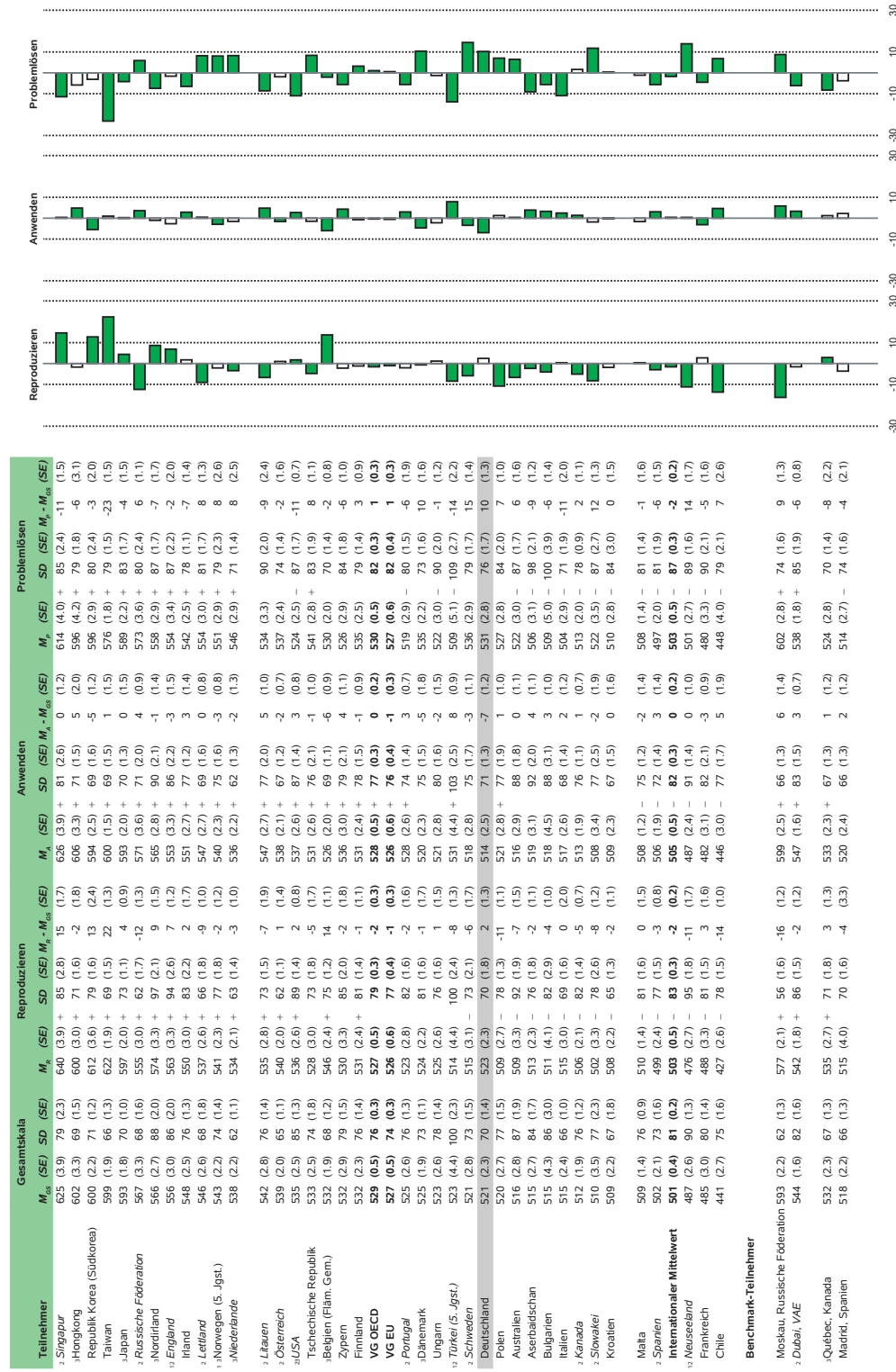
3.5.4 Kognitive Anforderungsbereiche

In diesem Abschnitt werden die Testergebnisse in Bezug auf die kognitiven Anforderungsbereiche dargestellt und im Vergleich zu den drei vorangegangenen Zyklen eingeordnet.

Kompetenzen in den Anforderungsbereichen im internationalen Vergleich: Abbildung 3.10 zeigt die nach *Reproduzieren*, *Anwenden* und *Problemlösen* gegliederten Leistungsmittelwerte der Teilnehmenden. Die Gestaltung der Abbildung orientiert sich an der von Abbildung 3.8.

Die Analyse zeigt, dass Schülerinnen und Schüler in Deutschland am Ende ihrer Grundschulzeit im Bereich *Reproduzieren* 523 Punkte, im Bereich *Anwenden* 514 Punkte und im Bereich *Problemlösen* 531 Punkte erzielen. Somit liegen die Testleistungen im Bereich *Problemlösen* signifikant über dem Gesamtmittelwert Deutschlands in Mathematik (521), während die Testleistungen aus dem Bereich *Anwenden* signifikant unterhalb dieses Wertes einzuordnen sind. Der Unterschied zwischen *Problemlösen* und *Anwenden* beträgt 17 Punkte.

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



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

Keine statistisch signifikanten Unterschiede zur Gesamtskala ($p > .05$)
 Statistisch signifikante Unterschiede zur Gesamtskala ($p \leq .05$)
 + = Mittelwert statistisch signifikant höher als der Mittelwert von Deutschland ($p \leq .05$)
 - = Mittelwert statistisch signifikant niedriger als der Mittelwert von Deutschland ($p \leq .05$)

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Im Vergleich zu den Teilnehmerstaaten, die auf der Gesamtskala Mathematik in einer Leistungsgruppe mit Deutschland liegen, ist der Leistungsunterschied zwischen stärkstem und schwächstem Bereich in sechs Staaten geringer. So liegt für Dänemark eine Differenz von 15, für Australien von 13, für Italien von 13, für Bulgarien von 9, für Portugal von 9 und für Ungarn eine Differenz von lediglich 3 Punkten vor. Für drei Teilnehmerstaaten sind größere Differenzen festzustellen: Türkei (22), Schweden (21) und Polen (18).

Ferner zeigt sich, dass zwölf Staaten signifikant höhere Leistungswerte in allen drei kognitiven Anforderungsbereichen erzielen. Hingegen sind für fünf Teilnehmerstaaten in allen Bereichen signifikant niedrigere Leistungswerte festzustellen.

Eine separate Analyse der einzelnen kognitiven Anforderungsbereiche zeigt überdies, dass beim *Reproduzieren* für 17, beim *Anwenden* für 24 und beim *Problemlösen* für 13 Staaten signifikant bessere Testleistungen als für Deutschland festzustellen sind. Hingegen sind signifikant schwächere Testleistungen bei 13 Teilnehmerstaaten im Bereich *Reproduzieren*, bei 5 in *Anwenden* und 15 in *Problemlösen* zu verzeichnen.

Im Vergleich zu den jeweiligen Leistungsmittelwerten in den kognitiven Anforderungsbereichen der VG_{EU} und VG_{OECD} zeigt sich, dass im *Reproduzieren* und *Problemlösen* keine signifikanten Unterschiede festzustellen sind. Hingegen sind die Leistungswerte im *Anwenden* für Deutschland signifikant niedriger als in beiden Vergleichsgruppen.

Vergleich der Leistungswerte in den kognitiven Anforderungsbereichen 2007, 2011, 2015 und 2019: In der Tabelle 3.17 und der Abbildung 3.11 werden die Testergebnisse aus den vier Zyklen in den Bereichen *Reproduzieren*, *Anwenden* und *Problemlösen* berichtet. Für die Auswahl der Teilnehmerstaaten und die Gestaltung der Tabelle und der Abbildung werden die für die Tabelle 3.16 und die Abbildung 3.9 beschriebenen Kriterien herangezogen.

Abbildung 3.11 ist zu entnehmen, dass in Deutschland im Jahr 2019 beim *Reproduzieren* signifikant bessere Testergebnisse erzielt wurden als im Jahr 2007, während im Vergleich zu den Studien in 2011 und 2015 keine nennenswerten Unterschiede zu konstatieren sind. Im Bereich *Anwenden* wurden in 2019 signifikant schlechtere Ergebnisse erzielt als in den Jahren 2007 und 2011; im Vergleich zu 2015 ist kein bedeutsamer Unterschied festzustellen. Für den Bereich *Problemlösen* sind im Vergleich der Resultate von 2019 keine statistisch signifikanten Unterschiede zu den vorangegangenen TIMS-Studien auszumachen.

Im internationalen Vergleich der Daten von 2019 und 2007 zeigt sich folgendes Bild: 14 Teilnehmerstaaten schneiden in 2019 signifikant besser und zwei signifikant schlechter in *Reproduzieren* ab. Für *Anwenden* gilt dies für 14 beziehungsweise einen Teilnehmerstaat und für *Problemlösen* für 13 beziehungsweise keinen Staat. Zehn Teilnehmerstaaten schneiden in 2019 in allen drei kognitiven Anforderungsbereichen signifikant besser ab als noch in 2007; drei Teilnehmerstaaten in zwei Bereichen und fünf Teilnehmerstaaten – so auch Deutschland – in einem Bereich. Überdies sind für lediglich drei Staaten in 2019 in einem Bereich signifikant niedrigere Werte als in 2007 festzustellen. Dies trifft ebenfalls auf Deutschland zu.

Der Vergleich der Daten der Studien von 2019 und 2011 zeigt, dass zehn Teilnehmerstaaten in 2019 signifikant besser und drei signifikant schlechter in *Reproduzieren* abschneiden. Für *Anwenden* gilt dies für elf beziehungsweise zwei Teilnehmerstaaten und für *Problemlösen* für zehn Staaten beziehungsweise einen Staat. Sieben Staaten schneiden in allen drei kognitiven Anforderungsbereichen

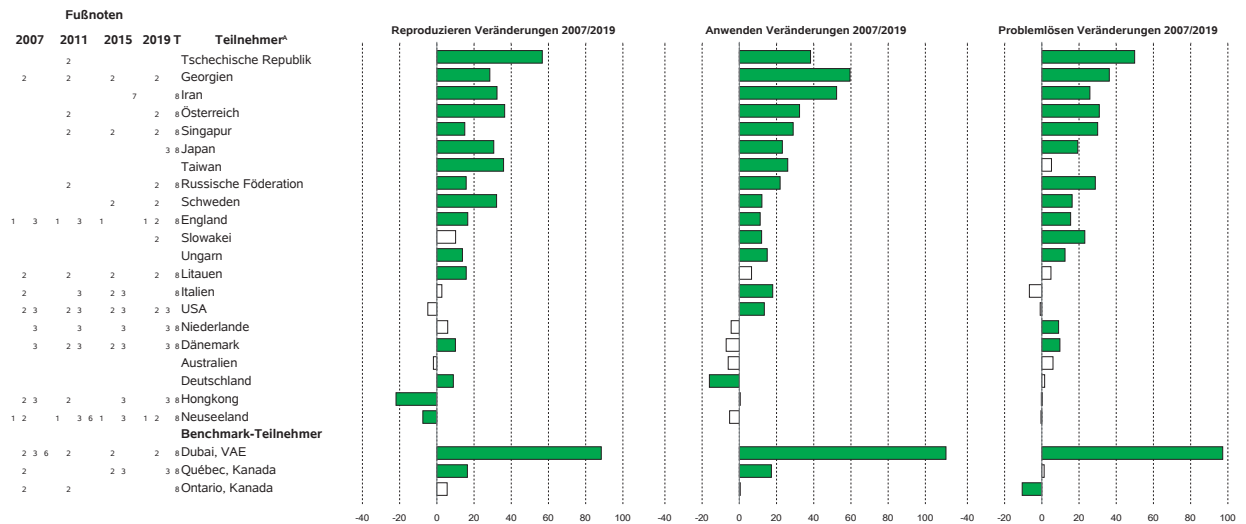
Tabelle 3.17: Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 in den kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* I

Fußnoten			Reproduzieren												Anwenden				Problemlösen											
2007	2011	2015	2019 T	Teilnehmer ^a			2007 ^B	2011 ^B	2015	2019	2007 ^B	2011 ^B	2015	2019	2007 ^B	2011 ^B	2015	2019												
							M_{σ^2}	(SE)	M_{11}	(SE)	M_{15}	(SE)	M_{19}	(SE)	M_{σ^2}	(SE)	M_{11}	(SE)	M_{15}	(SE)	M_{19}	(SE)								
2	2	2	2		Tschechische Republik		472	(2.5)	502	(2.4)	519	(2.5)	528	(3.0)	493	(2.9)	512	(2.8)	528	(2.4)	531	(2.6)	491	(3.7)	523	(2.7)	544	(3.0)	541	(2.8)
							445	(4.2)	449	(3.7)	466	(4.0)	473	(3.9)	430	(4.7)	447	(3.4)	461	(4.1)	490	(3.6)	433	(4.6)	450	(3.5)	452	(4.4)	469	(4.5)
2	2	2	2	7	Iran		404	(3.8)	435	(3.8)	429	(3.2)	436	(3.9)	397	(3.9)	427	(3.6)	435	(2.9)	450	(4.0)	401	(4.3)	423	(3.0)	426	(3.3)	426	(4.3)
							504	(2.2)	507	(2.7)		540	(2.0)		505	(2.0)	506	(2.8)		538	(2.1)		506	(2.4)	513	(3.3)		537	(2.4)	
2	2	2	2	3	Japan	Taiwan	625	(4.3)	629	(3.5)	631	(4.0)	640	(3.9)	597	(4.1)	602	(3.4)	619	(4.0)	626	(3.9)	584	(4.1)	588	(3.7)	603	(4.5)	614	(4.0)
							567	(2.4)	590	(1.7)	601	(2.4)	597	(2.0)	570	(2.2)	579	(1.6)	589	(2.1)	593	(2.0)	569	(2.2)	592	(2.0)	595	(2.7)	589	(2.2)
1	3	1	3	1	Russische Föderation	Schweden	586	(1.9)	599	(2.1)	620	(2.3)	622	(1.9)	574	(1.9)	593	(2.0)	593	(2.1)	600	(1.5)	571	(2.0)	577	(2.5)	576	(3.1)	576	(1.8)
							539	(4.9)	541	(3.4)	556	(3.4)	555	(3.0)	549	(5.3)	539	(3.9)	566	(3.7)	571	(3.6)	544	(5.3)	548	(3.6)	570	(4.0)	573	(3.6)
2	2	2	2	2	England	Slowakei	483	(2.6)	489	(2.2)	501	(3.4)	515	(3.1)	506	(2.3)	507	(2.2)	521	(2.7)	518	(2.8)	519	(2.8)	520	(3.0)	542	(3.3)	536	(2.9)
							546	(3.7)	552	(4.3)	554	(3.3)	563	(3.3)	542	(3.3)	542	(3.7)	544	(3.2)	553	(3.3)	539	(3.4)	531	(3.7)	540	(3.2)	554	(3.4)
2	2	2	2	2	Litauen	Ungarn	491	(4.3)	506	(3.8)	491	(2.4)	502	(3.3)	496	(4.4)	505	(4.0)	497	(2.5)	508	(3.4)	499	(4.5)	511	(3.9)	515	(2.9)	522	(3.5)
							511	(3.6)	519	(3.8)	532	(3.1)	525	(2.6)	506	(3.8)	513	(3.3)	526	(3.3)	521	(2.8)	510	(4.2)	514	(3.7)	529	(3.6)	522	(3.0)
2	3	2	3	3	Italien	USA	520	(2.8)	525	(3.0)	534	(2.8)	535	(2.8)	540	(2.7)	540	(2.5)	538	(2.9)	547	(2.7)	529	(2.8)	536	(2.5)	536	(3.2)	534	(3.3)
							512	(3.5)	510	(2.7)	511	(2.9)	515	(3.0)	499	(3.1)	506	(2.8)	504	(2.5)	517	(2.6)	511	(3.3)	505	(3.4)	503	(3.3)	504	(2.9)
3	3	3	3	3	Niederlande	Dänemark	541	(2.8)	556	(2.1)	547	(2.3)	536	(2.6)	524	(2.8)	539	(2.1)	537	(2.4)	537	(2.6)	525	(2.4)	525	(2.2)	531	(2.5)	524	(2.5)
							528	(2.4)	537	(2.0)	521	(1.8)	534	(2.1)	540	(2.2)	540	(1.6)	531	(1.7)	536	(2.2)	537	(2.5)	543	(2.6)	543	(2.6)	546	(2.9)
3	2	3	2	3	Australien	Deutschland	527	(2.8)	539	(2.9)	538	(2.8)	520	(2.3)	525	(2.5)	543	(2.7)	548	(3.2)	535	(2.2)	525	(2.4)	513	(2.6)	523	(3.0)	522	(3.0)
							511	(4.4)	516	(3.5)	509	(3.5)	509	(3.3)	522	(3.8)	519	(3.0)	521	(3.0)	516	(2.9)	516	(3.7)	513	(2.6)	523	(3.0)	522	(3.0)
2	3	3	3	3	Hongkong	Neuseeland	515	(2.1)	524	(2.3)	524	(2.3)	523	(2.3)	530	(2.4)	528	(2.3)	515	(2.2)	514	(2.5)	530	(2.9)	532	(3.0)	535	(2.4)	531	(2.8)
							622	(3.7)	619	(3.2)	618	(3.1)	600	(3.0)	606	(3.8)	597	(3.2)	621	(3.1)	606	(3.3)	596	(3.8)	589	(3.4)	600	(3.2)	596	(4.2)
1	2	1	3	6	1	3	484	(2.7)	476	(3.2)	475	(2.6)	476	(2.7)	493	(2.6)	490	(2.4)	497	(2.5)	487	(2.4)	502	(2.8)	490	(2.5)	504	(2.7)	501	(2.7)
							Benchmark-Teilnehmer																							
2	3	6	2	2	Dubai, VAE		454	(2.4)	472	(2.4)	514	(2.0)	542	(1.8)	436	(1.8)	465	(2.3)	510	(1.8)	547	(1.6)	441	(2.9)	464	(2.2)	507	(1.7)	538	(1.8)
							519	(3.1)	536	(2.6)	542	(4.3)	535	(2.7)	516	(2.9)	529	(2.6)	533	(4.1)	533	(2.3)	523	(3.2)	534	(2.5)	536	(4.9)	524	(2.8)
2	2	2	3	3	Québec, Kanada		498	(3.4)	510	(3.5)	505	(2.5)	504	(3.7)	513	(3.3)	521	(3.5)	513	(2.3)	514	(3.4)	526	(2.9)	522	(3.1)	524	(2.6)	516	(3.5)
							2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
2 = Der Auschlussgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
3 = Die Testinhaltequoten auf Schul- und/oder Schullebene erreichen nicht die internationalen Vorgaben.
4 = Abweichender Testzeitpunkt
5 = Abweichende Testinhalte
6 = Abweichende Testzeitpunkte
7 = Teilnahme an TIMSS 2015 und TIMSS Numeracy
8 = Eingeschränkte Vergleichbarkeit aufgrund variierender Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019
A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Schulzyklen 2007, 2011, 2015 und 2019 nicht berichtet.
B = Abweichungen in den berichteten Standardfehlern zur internationalen Berichterstattung sind in einem anderen Berechnungsverfahren begründet.

Abbildung 3.11: Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 in den kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* II

Fußnoten					Teilnehmer ^a	Reproduzieren Veränderungen ^b			Anwenden Veränderungen ^b			Problemlösen Veränderungen ^b			
2007	2011	2015	2019	T		$M_{2007}-M_{2019}$ (SE)	$M_{2011}-M_{2019}$ (SE)	$M_{2015}-M_{2019}$ (SE)	$M_{2007}-M_{2019}$ (SE)	$M_{2011}-M_{2019}$ (SE)	$M_{2015}-M_{2019}$ (SE)	$M_{2007}-M_{2019}$ (SE)	$M_{2011}-M_{2019}$ (SE)	$M_{2015}-M_{2019}$ (SE)	
2					2 Tschechische Republik	57	(3.9) +	26 (3.8) +	9 (3.9) +	38 (3.9) +	20 (3.8) +	3 (3.5)	50 (4.6) +	19 (3.9) +	-2 (4.1)
					2 Georgien	28	(5.7) +	24 (5.4) +	8 (5.5)	59 (5.9) +	42 (5.0) +	29 (5.5) +	36 (6.4) +	19 (5.7) +	17 (6.2) +
				7	Iran	32	(5.5) +	1 (5.5)	7 (5.1)	52 (5.6) +	22 (5.4) +	15 (5.0) +	26 (6.1) +	4 (5.3)	0 (5.5)
					2 Österreich	36	(3.0) +	33 (3.3) +		32 (2.9) +	32 (3.5) +		31 (3.4) +	24 (4.1) +	
				2	2 Singapur	15	(5.8) +	11 (5.3) +	10 (5.6)	29 (5.7) +	24 (5.2) +	6 (5.6)	30 (5.7) +	26 (5.4) +	11 (6.0)
				3	Japan	31	(3.1) +	7 (2.6) +	-4 (3.1)	23 (3.0) +	14 (2.6) +	4 (2.9)	19 (3.2) +	-3 (3.0)	-6 (3.5)
					Taiwan	36	(2.7) +	23 (2.8) +	1 (3.0)	26 (2.4) +	7 (2.5) +	7 (2.6) +	5 (2.7)	-1 (3.1)	1 (3.6)
				2	2 Russische Föderation	16	(5.7) +	14 (4.5) +	-2 (4.5)	22 (6.4) +	31 (5.3) +	4 (5.1)	29 (6.4) +	25 (5.1) +	3 (5.4)
				2	2 Schweden	32	(4.1) +	26 (3.8) +	15 (4.6) +	12 (3.6) +	10 (3.6) +	-3 (3.9)	16 (4.0) +	16 (4.2) +	-6 (4.4)
1	3	1	3	1	1 England	17	(4.9) +	11 (5.4) +	9 (4.6)	11 (4.7) +	11 (4.9) +	9 (4.6)	15 (4.8) +	23 (5.0) +	14 (4.7) +
					2 Slowakei	10	(5.4)	-5 (5.0)	11 (4.1) +	12 (5.5) +	3 (5.2)	11 (4.2) +	23 (5.7) +	11 (5.2) +	6 (4.5)
					Ungarn	14	(4.4) +	5 (4.6)	-8 (4.1)	15 (4.7) +	8 (4.4)	-5 (4.4)	12 (5.2) +	8 (4.8)	-7 (4.7)
2	2	2	2	2	2 Litauen	16	(4.0) +	10 (4.1) +	2 (4.0)	7 (3.8)	7 (3.7)	9 (4.0) +	5 (4.3)	-3 (4.1)	-2 (4.6)
2	3	2	3	2	3 Italien	3	(4.6)	5 (4.1)	4 (4.2)	18 (4.1) +	11 (3.9) +	13 (3.6) +	-7 (4.4)	-1 (4.5)	1 (4.3)
2	3	2	3	2	3 USA	-5	(3.8)	-19 (3.4) -	-11 (3.5) -	13 (3.9) +	-1 (3.4)	0 (3.6)	-1 (3.5)	-2 (3.3)	-7 (3.5)
	3	3	3		3 Niederlande	6	(3.1)	-3 (2.9)	14 (2.7) +	-4 (3.1)	-4 (2.8)	5 (2.8)	9 (3.9) +	2 (3.9)	3 (4.0)
3	2	3	2	3	3 Dänemark	10	(3.6) +	-8 (3.4) -	-12 (3.9) -	-7 (3.6)	-19 (3.8) -	-18 (3.6) -	10 (3.1) +	-8 (3.5) -	-13 (3.9) -
					Australien	-2	(5.5)	-7 (4.8)	0 (4.8)	-6 (4.8)	-3 (4.1)	-5 (4.2)	6 (4.8)	9 (4.0) +	-1 (4.2)
					Deutschland	9	(3.1) +	0 (3.2)	0 (3.2)	-16 (3.4) -	-14 (3.4) -	-1 (3.3)	2 (4.0)	-1 (4.1)	-4 (3.6)
2	3	2	3	3	3 Hongkong	-22	(4.8) -	-19 (4.4) -	-18 (4.4) -	0 (5.0)	9 (4.7)	-14 (4.6) -	0 (5.7)	7 (5.4)	-4 (5.3)
1	2	1	3	6	1 Neuseeland	-8	(3.8) -	0 (4.1)	1 (3.8)	-5 (3.5)	-2 (3.4)	-10 (3.5) -	0 (3.9)	11 (3.6) +	-2 (3.8)
Benchmark-Teilnehmer															
2	3	6	2	2	2 Dubai, VAE	88	(3.0) +	71 (3.0) +	29 (2.6) +	111 (2.5) +	82 (2.8) +	37 (2.4) +	97 (3.4) +	74 (2.9) +	31 (2.5) +
2				2	3 Québec, Kanada	16	(4.1) +	-1 (3.7)	-7 (5.1)	17 (3.7) +	4 (3.5)	1 (4.7)	1 (4.2)	-10 (3.8) -	-13 (5.7) -
2	2				2 Ontario, Kanada	6	(5.0)	-6 (5.1)	-1 (4.5)	0 (4.7)	-7 (4.9)	1 (4.1)	-11 (4.6) -	-6 (4.7)	-9 (4.4) -



□ Keine statistisch signifikanten Veränderungen zwischen 2007 und 2019 ($p > .05$)

■ Statistisch signifikante Veränderungen zwischen 2007 und 2019 ($p \leq .05$)

+ = Mittelwert in 2019 statistisch signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

- = Mittelwert in 2019 statistisch signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

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

6 = Abweichender Testzeitpunkt

7 = Teilnahme an TIMSS 2015 und TIMSS Numeracy

8 = Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

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

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

in 2019 besser als in 2011 ab. Vier Teilnehmerstaaten schneiden in mindestens einem Bereich signifikant schlechter ab. Dazu zählt auch Deutschland.

Abschließend werden die Daten der Studien von 2019 und 2015 verglichen. Es zeigt sich, dass vier Teilnehmerstaaten in 2019 signifikant höhere und drei signifikant niedrigere Leistungswerte in *Reproduzieren* erzielten. Für *Anwenden* gilt dies für sechs beziehungsweise drei Staaten und für *Problemlösen* für zwei Staaten beziehungsweise einen Staat. Zehn Teilnehmerstaaten schneiden in mindestens einem kognitiven Anforderungsbereich in 2019 besser als in 2015 ab. Auf der anderen Seite sind für vier Staaten in mindestens einem Bereich signifikant schwächere Leistungen festzustellen. Für Deutschland unterscheiden sich die Leistungen in keinem Bereich nennenswert.

3.5.5 Einstellungen und Selbstkonzept

Zusätzlich zu den TIMSS-Leistungstests wurden im Rahmen der Hintergrundbefragung Informationen zu den Einstellungen zur Mathematik und zum Selbstkonzept der Schülerinnen und Schüler erhoben. In diesem Abschnitt werden zunächst die TIMSS 2019-Ergebnisse für Deutschland dargestellt und anschließend wird ein Vergleich zu den Studienzyklen von 2007, 2011 und 2015 angestellt. Auf die Darstellung des internationalen Vergleichs der Ergebnisse wird – wie bereits in den bisherigen TIMSS-Berichtsbänden – verzichtet, um mögliche Referenzgruppeneffekte zu vermeiden. Darunter versteht man Auswirkungen, die durch den Vergleich eigener Fähigkeiten mit denen einer Bezugsgruppe entstehen können. Für schulische Kontexte bedeutet dies, dass die Einschätzung der eigenen Leistung davon abhängig sein kann, ob eine leistungsstarke beziehungsweise leistungsschwache Referenzgruppe vorliegt. Hieraus können sich Effekte auf die fachbezogene Einstellung und auf die fachbezogenen Selbstkonzepte der Kinder ergeben. Da die TIMS-Studien durch internationale Vergleiche von Leistungsdaten geprägt sind, wären die Ergebnisse von Referenzgruppeneffekten betroffen. Daher werden im Folgenden lediglich die nationalen Befunde berichtet.

Einstellung zur Mathematik: Im Schülerfragebogen geben die teilnehmenden Viertklässlerinnen und Viertklässler an, ob sie den folgenden Aussagen völlig zustimmen, eher zustimmen, eher nicht zustimmen oder überhaupt nicht zustimmen.

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

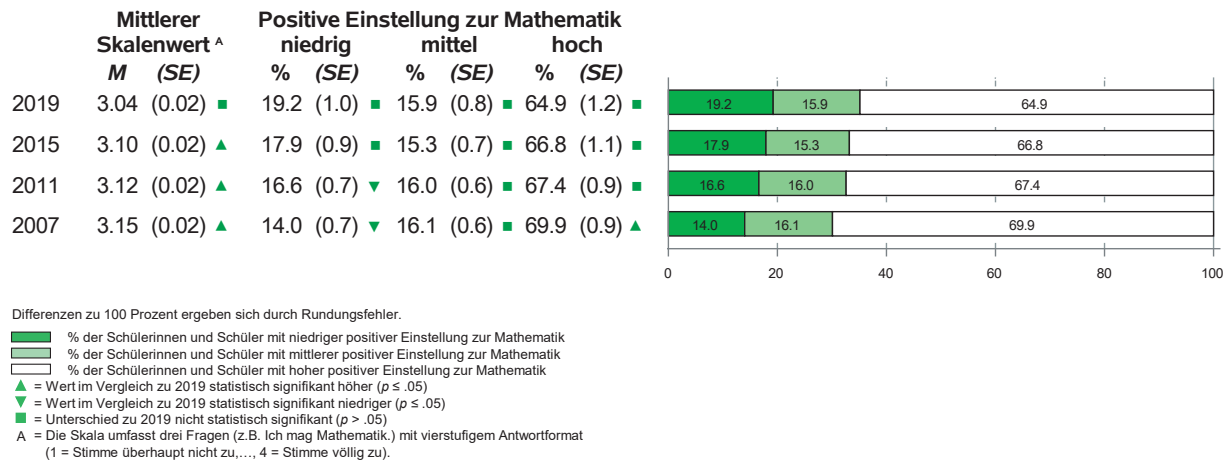
Für die Darstellung der Ergebnisse werden die Angaben der Schülerinnen und Schüler bei Aussage 1 und 3 mit 1 (*stimme überhaupt nicht zu*) bis 4 (*stimme völlig zu*) Punkten kodiert, bei Aussage 2 hingegen in umgekehrter Reihenfolge. Daraufhin wird der Mittelwert dieser drei Aussagen für jedes Kind berechnet. Der daraus resultierende Skalenwert gibt an, inwieweit das jeweilige Kind eine positive oder negative Einstellung zur Mathematik vertritt. Die interne Konsistenz der Skala liegt bei einem Cronbachs Alpha von $\alpha = .86$.

Die Gesamtheit der Schülerinnen und Schüler wird überdies in drei Gruppen eingeteilt. Schülerinnen und Schüler, deren Antworten auf obige Aussagen einen Mittelwert (M) von $1 \leq M < 2$ aufweisen, werden als *niedrig* bezüglich ihrer positiven Einstellung zur Mathematik klassifiziert. Hingegen werden Mittelwerte von $2 \leq M < 3$ als *mittel* und $3 \leq M \leq 4$ als *hoch* eingestuft.

Abbildung 3.12 zeigt die Einstellung der Schülerinnen und Schüler aus Deutschland zum Fach Mathematik in den drei bisherigen TIMSS-Zyklen in 2007, 2011 und 2015 sowie im aktuellen Studienzyklus TIMSS 2019. Die Abbildung zeigt auch die Anteile derjenigen Kinder, die eine *niedrige*, *mittlere* und *hohe* positive Einstellung zum Fach Mathematik aufweisen.

In TIMSS 2019 liegt der mittlere Skalenwert bei 3.04. Dieser Wert ist signifikant niedriger als die mittleren Skalenwerten aus TIMSS 2007 (3.15), TIMSS 2011 (3.12) sowie TIMSS 2015 (3.10). Somit weisen die Schülerinnen und Schüler aus Deutschland in 2019 eine negativere Einstellung zum Fach Mathematik auf als in allen bisherigen Studienzyklen. Gleichwohl bildet der mittlere Skalenwert eine positiv einzuschätzende Einstellung ab. Die Zustimmung der

Abbildung 3.12: Mittlere positive Einstellung von Schülerinnen und Schülern zur Mathematik sowie prozentuale Verteilungen nach niedriger, mittlerer und hoher Einstellung – TIMSS 2007, 2011, 2015 und 2019 im Vergleich



IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

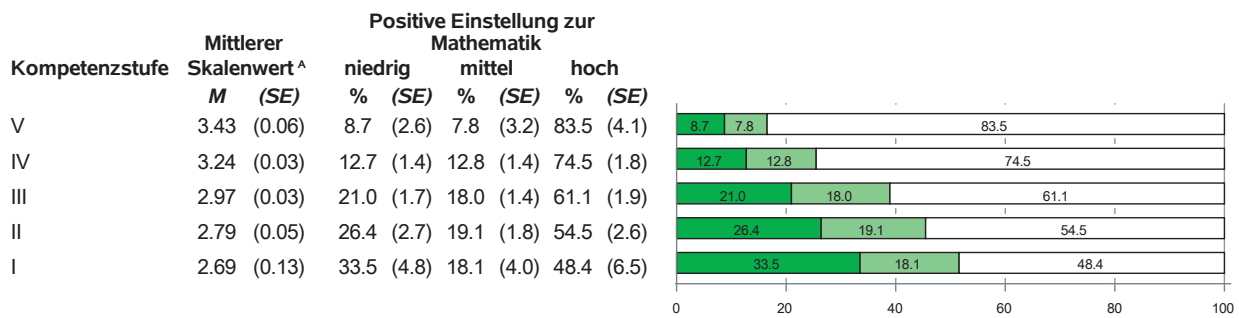
Schülerinnen und Schüler in Deutschland zu den oben angegebenen Aussagen ist somit im Mittel nach wie vor hoch.

Neben den mittleren Skalenwerten zeigt Abbildung 3.12, wie viel Prozent der Lernenden eine *niedrige*, *mittlere* oder eine *hohe* positive Einstellung zum Fach Mathematik zeigen. 19 Prozent der Viertklässlerinnen und Viertklässler weisen in TIMSS 2019 eine *niedrige* positive Einstellung auf. Während sich dieser Wert im Vergleich zu 2015 (18%) nicht nennenswert unterscheidet, liegen im Vergleich zu den Studienzyklen von 2011 (17%) und 2007 (14%) signifikante Unterschiede vor. Im Vergleich zu diesen Studienzyklen ist der Anteil der Schülerinnen und Schüler mit *niedriger* positiver Einstellung zur Mathematik gestiegen.

Der Anteil an Schülerinnen und Schülern mit *hoher* positiver Einstellung beträgt in TIMSS 2019 65 Prozent. Im Vergleich zu TIMSS 2011 sowie TIMSS 2015 (jeweils 67%) liegen nur marginale Unterschiede vor; im Vergleich zu TIMSS 2007 (70%) ist hingegen ein signifikanter Unterschied zu konstatieren. Der Anteil an Schülerinnen und Schülern mit *hoher* positiver Einstellung ist 2019 niedriger als noch in 2007. Hinsichtlich der Schülerinnen und Schüler mit *mittlerer* positiver Einstellung liegt für TIMSS 2019 ein Wert von 16 Prozent vor, der sich nicht signifikant von den Werten der bisherigen Zyklen unterscheidet.

Nachdem die *mittlere* positive Einstellung von Viertklässlerinnen und Viertklässlern aus Deutschland im Trend analysiert wurden, zeigt die nachfolgende Abbildung 3.13 die prozentuale Verteilung der Testleistungen aus TIMSS 2019 auf den Kompetenzstufen nach positiver Einstellung. Damit ist das Ziel verbunden, potenzielle Zusammenhänge zwischen Einstellung zur Mathematik und mathematischen Kompetenzen zu erfassen. Die Daten zeigen einen positiven Zusammenhang zwischen mathematischer Kompetenz der Kinder und positiver Einstellung zur Mathematik. Die Korrelation zwischen den Leistungswerten und den Werten der Skala *Einstellung zur Mathematik* beträgt für Deutschland $r = .23$ ($p \leq .01$).

Abbildung 3.13 bildet für die fünf Kompetenzstufen den jeweiligen Anteil der Schülerinnen und Schüler mit *niedriger*, *mittlerer* beziehungsweise *hoher* positiver Einstellung ab und illustriert diesen Zusammenhang. Es zeigt sich, dass der Anteil an Schülerinnen und Schülern mit *hoher* positiver Einstellung auf höheren Kompetenzstufen deutlich größer ist (KS V: 84%, KS IV: 75%,

Abbildung 3.13: Prozentuale Verteilung von Schülerinnen und Schülern auf den fünf Kompetenzstufen nach Einstellung zur Mathematik

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

■ % der Schülerinnen und Schüler mit niedriger positiver Einstellung zur Mathematik
 ■ % der Schülerinnen und Schüler mit mittlerer positiver Einstellung zur Mathematik
 ■ % der Schülerinnen und Schüler mit hoher positiver Einstellung zur Mathematik
 A = Die Skala umfasst drei Fragen (z.B. Ich mag Mathematik.) mit vierstufigem Antwortformat
 (1 = Stimme überhaupt nicht zu, ..., 4 = Stimme völlig zu).

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

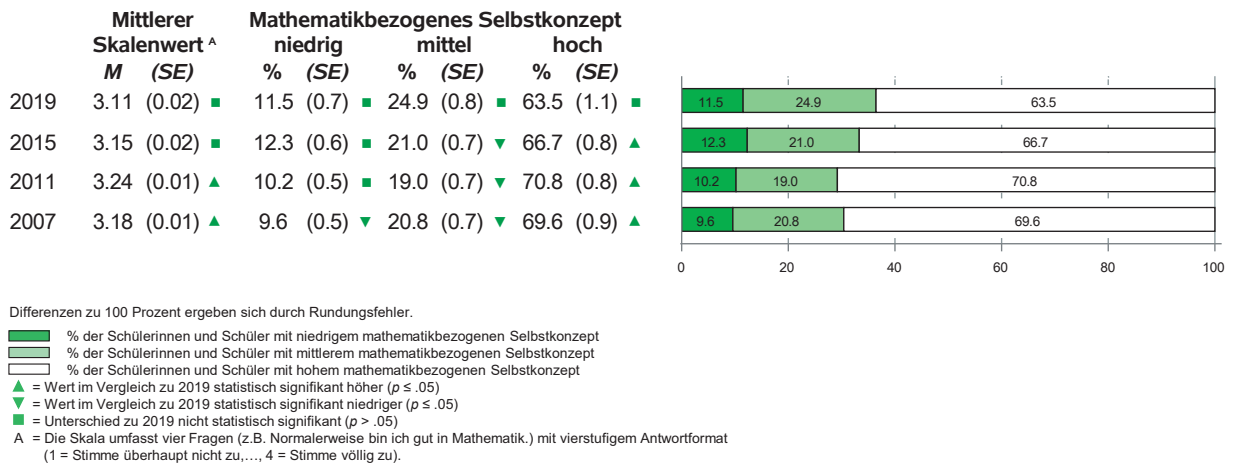
KS III: 61 %, KS II: 55 %, KS I: 48 %). Auf der anderen Seite ist der Anteil an Schülerinnen und Schülern mit *niedriger* positiver Einstellung zur Mathematik auf den niedrigen Kompetenzstufen höher (KS V: 9 %, KS IV: 13 %, KS III: 21 %, KS II: 26 %, KS I: 34 %). In diesem Zusammenhang ist zu erwähnen, dass etwa 10 Prozent der Schülerinnen und Schüler, die den Kompetenzstufen IV und V zuzuordnen sind, eine *niedrige* positive Einstellung zur Mathematik vertreten. Auf der anderen Seite artikuliert etwa die Hälfte der Lernenden, deren mathematische Kompetenzen den Kompetenzstufen I und II entsprechen, eine *hohe* positive Einstellung zur Mathematik. Dies ist insofern erwähnenswert, als dass Schülerinnen und Schüler mit vergleichsweise ausbaufähigen mathematischen Kompetenzen eine *hohe* positive Einstellung zum Fach Mathematik haben können.

Mathematikbezogenes Selbstkonzept: Neben der Einstellung der Schülerinnen und Schüler zur Mathematik wird in TIMSS auch das mathematikbezogene Selbstkonzept im Schülerfragebogen erfasst. Analog zum beschriebenen Vorgehen bei der Untersuchung der Einstellung wurden den Kindern Aussagen vorgelegt, und es wurde gefragt, ob sie diesen völlig zustimmen, eher zustimmen, eher nicht zustimmen oder überhaupt nicht zustimmen. Folgende Aussagen wurden herangezogen:

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

Bei den Aussagen 1 und 4 werden die Antworten der Schülerinnen und Schüler mit 1 (*stimme überhaupt nicht zu*) bis 4 (*stimme völlig zu*) kodiert. Die Kodierung der Angaben zu den gegenteilig formulierten Aussagen 2 und 3 erfolgt hingegen invertiert. Aus den Angaben der Kinder zu den vier obigen Aussagen wird anschließend der Mittelwert zum mathematikbezogenen Selbstkonzept gebildet, woraufhin die Einordnung in eine von drei Gruppen erfolgt. Bei einem Mittelwert von $1 \leq M < 2$ wird ein *niedriges*, bei $2 \leq M < 3$ ein *mittleres* und bei $3 \leq M \leq 4$ ein *hohes* mathematikbezogenes Selbstkonzept angenommen. Die interne Konsistenz der Skala liegt bei einem Cronbachs Alpha von $\alpha = .82$.

Abbildung 3.14: Mittleres mathematikbezogenes Selbstkonzept von Schülerinnen und Schülern sowie prozentuale Verteilungen nach niedrigem, mittlerem und hohem Selbstkonzept – TIMSS 2007, 2011, 2015 und 2019 im Vergleich



IEA: Trends in International Mathematics and Science Study

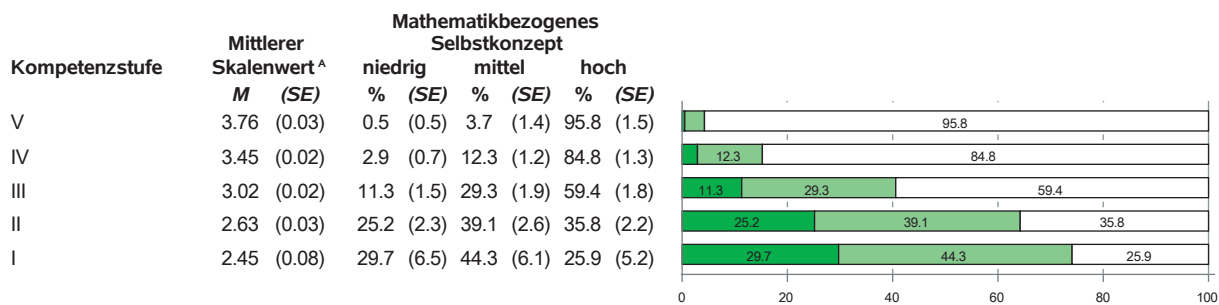
© TIMSS 2019

Abbildung 3.14 zeigt anhand der mittleren Skalenwerte, wie das mathematikbezogene Selbstkonzept der Viertklässlerinnen und Viertklässler aus Deutschland in den vier TIMSS-Erhebungszyklen ausfällt. Außerdem ist die Verteilung der Schülerinnen und Schüler auf die Gruppen derjenigen Kinder mit *niedrigem*, *mittlerem* und *hohem* mathematikbezogenen Selbstkonzept dargestellt.

Der mittlere Skalenwert in TIMSS 2019 liegt bei 3.11. Dieser Wert ist im Vergleich zu den Werten aus TIMSS 2011 (3.24) und TIMSS 2007 (3.18) signifikant niedriger. Hingegen sind keine bedeutsamen Unterschiede im Vergleich zu TIMSS 2015 (3.15) zu verzeichnen.

Der Vergleich der Ergebnisse der vier Zyklen zeigt zudem, dass der Anteil der Kinder mit *hohem* mathematikbezogenen Selbstkonzept bei TIMSS 2019 bei 64 Prozent liegt und damit signifikant niedriger ist als bei allen bisherigen Studienzyklen. In TIMSS 2007 war für 70 Prozent ein *hohes* Selbstkonzept festzustellen, in 2011 für 71 Prozent und in 2015 für 67 Prozent. Darüber hinaus ist der Anteil der Schülerinnen und Schüler mit *mittlerem* mathematikbezogenen Selbstkonzept im Vergleich zu allen bisherigen Studienzyklen signifikant größer. In TIMSS 2019 trifft dies auf 25 Prozent zu, während es in 2007 mit 21 Prozent, 2011 mit 19 Prozent und 2015 mit 21 Prozent anteilig jeweils weniger Kinder waren. Hinsichtlich der Gruppe derjenigen Kinder mit *niedrigem* mathematischen Selbstkonzept lassen sich zwischen den Zyklen von 2019 (12%), 2015 (12%) und 2011 (10%) keine bedeutenden Unterschiede feststellen. Lediglich in 2007 (9.6%) sind im Vergleich zu 2019 anteilig signifikant weniger Kinder mit *niedrigem* mathematikbezogenen Selbstkonzept zu verzeichnen.

Wie auch bei der Einstellung zur Mathematik sollen nachfolgend potenzielle Zusammenhänge zwischen mathematikbezogenem Selbstkonzept und mathematischen Kompetenzen untersucht werden. Abbildung 3.15 zeigt hierzu die prozentuale Verteilung der Testleistungen aus TIMSS 2019 auf den Kompetenzstufen nach *niedrigem*, *mittlerem* und *hohem* mathematikbezogenen Selbstkonzept. Analog zu Abbildung 3.13 bildet Abbildung 3.15 für die fünf Kompetenzstufen den jeweiligen Anteil der Schülerinnen und Schüler mit *niedrigem*, *mittlerem* beziehungsweise *hohem* mathematikbezogenen Selbstkonzept ab. Die Daten verweisen auf einen positiven Zusammenhang zwischen *positivem* mathematik-

Abbildung 3.15: Prozentuale Verteilung von Schülerinnen und Schülern in Mathematik auf den fünf Kompetenzstufen nach mathematikbezogenem Selbstkonzept

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

■ % der Schülerinnen und Schüler mit niedrigem mathematikbezogenen Selbstkonzept

■ % der Schülerinnen und Schüler mit mittlerem mathematikbezogenen Selbstkonzept

■ % der Schülerinnen und Schüler mit hohem mathematikbezogenen Selbstkonzept

A = Die Skala umfasst vier Fragen (z.B. Normalerweise bin ich gut in Mathematik.) mit vierstufigem Antwortformat (1 = Stimme überhaupt nicht zu, ..., 4 = Stimme völlig zu).

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

bezogenen Selbstkonzept und mathematischen Kompetenzen. Die Korrelation beträgt für Deutschland $r = .49$ ($p \leq .01$). Der Anteil an Schülerinnen und Schülern mit *hohem* mathematikbezogenen Selbstkonzept fällt von 96 Prozent auf Kompetenzstufe V auf 26 Prozent auf Kompetenzstufe I. Auf der anderen Seite nimmt der Anteil der Kinder mit *niedrigem* beziehungsweise *mittlerem* mathematikbezogenen Selbstkonzept von Kompetenzstufe V (1% bzw. 4%) zu Kompetenzstufe I (30% bzw. 44%) zu.

3.6 Zusammenfassung

Abschließend werden nun zentrale Befunde des Studienzyklus TIMSS 2019 zusammenfassend dargestellt und interpretiert.

Leistungsmittelwerte: Viertklässlerinnen und Viertklässler in Deutschland erzielen im internationalen Vergleich einen Leistungsmittelwert von 521 Punkten. Anders als in den Studienzyklen von 2007 und 2011 befinden sie sich damit nicht mehr im oberen Drittel der Rangreihe, sondern – wie im Jahr 2015 – nur noch im Mittelfeld der insgesamt 58 teilnehmenden Staaten. Dabei liegen die Testergebnisse für Deutschland zwar signifikant über dem internationalen Mittelwert von 501 Punkten. Im Vergleich zu den Mittelwerten aller teilnehmenden EU-Staaten (527) sowie aller teilnehmenden OECD-Staaten (529) schneidet Deutschland aber wie bei TIMSS 2015 signifikant schwächer ab. Deutschland liegt zusammen mit Portugal (525), Dänemark (525), Ungarn (523), der Türkei (5. Jahrgangsstufe, 523), Schweden (521), Polen (520), Australien (516), Aserbaidschan (515), Bulgarien (515) und Italien (515) in einer Staatengruppe, deren Testergebnisse sich bezüglich des Leistungsmittelwertes statistisch nicht signifikant voneinander unterscheiden.

Zu den asiatischen Teilnehmerstaaten am oberen Ende der Rangreihe besteht ein sehr großer Abstand. Die in TIMSS 2019 leistungsstärksten fünf Staaten sind Singapur (625), Hongkong (602), die Republik Korea (600), Taiwan (599) sowie Japan (593). Signifikant schwächere Testergebnisse haben sich für Deutschland auch im Vergleich zu einigen europäischen Teilnehmerstaaten ergeben. Hier sind unter anderem Nordirland (566), England (556), Irland (548), Lettland (546),

Österreich (539), die Niederlande (538) oder die Flämische Gemeinschaft in Belgien (532) zu nennen. Die Ergebnisse zeigen also, dass diverse EU-Staaten signifikant besser abschneiden als Deutschland.

Leistungsstreuung: In den vorangegangenen TIMS-Studien konnte für Deutschland eine im Vergleich zu vielen anderen Teilnehmerstaaten geringe Leistungsstreuung verzeichnet werden. Dieser Befund setzt sich tendenziell auch in TIMSS 2019 fort. Dafür spricht sowohl die Größe der Standardabweichung (70) als auch die Leistungsspanne zwischen den jeweils 5 Prozent schwächsten und 5 Prozent stärksten Viertklässlerinnen und Viertklässlern. Es gibt zwölf Staaten, die eine vergleichbare oder gar extremere ‚Leistungshomogenität‘ aufweisen, während in 45 Teilnehmerstaaten eine größere Standardabweichung zu beobachten ist.

Leistungsmittelwerte im Trendvergleich: Weil Deutschland im Jahr 2019 zum vierten Mal an TIMSS-Grundschule teilgenommen hat, können über einen Zeitraum von zwölf Jahren Tendaussagen zur mathematischen Kompetenz der Viertklässlerinnen und Viertklässler gemacht werden. Dabei zeigt sich, dass die mathematische Kompetenz der Lernenden in Deutschland in 2019 mit 521 Punkten keine bedeutenden Unterschiede zu den Untersuchungen in den Jahren 2007 (526 Punkte) und 2015 (522 Punkte) aufweist. Im Vergleich zu TIMSS 2011 hingegen (528 Punkte) ist ein signifikant schlechteres Ergebnis zu konstatieren.

In TIMSS 2007 waren Dänemark, Litauen und die USA diejenigen Staaten, deren Leistungsmittelwerte in Mathematik nur geringfügig von den Ergebnissen Deutschlands abwichen. In TIMSS 2019 liegen die Leistungsmittelwerte der USA und Litauens auf der Gesamtskala Mathematik signifikant oberhalb des deutschen Mittelwertes, während sich Dänemark nach wie vor in der ‚Leistungsgruppe Deutschland‘ befindet. Betrachtet man alle Teilnehmerstaaten, die an allen vier Untersuchungen teilnahmen (und Österreich), so ist festzuhalten, dass 14 der 21 Teilnehmerstaaten in 2019 im Vergleich zu 2007 signifikante Leistungszuwächse erreichen. Deutschland gelingt dies wie auch sechs weiteren Staaten nicht.

Im TIMSS-Zyklus von 2011 waren Litauen, Portugal und Irland in der ‚Leistungsgruppe Deutschland‘. Die Leistungsmittelwerte von Litauen und Irland sind in 2019 signifikant oberhalb des entsprechenden Wertes für Deutschland, während sich Portugal noch in der ‚Leistungsgruppe Deutschland‘ befindet. Eine Analyse aller Teilnehmerstaaten, die an allen vier Untersuchungen teilnahmen (und Österreich), zeigt, dass zwölf von 21 Teilnehmerstaaten in 2019 signifikant höhere Leistungsmittelwerte als in 2011 erzielen, in sieben Staaten sind keine signifikanten Unterschiede festzustellen. Für Deutschland und Dänemark sind signifikante Leistungsabnahmen zu konstatieren.

Im Vergleich der beiden letzten Studienzyklen 2019 und 2015 sind in sechs Staaten signifikante Zunahmen, in zwölf Staaten keine bedeutsamen Veränderungen und in zwei Staaten (Dänemark, Hongkong) signifikante Rückgänge zu verzeichnen. Das Beispiel Hongkong zeigt, dass es nicht nur um Tendenzen, sondern auch um Ausgangsniveaus geht. Gleichwohl ist zusammenfassend festzuhalten, dass es in vergleichsweise vielen Staaten positive Entwicklungen, zumindest im Vergleich zu den Zyklen 2007 und 2011, zu verzeichnen gibt, während die Leistungswerte in Deutschland stagnieren.

Extreme Leistungsgruppen: In Deutschland erreichten bei TIMSS 2019 die 5 Prozent leistungsschwächsten Schülerinnen und Schüler höchstens 405 Punkte. Im Vergleich der 20 Staaten, die an den letzten vier Zyklen teilnahmen (und Österreich), sind in sieben Staaten signifikant höhere und in sieben Staaten signifikant niedrigere Werte zu beobachten. Die 5 Prozent leistungsstärksten Lernenden erreichen in Deutschland mindestens 630 Punkte. Hier werden in 13 Staaten signifikant höhere (u. a. England, Dänemark oder Tschechische Republik) und in zwei Staaten signifikant niedrigere Werte erzielt (Georgien, Iran). Die Ergebnisse deuten darauf hin, dass in Deutschland die Förderung leistungsstarker Lernender noch weniger gut gelingt als die Förderung der leistungsschwächeren Lernenden.

Leistungsgruppen im Trendvergleich: Insgesamt zeigt sich im Vergleich der Entwicklungen der den Perzentilen 5, 25, 75 und 95 zugeordneten Punktwerte, dass für Deutschland im Unterschied zu 2015 keine und zu 2011 sowie 2007 jeweils signifikant niedrigere Werte im Perzentil 25 zu beobachten sind. Diese weitgehende Stagnation oder leichte Verschlechterung lässt sich im Trendvergleich nur bei wenigen Staaten wie Dänemark oder Australien beobachten. In den meisten Staaten sind hingegen statistisch positive Unterschiede zu verzeichnen. Im Vergleich von 2019 und 2007 zeigen 15 Teilnehmer in mindestens einem und acht Teilnehmende in allen Perzentilen signifikant höhere Punktwerte.

Verteilung auf Kompetenzstufen: Für eine inhaltliche Beschreibung der Mathematikleistung wurden die Leistungswerte der Lernenden jeweils einer von fünf Kompetenzstufen zugeordnet. In der Untersuchung TIMSS 2019 erreichten 96 Prozent der Schülerinnen und Schüler in Deutschland mindestens Stufe II, 75 Prozent mindestens Stufe III, 36 Prozent mindestens Stufe IV und 6 Prozent die höchste Stufe V.

Dementsprechend ist festzuhalten, dass etwa ein Viertel der Schülerinnen und Schüler am Ende ihrer Grundschulzeit nur die erste oder zweite Kompetenzstufe erreicht und damit allenfalls über einfache mathematische Fertigkeiten und Fähigkeiten verfügt. Zudem ist es bedenklich, dass 4 Prozent der Schülerinnen und Schüler nur die unterste Kompetenzstufe erreichen und lediglich über rudimentäres mathematisches Wissen verfügen. Fast alle an TIMSS 2019 teilnehmenden Staaten, die signifikant bessere Testleistungen auf der Gesamtskala Mathematik erzielen, haben in diesem Bereich anteilig weniger Schülerinnen und Schüler – die wenigen Ausnahmen waren die USA (7%) sowie Finnland und Zypern (jeweils 5%). Demnach gelingt es fast allen diesen Staaten, höhere Leistungen zu erreichen, ohne dass dies auf Kosten der leistungsschwachen Kinder geht. In der ‚Leistungsgruppe Deutschland‘ weist Deutschland allerdings den niedrigsten Wert auf. Auch die Vergleichswerte der VG_{EU} (6%) und VG_{OECD} (7%) liegen knapp über dem Wert in Deutschland.

Die 6 Prozent der Schülerinnen und Schüler auf der höchsten Kompetenzstufe V nutzen ihre mathematischen Fertigkeiten und Fähigkeiten verständlich beim Lösen verhältnismäßig komplexer Probleme und können ihre Begründungen erklären. Der Anteil dieser Viertklässlerinnen und Viertklässler ist jedoch nicht nur bei den Staaten am oberen Ende der Leistungsskala größer, sondern auch bei Staaten aus der ‚Leistungsgruppe Deutschland‘, wie beispielsweise der Türkei (5. Jahrgangsstufe, 15%) und Australien (10%). Die Vergleichswerte VG_{EU} und VG_{OECD} liegen mit 9 Prozent beziehungsweise 12 Prozent ebenfalls über dem Wert für Deutschland. Anteile auf der höchsten Kompetenzstufe von mehr als

35 Prozent konnten in Singapur (54%), Hongkong (38%), der Republik Korea (37%) und Taiwan (37%) festgestellt werden. Von Werten dieser Größenordnung ist Deutschland weit entfernt.

Angesichts dieser Verteilung der Schülerinnen und Schüler in Deutschland auf die Kompetenzstufen wird Handlungsbedarf deutlich. Für die 4 Prozent der Schülerinnen und Schüler auf der untersten Kompetenzstufe ist anschlussfähiges Lernen in der Sekundarstufe I extrem schwierig. Dass in Deutschland mit 6 Prozent selbst im europäischen Vergleich nur vergleichsweise wenige Lernende die höchste Kompetenzstufe erreichen, deutet darauf hin, dass eine Potenzialförderung durch kognitiv aktivierenden Mathematikunterricht in Deutschland noch nicht in ausreichendem Maße stattzufinden scheint.

Verteilung der Kompetenzstufen im Trendvergleich: Vergleicht man die Ergebnisse zur prozentualen Besetzung der Kompetenzstufen von 2019 mit denen der vorangegangenen Untersuchungszyklen, so lassen sich für die Kompetenzstufen IV und V keine statistisch bedeutenden Unterschiede erkennen. Jeweils zwischen 29 und 32 Prozent erreichen die Kompetenzstufe IV, jeweils 5 bis 6 Prozent erreichen die Stufe V. Insbesondere im Vergleich zu TIMSS 2007 und TIMSS 2011 gelingt es einer ganzen Reihe von Staaten, hier signifikante Steigerungen der Anteile der Lernenden in diesen beiden Kompetenzstufen aufzuweisen.

In Bezug auf die Kompetenzstufe III (durchschnittlich) ergibt sich für Deutschland im Vergleich der Ergebnisse von TIMSS 2019 und TIMSS 2007 keine, jedoch im Vergleich zu den anderen beiden vorangegangenen Untersuchungen eine jeweils signifikante Veränderung in Form eines Rückgangs von jeweils 43 auf 39 Prozent der Schülerinnen und Schüler, die sich auf dieser Kompetenzstufe befinden.

Mit Blick auf die Kompetenzstufen I und II schließlich lassen sich im Vergleich von TIMSS 2019 zu TIMSS 2015 keine statistisch bedeutsamen Unterschiede konstatieren. Im Vergleich zu TIMSS 2007 und TIMSS 2011 ist jedoch festzuhalten, dass es zu einem signifikanten Anstieg des Anteils der Lernenden in beiden (2011) beziehungsweise in der Kompetenzstufe II (2007) gekommen ist. Einer ganzen Reihe von Staaten gelingt hingegen eine signifikante Reduktion dieser Anteile. Während es in 2007 und 2011 20 beziehungsweise 21 Prozent der Lernenden waren, sind es in 2015 und 2019 jeweils 25 Prozent der Lernenden in Deutschland, deren Testergebnisse den Rückschluss zulassen, dass sie sich in den Kompetenzstufen I und II befinden.

Mathematische Inhaltsbereiche: In jedem der drei Inhaltsbereiche liegt der Leistungsmittelwert in Deutschland signifikant über dem internationalen Mittelwert. Die Viertklässlerinnen und Viertklässler in Deutschland erzielen im Bereich *Arithmetik* 516 Punkte, im Bereich *Messen und Geometrie* 530 Punkte und im Bereich *Daten* 515 Punkte. Die Mittelwerte in den Inhaltsbereichen *Arithmetik* und *Daten* fallen signifikant niedriger und im Inhaltsbereich *Messen und Geometrie* signifikant höher aus als der Gesamtmittelwert Deutschlands (521).

Mathematische Inhaltsbereiche im Trendvergleich: Im Jahr 2019 zeigen sich im Bereich *Arithmetik* signifikant schwächere Testleistungen als in 2007, was im Vergleich von 2019 mit den Erhebungen in 2011 beziehungsweise 2015 nicht der Fall ist. Im Bereich *Messen und Geometrie* sind im jeweiligen Vergleich zu den vorangegangenen TIMS-Studien keine statistisch bedeutsamen Veränderungen zu

verzeichnen. Hingegen sind im Bereich *Daten* im Vergleich der Resultate von 2019 mit den anderen drei Zeitpunkten signifikant schlechtere Punktwerte zu verzeichnen. Insgesamt zeigen die Viertklässlerinnen und Viertklässler damit eine relative Stärke im Bereich *Messen und Geometrie*, und diese Leistung ist stabil. Für die anderen beiden Bereiche zeigen sich aktuell relative Schwächen, die auch im Trendvergleich nicht ignoriert werden sollten.

Kognitive Anforderungsbereiche: In jedem der drei kognitiven Anforderungsbereiche liegt der Leistungsmittelwert in Deutschland signifikant oberhalb des internationalen Mittelwertes. Die Leistungswerte der Bereiche *Reproduzieren* (523) und *Problemlösen* (531) liegen dabei signifikant oberhalb des Gesamtmittelwertes Mathematik (521). Demgegenüber liegt der Leistungsmittelwert im Bereich *Anwenden* (514) signifikant unterhalb dieses Wertes. Daraus geht hervor, dass die Schülerinnen und Schüler in Deutschland ihre relativen Schwächen im Bereich *Anwenden* aufweisen und ihre relativen Stärken in den anderen beiden Bereichen.

Kognitive Anforderungen im Trendvergleich: Im Vergleich der Studienzyklen zeigt sich, dass 2019 beim *Reproduzieren* signifikant bessere Testergebnisse erzielt wurden als im Jahr 2007, während im Vergleich zu den anderen beiden Erhebungszeitpunkten keine nennenswerten Differenzen zu berichten sind. Im Bereich *Anwenden* ergibt der Vergleich mit 2015 keine signifikant unterschiedlichen, jedoch im Vergleich zu 2007 und 2011 signifikant schlechtere Ergebnisse. Für den Bereich *Problemlösen* sind im Vergleich von TIMSS 2019 mit den vorangegangenen Studien keine statistisch signifikanten Unterschiede auszumachen.

Einstellungen: Auf einer Skala von 1 bis 4 zeigt der Vergleich der mittleren Skalenwerte der vier TIMSS-Zyklen, dass in TIMSS 2019 (3.04) ein signifikant niedrigerer Wert als in allen anderen Studienzyklen vorliegt (2007: 3.15; 2011: 3.12; 2015: 3.10). Die Schülerinnen und Schüler zeigen also im aktuellen Studienzyklus eine weniger positive Einstellung zum Fach Mathematik als in den vorangegangenen Zyklen.

Trotz der Tatsache, dass anteilig weniger Grundschulkinder eine *hohe* und mehr Kinder eine *niedrige* positive Einstellung zur Mathematik aufweisen, ist die Einstellung der Grundschulkinder zur Mathematik insgesamt als positiv einzustufen. Dies belegt insbesondere der Anteil von 65 Prozent der Kinder, die eine *hohe* positive Einstellung zur Mathematik ausdrücken. Ferner weisen 19 Prozent der Kinder eine *niedrige* und 16 Prozent eine *mittlere* positive Einstellung zum Fach Mathematik auf. Zudem konnte ein positiver Zusammenhang zwischen der Einstellung zur Mathematik und den mathematischen Kompetenzen für Deutschland in TIMSS 2019 nachgewiesen werden.

Signifikante Unterschiede über die Zyklen hinweg lassen sich bei der Gruppe der Kinder mit *niedriger* und *hoher* positiver Einstellung finden. Es konnte gezeigt werden, dass in 2019 anteilig signifikant weniger Kinder eine *hohe* positive Einstellung aufweisen als noch in 2007. In der Gruppe der Schülerinnen und Schüler mit *niedriger* positiver Einstellung zur Mathematik liegt für 2019 ein signifikant größerer Wert als für 2011 und 2007 vor.

Mathematikbezogenes Selbstkonzept: Der mittlere Skalenwert für den aktuellen Zyklus liegt bei 3.11. Im Vergleich zu TIMSS 2015 (3.15) ergeben sich keine bedeutenden Unterschiede. Signifikant höhere Skalenmittelwerte lagen jedoch

bei den Zyklen von 2011 (3.24) und 2007 (3.18) vor. Dementsprechend hat sich der mittlere Skalenwert für das mathematikbezogene Selbstkonzept in 2019 im Vergleich zu 2015 kaum verändert, im Vergleich zu 2011 und 2007 ist er jedoch niedriger geworden. Er ist aber nach wie vor als hoch zu bezeichnen. Darüber hinaus ist ein positiver Zusammenhang zwischen dem mathematikbezogenen Selbstkonzept und mathematischen Kompetenzen für Deutschland in TIMSS 2019 zu verzeichnen.

Der Anteil der Grundschulkinder mit einem *hohen* mathematikbezogenen Selbstkonzept liegt in TIMSS 2019 bei 64 Prozent. Dieser Wert ist signifikant niedriger als in allen anderen Zyklen (2015: 67%; 2011: 71%; 2007: 69%). Auf der anderen Seite ist der Anteil der Schülerinnen und Schüler mit *mittlerem* mathematikbezogenen Selbstkonzept in TIMSS 2019 (25%) im Vergleich zu allen bisherigen Studienzyklen signifikant größer geworden (2015: 21%; 2011: 19%; 2007: 21%). Anteilig 12 Prozent der Kinder weisen in TIMSS 2019 ein *niedriges* Selbstkonzept auf. Dieser Wert unterscheidet sich nicht von den Werten der Zyklen von 2015 (12%) und 2011 (10%). Mit 9.6 Prozent artikulierten jedoch in TIMSS 2007 anteilig signifikant weniger Kinder ein *niedriges* mathematikbezogenes Selbstkonzept als in TIMSS 2019.

Abschließend bleibt festzuhalten, dass die Testleistungen der Schülerinnen und Schüler in Deutschland am Ende ihrer Grundschulzeit wie auch schon in den vorangegangenen Studien als vergleichsweise gut eingeordnet werden können, wenn man sich daran orientiert, dass der Leistungsmittelwert in Mathematik signifikant oberhalb des internationalen Mittelwertes aller Teilnehmerstaaten liegt. Die Leistungen der Schülerinnen und Schüler sind – wie im letzten Studienzyklus – im Mittelfeld der internationalen Rangreihe anzusiedeln. Im Vergleich zum Mittelwert der teilnehmenden EU-Staaten (VG_{EU} 527) beziehungsweise der OECD-Staaten (VG_{OECD} 529) fallen die Leistungen der Schülerinnen und Schüler aus Deutschland allerdings signifikant niedriger aus.

Insgesamt sind die Testleistungen jedoch in 2019 um 4 Punkte niedriger als bei der erstmaligen Teilnahme an TIMSS in 2007, um 7 Punkte niedriger und signifikant schlechter als in 2011 und um 1 Punkt niedriger im Vergleich zur letzten Teilnahme an TIMSS im Jahr 2015. Deutschland ist einer von sieben Staaten, für den im Vergleich der Resultate von 2019 mit 2007 keine positiv signifikanten Veränderungen der Testleistungen vorliegen. In 14 anderen Staaten ist dies der Fall. In der Tendenz gilt diese Grundaussage auch für den Vergleich von 2019 und 2011.

Ein Vergleich der Daten von 2007 und 2019 zeigt ferner, dass für die USA für Kompetenzstufe I sowie für Deutschland für Kompetenzstufe II als jeweils einziger Staat eine signifikante Steigerung der Anteile der Lernenden auf der Kompetenzstufe I beziehungsweise II zu verzeichnen ist. Im Kontrast dazu sind bei zwölf beziehungsweise zehn Staaten signifikante Reduzierungen der Anteile auf diesen Kompetenzstufen festzustellen. Bei acht beziehungsweise elf Teilnehmerstaaten können signifikante Steigerungen der Anteile auf der Kompetenzstufe IV beziehungsweise V verzeichnet werden, Deutschland gehört nicht dazu. Der Vergleich zwischen 2011 und 2019 zeigt in der Tendenz ein ähnliches Ergebnis, wohingegen zwischen 2015 und 2019 keine vergleichbar klaren Unterschiede zu konstatieren sind.

Die Kernbefunde für Deutschland im internationalen Vergleich lauten:

- in der Gesamtschau mittelmäßige Leistungen,
- ein hoher Anteil von 25 Prozent der Lernenden auf den Kompetenzstufen I und II,
- ein niedriger Anteil von 6 Prozent der Lernenden auf der Kompetenzstufe V.

Diese Befunde sind über die Untersuchungszeitpunkte hinweg seit 2007 im Wesentlichen stabil. Geht man einerseits davon aus, dass sich die Rahmenbedingungen für das Mathematiklernen in der Schule in den letzten Jahren durch zusätzliche Anforderungen, die sich beispielsweise durch Inklusion oder Integration, Personalmangel oder Unterrichtsausfall, Konzentrationsschwierigkeiten oder geringere Lernvoraussetzungen (Skorsetz, Bonanati & Kucharz, 2020; Stanat, Schipolowski, Rjosk, Weirich & Haag, 2017) ergeben, sicherlich nicht zum Positiven verändert haben, so ist diese Stabilität auf den ersten Blick nicht unerfreulich. Zu konstatieren ist jedoch andererseits, dass es viele andere Staaten besser schaffen, im Hinblick auf viele der untersuchten Aspekte Fortschritte zu machen, während die Leistungen in Deutschland mehr oder weniger stagnieren. Als vierter zentraler Befund ist also festzuhalten:

- eine weitgehende Stagnation in Deutschland bei positiven Entwicklungen in vielen anderen Staaten (auch Nachbarstaaten).

Zur Steigerung der Leistungen insgesamt, insbesondere aber zur deutlichen Verringerung der Lernenden mit ungünstigen Voraussetzungen zum Weiterlernen im Fach Mathematik und auch im Hinblick auf die Steigerung des Anteils der leistungsstarken Lernenden, bedarf es zum einen der besseren Unterstützung der Lehrkräfte bei der Bewältigung der oben beschriebenen, nicht allein fachbezogenen Anforderungen. Zum anderen sind Maßnahmen zur Weiterentwicklung des Mathematikunterrichts in der Primarstufe zu intensivieren und zu systematisieren. Dabei ist sowohl die Lehrkräfteausbildung noch konsequenter auf die professionsbezogenen Anforderungen auszurichten als auch das fachbezogene, kontinuierliche Weiterlernen im Beruf beständig zu unterstützen. Notwendig hierzu ist insbesondere ein System unterrichtsrelevanter, forschungsbasierter, fachbezogener Fortbildungs- und Unterrichtsentwicklungsprogramme, welche einerseits zur effektiveren Sicherung der mathematischen Basiskompetenzen und andererseits zur besseren, kontinuierlichen Unterstützung der Lernenden beim Erwerb prozessbezogener Kompetenzen beitragen.

Literatur

- 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. <https://doi.org/10.1007/978-3-322-83412-6>
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O. & Neubrand, J. (Hrsg.). (1997). *TIMSS – mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: Deskriptive Befunde*. Opladen: Leske + Budrich. <https://doi.org/10.1007/978-3-322-95096-3>
- Bonsen, M., Lintorf, K., Bos, W. & Frey, K. A. (2008). TIMSS 2007 Grundschule – Eine Einführung in die Studie. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 19–48). Münster: Waxmann.

- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Wendt, H., Köller, O., Selter, C., Schwippert, K. & Kasper, D. (2016). Wichtige Ergebnisse im Überblick. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 13–29). Münster: Waxmann.
- KMK (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland). (2005). *Bildungsstandards im Fach Mathematik für den Primarbereich (Jahrgangsstufe 4). Beschluss vom 15.10.2004*. München: Wolters Kluwer.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (Hrsg.). (2008). *Richtlinien und Lehrpläne für die Grundschule in Nordrhein-Westfalen*. Frechen: Ritterbach.
- Mullis, I. V. S. & Martin, M. O. (Hrsg.). (2013). *TIMSS 2015 assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S. & Martin, M. O. (Hrsg.). (2017). *TIMSS 2019 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Verfügbar unter: <http://timssandpirls.bc.edu/timss2019/international-results/>
- Selter, C., Walter, D., Walther, G. & Wendt, H. (2016). Mathematische Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 79–136). Münster: Waxmann.
- Selter, C., Walther, G., Wessel, J. & Wendt, H. (2012). Mathematische Kompetenz im internationalen Vergleich. Testkonzeption und Ergebnisse. In W. Bos, H. Wendt, O. Köller & C. Selter (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 69–122). Münster: Waxmann.
- Skorsetz, N., Bonanati, M. & Kucharz, D. (Hrsg.). (2020). *Diversität und soziale Ungleichheit: Herausforderungen an die Integrationsleistung der Grundschule*. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-27529-7>
- Stanat, P., Schipolowski S., Rjosk, C., Weirich, S. & Haag, N. (Hrsg.). (2017). *IQB Bildungstrend 2015. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Münster: Waxmann.
- Walther, G., Selter, C., Bensen, M. & Bos, W. (2008). Mathematische Kompetenz im internationalen Vergleich. Testkonzeption und Ergebnisse. In W. Bos, M. Bensen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 49–85). Münster: Waxmann.
- Wendt, H., Bos, W., Kasper, D., Walzebug, A., Goy, M. & Jusufi, D. (2016). Ziele, Anlage und Durchführung der Trends in International Mathematics and Science Study (TIMSS 2015). In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 31–77). Münster: Waxmann.

Kapitel 4

Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse

Mirjam Steffensky, Luise A. Scholz, Daniel Kasper und Olaf Köller

4.1 Einleitung

In der Grundschule sind die Naturwissenschaften kein eigenständiges Fach beziehungsweise keine eigenständigen Fächer, sondern gehören zu dem mehrperspektivischen Fach Sachunterricht¹. Neben der naturwissenschaftlichen Perspektive, die die Bereiche *Biologie* (belebte Natur) und *Physik/Chemie* (unbelebte Natur) umfasst, gehören die sozialwissenschaftliche (Politik – Wirtschaft – Soziales), die technische (Technik – Arbeit), die geografische (Räume – Naturgrundlagen – Lebenssituationen) und die historische Perspektive (Zeit – Wandel) zum Sachunterricht. Der Sachunterricht versteht sich als perspektivenintegrierendes Fach, in dem naturwissenschaftlich-technische und sozialwissenschaftliche Perspektiven verbunden werden, um so Kinder zu unterstützen, sich ihre natürliche und kulturelle und soziale Umwelt zu erschließen und darin zu handeln (Gesellschaft für die Didaktik des Sachunterrichts [GDSU], 2013).

Auch wenn sich dieser Ansatz eines integrierten Faches von den meisten anderen Ländern, in denen Naturwissenschaften ein eigenes Fach in der Grundschule darstellen, unterscheidet, nimmt Deutschland seit 2007 an der internationalen Studie *Trends in Mathematics and Science Study* (TIMSS) teil. Dadurch liegen repräsentative und belastbare Befunde über die Studienzyklen 2007 (Bos et al., 2008), 2011 (Bos, Wendt, Köller & Selter, 2012), 2015 (Wendt, Bos, Selter et al., 2016) und 2019 vor, die es ermöglichen, die naturwissenschaftlichen Kompetenzen von Viertklässlerinnen und Viertklässlern in Deutschland international und zwischen den Erhebungszyklen zu vergleichen. Im Folgenden werden zunächst wesentliche Befunde der ersten drei Studienzyklen zusammenge-

¹ In Bayern und Thüringen wird das Fach als Heimat- und Sachunterricht beziehungsweise Heimat- und Sachkunde bezeichnet. Der Einfachheit halber wird in diesem Kapitel von dem Fach Sachunterricht gesprochen.

fasst (Kleickmann, Brehl, Saß, Prenzel & Köller, 2012; Steffensky, Kleickmann, Kasper & Köller, 2016; Wittwer, Saß & Prenzel, 2008).

1. *Naturwissenschaftliche Kompetenzen*: Zwischen 2007 und 2015 bestand eine hohe Stabilität in den naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler aus Deutschland. So war der Mittelwert der naturwissenschaftlichen Kompetenzen identisch in den drei Studienzyklen. Er lag über dem Mittelwert aller teilnehmenden Staaten und war vergleichbar mit den Mittelwerten aus anderen EU- oder OECD-Staaten. Andere Staaten, zum Beispiel die Republik Korea, Japan oder Finnland, erreichten allerdings deutlich höhere Kompetenzmittelwerte. Differenziert man die Kompetenzen nach Inhaltsbereichen, so zeigten sich keine gravierenden Unterschiede zwischen der *Biologie*, *Physik/Chemie* und (Natur-) *Geografie*. Dies spricht dafür, dass die Schülerinnen und Schüler hinreichende Lerngelegenheiten in allen drei Bereichen erhalten haben. Gleiches gilt für die drei Anforderungsbereiche (*Reproduzieren*, *Anwenden* und *Problemlösen*). Die Streuung der Kompetenzen war in Deutschland vergleichsweise gering ausgeprägt, das heißt, die Kompetenzunterschiede der Schülerinnen und Schüler waren im internationalen Vergleich relativ klein.
2. *Leistungsstarke und -schwache Schülerinnen und Schüler*: Auch die Anteile der leistungsstarken und -schwachen Schülerinnen und Schüler waren sehr ähnlich in den drei Studienzyklen. So befand sich knapp ein Viertel der Schülerinnen und Schüler auf den beiden unteren Kompetenzstufen. Diese haben weniger gute Voraussetzungen für eine erfolgreiche Teilnahme am nachfolgenden naturwissenschaftlichen Unterricht der weiterführenden Schulen, weil ihnen wichtige Grundlagen fehlen. Gleichzeitig war der Anteil der Schülerinnen und Schüler, welche die höchste Kompetenzstufe erreichen, gering (7–10 %). Zum Vergleich: In Singapur waren es in allen drei Studienzyklen circa 35 Prozent.
3. *Einstellung zum Sachunterricht*: Der überwiegende Teil der Viertklässlerinnen und Viertklässler verfügte über eine ausgesprochen positive Einstellung zum Sachunterricht und ein hohes sachunterrichtsbezogenes Selbstkonzept eigener Fähigkeiten. Dabei ist zu berücksichtigen, dass hier nicht speziell nach den naturwissenschaftlichen Anteilen des Sachunterrichts gefragt wurde. Trotzdem ist es zu begrüßen, dass die Schülerinnen und Schüler eine positive Einschätzung zu dem Fach haben, in dem die Naturwissenschaften verortet sind.
4. *Geschlechterbedingte Disparitäten*: In den Studienzyklen 2007 und 2011 zeigten sich Leistungsunterschiede zugunsten der Jungen, die sich bei differenzierter Analyse auf Unterschiede in den Inhaltsbereichen *Physik/Chemie* und (Natur-) *Geografie* zurückführen ließen. Diese Unterschiede fanden sich in TIMSS 2015 nicht mehr. Dies liegt zum Teil an den sinkenden Leistungen der Jungen. Möglicherweise greifen hier auch Maßnahmen einer gezielten Förderung von Mädchen in den Naturwissenschaften, sodass sich deren Leistungen verbessert haben (Wendt, Steinmayr & Kasper, 2016).
5. *Herkunftsbedingte Disparitäten*: Auch in TIMSS 2015 zeigte sich die aus den vorangegangenen Studienzyklen und anderen Studien bekannte enge Kopplung zwischen sozialer Herkunft und den naturwissenschaftlichen Leistungen (Stubbe, Schwippert & Wendt, 2016). Die ausgeprägten migrationsbedingten Disparitäten wurden auch in TIMSS 2015 deutlich, im Vergleich zu TIMSS 2007 und TIMSS 2011, bei denen die Disparitäten auf ähnlichem Niveau lagen, hat der Effekt aber abgenommen (Wendt, Schwippert & Stubbe, 2016).

Zusammengefasst ergibt sich ein relativ stabiles Bild der naturwissenschaftlichen Leistungen über die Jahre 2007 bis 2015. In allen drei Studienzyklen wurde vor allem die gezielte Förderung von leistungsschwachen und leistungsstarken Schülerinnen und Schülern als zentrale Herausforderung für den naturwissenschaftsbezogenen Sachunterricht identifiziert. Durch die weitere Teilnahme von Deutschland an TIMSS 2019 lassen sich nun Trends in Bezug auf die naturwissenschaftlichen Kompetenzen von Viertklässlerinnen und Viertklässlern über den Zeitraum von zwölf Jahren untersuchen. Das hier vorliegende Kapitel stellt die naturwissenschaftsbezogenen Ergebnisse aus TIMSS 2019 dar und analysiert diese, auch vor dem Hintergrund der Ergebnisse aus den vorangegangenen Studienzyklen.

In den folgenden drei Abschnitten werden zunächst grundlegende Aspekte naturwissenschaftlicher Grundbildung in Deutschland (Abschnitt 4.2), die TIMSS-Rahmenkonzeption zur Erfassung naturwissenschaftlicher Kompetenzen (Abschnitt 4.3) sowie der TIMSS-Test zur Erfassung naturwissenschaftlicher Kompetenzen (Abschnitt 4.4) skizziert. Im abschließenden Abschnitt 4.5 werden die Ergebnisse des aktuellen Studienzyklus aus dem Jahr 2019 dargestellt. Die Darstellung orientiert sich dabei an den folgenden Fragestellungen, wobei die Unterschiede zu den Ergebnissen aus TIMSS 2007, 2011 und 2015 bei den jeweiligen Fragestellungen berücksichtigt werden:

1. Welches naturwissenschaftliche Kompetenzniveau zeigen Schülerinnen und Schüler in Deutschland am Ende der vierten Jahrgangsstufe im internationalen Vergleich? (Abschnitt 4.5.1)
2. Wie lassen sich die Leistungskennwerte in Deutschland auf die Kompetenzstufen im internationalen Vergleich einordnen? Wie groß sind die Gruppen der auffällig leistungsstarken und leistungsschwachen Kinder? (Abschnitt 4.5.2)
3. Welche Ergebnisse erzielen die deutschen Viertklässlerinnen und Viertklässler in den drei naturwissenschaftlichen Inhaltsbereichen (*Biologie, Physik/Chemie und Geografie*)? (Abschnitt 4.5.3)
4. Welche Ergebnisse erzielen die deutschen Viertklässlerinnen und Viertklässler in den drei kognitiven Anforderungsbereichen *Reproduzieren, Anwenden und Problemlösen*? (Abschnitt 4.5.4)
5. Wie hoch sind die Einstellung zu Naturwissenschaften und das naturwissenschaftsbezogene Selbstkonzept eigener Fähigkeiten der Schülerinnen und Schüler in Deutschland ausgeprägt? Zeigen sich Unterschiede in den naturwissenschaftsbezogenen Einstellungen und Selbstkonzepten zwischen Schülerinnen und Schülern verschiedener Kompetenzstufen? (Abschnitt 4.5.5)

Auf Ergebnisse zu Unterschieden in der naturwissenschaftlichen Kompetenz in Abhängigkeit von Geschlecht, sozialer Herkunft und Migrationshintergrund wird in den Kapiteln 8, 9 und 10 in diesem Band eingegangen.

4.2 Naturwissenschaftliche Grundbildung in Deutschland

Wie eingangs erwähnt, gehören die Naturwissenschaften zu dem multiperspektivischen Fach Sachunterricht. Während in den 1980/90er-Jahren Naturwissenschaften und insbesondere Themen der Physik und Chemie kaum eine Rolle im Sachunterricht spielten, deuten die aktuellen Lehrpläne, Schulbuchanalysen (Blaseio, 2009), die Befragung aus TIMSS 2015 (Steffensky et al., 2016) und auch die aktuellen Befragungen der Lehrkräfte aus TIMSS 2019 darauf hin, dass

die Naturwissenschaften im Sachunterricht wieder stärker in den Vordergrund getreten sind. So geben die Lehrkräfte im aktuellen Studienzyklus an, dass circa 40 Prozent der behandelten Themen der naturwissenschaftlichen Perspektive zugeordnet werden können (Tabelle 4.1). Das bedeutet, dass der größte Anteil der Stunden des Sachunterrichts nach Einschätzung der Lehrkräfte auf die Naturwissenschaften entfällt. Gleichzeitig fällt aber auf, dass die Streuung hier besonders groß ist. So geben die Lehrkräfte Werte zwischen 10 und 100 Prozent an. Die Ergebnisse sind ähnlich zu denen aus TIMSS 2015, sie decken sich auch hinsichtlich der Rangfolge der fünf Perspektiven mit denen aus dem vorangegangenen Studienzyklus. Bei der Interpretation der Daten ist allerdings zu berücksichtigen, dass es sich bei der Stichprobe der Lehrkräfte (anders als bei den Schülerinnen und Schülern) nicht um eine national repräsentative Stichprobe handelt. Auch ist die Validität von Selbstberichten zum Teil beeinträchtigt, weil sie beispielsweise durch soziale Erwünschtheit verzerrt werden können. Denkbar wäre, dass Lehrkräfte, die an einer Studie zu den Naturwissenschaften teilnehmen, hier besonders hohe Werte angeben. Wie belastbar die Daten also sind, lässt sich schwer einschätzen.

Tabelle 4.1: Einschätzung der Lehrkräfte zum prozentualen Anteil der Unterrichtsstunden in Klasse drei und vier in den fünf Perspektiven des Perspektivrahmens

Perspektiven	M	(SE)	SD	Min - Max
Sozialwissenschaftliche Perspektive Politik – Wirtschaft – Soziales	14.9	(0.5)	7.1	0 - 50
Naturwissenschaftliche Perspektive Belebte und unbelebte Natur	39.7	(0.9)	12.1	10 - 100
Geografische Perspektive Räume – Naturgrundlagen – Lebenssituationen	19.5	(0.6)	8.1	0 - 50
Historische Perspektive Zeit – Wandel	14.4	(0.8)	7.9	0 - 60
Technische Perspektive Technik – Arbeit	12.1	(0.4)	6.0	0 - 40

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Um ein differenziertes Bild der naturwissenschaftlichen Perspektive zu erhalten, wurden die Lehrkräfte auch gefragt, wie sich der von ihnen angegebene naturwissenschaftliche Teil des Sachunterrichts (mittlerer Anteil 39.7%) auf Themen der belebten Natur (*Biologie*) und solche der unbelebten Natur (*Physik/Chemie*) aufteilt (Tabelle 4.2). Die Lehrkräfte geben an, die Themen der belebten und unbelebten Natur im Verhältnis von drei zu zwei zu unterrichten. Dies erscheint mit dem Blick auf Lehrpläne und Studienschwerpunkte von Sachunterrichtslehrkräften nicht überraschend. Gleichzeitig deuten die Ergebnisse an, dass die früher beschriebene Schwerpunktsetzung auf die belebte Natur nicht mehr so stark ausgeprägt ist. So zeigten sich auch anders als in TIMSS 2007 in den letzten beiden TIMSS-Zyklen keine Unterschiede mehr in den Leistungen zwischen Inhalten der *Biologie* und der *Physik/Chemie*. Veränderungen in der

Ausbildung von Sachunterrichtslehrkräften und/oder Programme zur Weiterentwicklung des Bereichs *Mathematik, Informatik, Naturwissenschaft, Technik* (MINT), zum Beispiel *Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts* (SINUS) an Grundschulen, die den Fokus eher auf Themen der unbelebten Natur gelegt haben, zeigen hier möglicherweise Wirkungen.

Tabelle 4.2: Einschätzung der Lehrkräfte zur prozentualen Aufteilung der naturwissenschaftlichen Perspektive auf Themen der belebten und unbelebten Natur

Naturwissenschaftliche Perspektive	M	(SE)	SD	Min - Max
Belebte Natur	23.8	(0.9)	11.2	5 - 90
Unbelebte Natur	15.8	(0.4)	6.4	0 - 40

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Der (naturwissenschaftliche) Sachunterricht verfolgt multikriteriale Ziele, die sich unter anderem im Perspektivrahmen der GDSU ablesen lassen (GDSU, 2013). Dieser hat keinen bindenden Charakter, stellt aber einen wichtigen Orientierungsrahmen für die Weiterentwicklung von Bildungsplänen und didaktischen Materialien dar. Der Perspektivrahmen benennt übergreifende Ziele, zum Beispiel die Entwicklung von Interesse und von Kommunikations- und Reflexionskompetenzen, und spezifische Ziele für die fünf Perspektiven. Unter der naturwissenschaftlichen Perspektive werden drei zentrale Bereiche aufgeführt (GDSU, 2013):

1. Der inhaltliche Bereich, in dem Kompetenzen zu grundlegenden Konzepten und Zusammenhängen in den folgenden fünf Themenbereichen entwickelt werden sollen:
 - Unbelebte Natur – Eigenschaften von Stoffen/Körpern,
 - Unbelebte Natur – Stoffumwandlungen,
 - Unbelebte Natur – physikalische Vorgänge,
 - Belebte Natur – Pflanzen, Tiere und ihre Unterteilungen,
 - Belebte Natur – Entwicklungs- und Lebensbedingungen von Lebewesen.
2. Die Generierung von naturwissenschaftlichem Wissen: Hierzu werden folgende naturwissenschaftliche Denk-, Arbeits- und Handlungsweisen benannt:
 - Naturphänomene sachorientiert (objektiv) untersuchen,
 - Methoden zum Erkenntnisgewinn aneignen und anwenden,
 - Naturphänomene auf Regelmäßigkeiten zurückführen,
 - Konsequenzen aus naturwissenschaftlichen Erkenntnissen für das Alltags-handeln ableiten,
 - Naturwissenschaftliches Lernen bewerten und reflektieren.

Eng im Zusammenhang mit diesem prozessbezogenen Wissen steht auch die Frage nach der Beschaffenheit von naturwissenschaftlichem Wissen, das beispielsweise in weiten Teilen als vorläufig beschrieben wird. Dieses epistemische Wissen wird allerdings lediglich erwähnt und nicht weiter ausdifferenziert.

3. Orientierungen für einen verantwortlichen Umgang mit der Natur: Die Bedeutung dieses Bereiches wird auch durch das zentrale Querschnittsthema des Sachunterrichts ‚Bildung für eine nachhaltige Entwicklung‘ unterstrichen.

Im Rahmen der TIMSS-Konzeption ist darüber hinaus auch die Geografie ein Inhaltsbereich, wobei es hier vor allem um naturgeografische und nicht um humangeografische Inhalte geht. Hierzu findet sich im Perspektivrahmen unter der geografischen Perspektive der inhaltliche Schwerpunkt *Naturphänomene, natürliche Zyklen und Kreisläufe*, wozu Themen wie Wetter, Jahreszeiten oder der Mondlauf gehören. Eine der hier relevanten Denk-, Arbeits- und Handlungsweisen ist es, Ordnungsmuster zu räumlichen Situationen und zu Natur-Mensch-Beziehungen aufzubauen und weiterzuentwickeln.

Für die TIMSS-Konzeption ist außerdem das Thema Gesundheit relevant, das dem Inhaltsbereich *Biologie* zugeordnet wird. Gesundheit wird im Perspektivrahmen als wichtiger, perspektivenübergreifender Themenbereich genannt. In allen Bereichen geht es um die Entwicklung von anwendbarem Wissen, das an die lebensweltlichen Erfahrungen von Kindern anknüpft und gleichzeitig anschlussfähig für die spätere kumulative Weiterentwicklung des Wissens ist.

Für den Sachunterricht gibt es bislang keine Bildungsstandards, im Perspektivrahmen sind aber Kompetenzerwartungen zu den Inhalten und Prozessen formuliert, die die oben genannten Bereiche konkretisieren. Beispiele aus dem Bereich der belebten und unbelebten Natur sowie der perspektivbezogenen Denk-, Arbeits- und Handlungsweisen wären:

Die Schülerinnen und Schüler können

- morphologische Merkmale von Pflanzen (Teile der Pflanze) und Tieren (Körperbau) untersuchen, benennen, beschreiben und vergleichen (verschiedene Pflanzen-, Tierarten),
- Rosten und Verbrennung als Umwandlung von Stoffen beschreiben,
- einfache Ursache-Wirkungszusammenhänge erkennen (z. B. Verdrängung von Wasser durch Luft) und angemessen sprachlich darstellen.

Entsprechende Ziele und auch Themen der belebten und unbelebten Natur beziehungsweise Denk-, Arbeits- und Handlungsweisen werden in ähnlicher Weise auch in den Bildungsplänen der 16 Länder benannt. Fast alle Bildungspläne für den Sachunterricht wurden in den letzten zehn Jahren überarbeitet (eine Ausnahme stellen die Bildungspläne aus Bremen und Mecklenburg-Vorpommern dar, wobei der Bremer Bildungsplan aktuell überarbeitet wird). Die Bildungspläne formulieren Kompetenzerwartungen in der Regel für das Ende der zweiten und vierten Klasse. Dabei sind die prozessbezogenen und inhaltsbezogenen Kompetenzerwartungen in fast allen Lehrplänen integriert dargestellt. Dies entspricht auch der regulären Umsetzung im Unterricht. Die eher heuristische Trennung zwischen inhalts- und prozessbezogenem Wissen, wie man sie im Perspektivrahmen oder zum Beispiel im Hamburger Bildungsplan findet, verdeutlicht aber, dass die naturwissenschaftlichen Denk- und Arbeitsweisen selbst auch ein Lerngegenstand im naturwissenschaftlichen Unterricht sind.

Diese multikriterialen Ziele des Sachunterrichts, die inhalts- und prozessbezogenes Wissen, Interesse und Verantwortungsbewusstsein für unsere Umwelt umfassen, lassen sich gut in Einklang bringen mit dem Grundbildungskonzept *Scientific Literacy* (Bybee & McCrae, 2011). *Scientific Literacy* umfasst Wissen über zentrale Begriffe, Konzepte und Theorien (inhaltsbezogenes Wissen), Wissen über naturwissenschaftliche Denk- und Arbeitsweisen (prozessbezo-

genes Wissen) und epistemisches Wissen sowie motivationale Orientierungen und Einstellungen gegenüber den Naturwissenschaften. Diese verschiedenen Komponenten stellen im Zusammenspiel die Grundlage für weiteres Lernen und die aktive Mitgestaltung unserer Lebensumwelt dar. Inwiefern diese multikriterialen Ziele gleichermaßen im Schulalltag berücksichtigt werden oder ob der Fokus nicht vor allem auf dem inhaltlichen Wissen und den dazugehörigen Kompetenzen liegt, ist unklar.

Neben der Frage, welche Inhalte in welcher Quantität umgesetzt werden, ist für die Entwicklung der Schülerinnen und Schüler vor allem die Qualität der Lernumgebungen entscheidend. Hierzu gibt es bislang nur wenige Befunde aus dem Sachunterricht (Fauth, Decristan, Rieser, Klieme & Büttner, 2014; Kleickmann, Steffensky & Praetorius, 2020), die auf Schülerratings beruhen. In Kapitel 6 in diesem Band werden die Befunde aus dem aktuellen TIMSS-Zyklus dargestellt. Neben den generischen Basisdimensionen der Unterrichtsqualität ist aus der fachspezifischen Perspektive die Frage wichtig, inwiefern kumulative Lernwege, also Lernwege, bei denen einzelne Themen miteinander verknüpft und zentrale Konzepte angebahnt werden, im Sachunterricht umgesetzt werden, indem zum Beispiel übergeordnete Konzepte wie Materie gezielt adressiert werden oder explizite Lernumgebungen für die Auseinandersetzung mit naturwissenschaftlichen Denk- und Arbeitsweisen, zum Beispiel *Was ist Beobachten?*, in den Unterricht integriert werden. Solche Lerngelegenheiten gelten als wichtig für die horizontale und vertikale Vernetzung des Wissens und damit auch für das Weiterlernen in den weiterführenden Schulen (Harlen, 2015; Smith, Wiser, Anderson & Krajcik, 2006).

4.3 TIMSS-Rahmenkonzeption zur Erfassung der naturwissenschaftlichen Kompetenzen

Die TIMSS-Rahmenkonzeption (Mullis & Martin, 2017) beschreibt die für die Studie relevanten naturwissenschaftlichen Inhaltsbereiche, die mit dem TIMSS-Test erfasst werden, sowie die zentralen Denkprozesse, die dem Lösen der Aufgaben zugrunde liegen. Letztere werden als kognitive Anforderungsbereiche bezeichnet. Wie auch in den vorangegangenen Studienzyklen umfassen die Inhaltsbereiche in TIMSS 2019 *Biologie*, *Physik/Chemie* und *Geografie* und die kognitiven Anforderungsbereiche *Reproduzieren* von Wissen, *Anwenden* von Wissen sowie *Problemlösen*. Die für die naturwissenschaftliche Grundbildung wichtigen prozessbezogenen Kompetenzen (im Perspektivrahmen als Denk-, Arbeits- und Handlungsweisen benannt) werden bei TIMSS nicht separat erhoben, sondern sind in die Inhaltsbereiche integriert.

Ziel der Rahmenkonzeption ist es, naturwissenschaftliche Kompetenz einerseits breit zu erfassen, andererseits aber Analysen von Teilkompetenzen zu ermöglichen, die ein differenziertes Bild zulassen. Hierfür ist die Aufteilung in die drei Inhaltsbereiche und drei Anforderungsbereiche grundlegend. Gleichzeitig erfordert die Rahmenkonzeption von TIMSS die Abdeckung zentraler Bereiche der nationalen Curricula der teilnehmenden Staaten, da TIMSS anders als das *Programme for International Student Assessment* (PISA) (OECD, 2016; OECD Global Science Forum, 2006) das Wissen über die curricularen Inhalte untersucht, während bei PISA die Anwendung von naturwissenschaftlichem Wissen in Alltagskontexten im Vordergrund steht. Um längerfristige Trends in den Kompetenzständen im nationalen und internationalen Vergleich untersuchen zu

können, ist die Vergleichbarkeit der Studienzyklen notwendig. Entsprechend gibt es große Überschneidungen zu den zentralen Bereichen der vorangegangenen Rahmenkonzeption. Gleichzeitig erfordern Änderungen in Curricula immer wieder kleinere Anpassungen in der Rahmenkonzeption. Zudem ermöglicht die Umstellung auf die computerbasierte Testung die stärkere Berücksichtigung von Aufgaben, mit denen das Verständnis und die Anwendung naturwissenschaftlicher Denk- und Arbeitsweisen erhoben werden können, die in vielen Curricula eine zunehmende Bedeutung erfahren. Das bedeutet, dass die Rahmenkonzeption und Aufgabenentwicklung einerseits anschlussfähig an die vorangegangenen Rahmenkonzeptionen sein muss, andererseits aber auch Entwicklungen abbilden muss. Im Folgenden wird ein detaillierter Überblick über die Rahmenkonzeption gegeben, wobei auch entsprechende Änderungen zu TIMSS 2015 skizziert werden.

4.3.1 Naturwissenschaftliche Inhaltsbereiche

Die drei übergeordneten Inhaltsbereiche *Biologie*, *Physik/Chemie* und *Geografie* sind bei TIMSS 2019 durch jeweils drei bis fünf Themengebiete beschrieben, die wiederum nochmal in einzelne Themen differenziert werden (siehe Tabelle 4.3). Zur Konkretisierung der Erwartungen gibt es zu allen Themen Kompetenzerwartungen, ausführliche Informationen finden sich bei Mullis und Martin (2017).

Der Inhaltsbereich *Biologie* umfasst fünf Themengebiete: 1. Kennzeichen des Lebendigen und Lebensvorgänge von Organismen, 2. Lebenszyklen, Fortpflanzung und Vererbung, 3. Organismen in verschiedenen Lebensräumen, 4. Ökosysteme und 5. Gesundheit. Von den Schülerinnen und Schülern wird grundlegendes Wissen über Lebewesen und ihre Eigenschaften sowie über ihre Entwicklung und wie sie mit anderen Lebewesen und ihrer Umgebung interagieren, erwartet. Das Themengebiet Gesundheit fokussiert auf alltagspraktisches Wissen über die Verbreitung von und den Schutz vor ansteckenden Krankheiten sowie auf gesunde Ernährungs- und Verhaltensweisen.

Der Inhaltsbereich *Physik/Chemie* umfasst drei Themengebiete: 1. Stoffe (Klassifizierung von Stoffen, Stoffeigenschaften und Veränderungen von Stoffen), 2. Energieformen und Energieumwandlung sowie 3. Kräfte und Bewegungen. Von den Schülerinnen und Schülern wird grundlegendes Wissen über Stoffklassen, die drei Aggregatzustände und einfache physikalische und chemische Veränderungen von Stoffen erwartet. Außerdem sollen sie gängige Energieformen und -quellen, zum Beispiel Sonne, Kohle, aufgestautes Wasser, und ihre praktische Anwendung kennen. Zudem sollen sie ein erstes phänomenbasiertes Wissen über Licht, Schall, Elektrizität und Magnetismus haben. Auch das dritte Themengebiet Kräfte und Bewegungen zielt ‚lediglich‘ auf ein phänomenbasiertes Verständnis von Kräften und Bewegungen ab, zum Beispiel Änderungen von Bewegungen durch Ziehen oder Schieben erklären. Für diesen Bereich auch relevant sind erste Kenntnisse über einfache Maschinen, mit denen sich Kraft sparen lässt.

Der Inhaltsbereich *Geografie* umfasst drei Themengebiete: 1. Charakteristische Landschaftsmerkmale, Ressourcen und Erdgeschichte, 2. Wetter und Klima sowie 3. die Erde im Sonnensystem. Von den Schülerinnen und Schülern wird grundlegendes Wissen über Merkmale der Erde und langfristige Veränderungen von Landschaften erwartet. Außerdem sollen sie Wetter-

Tabelle 4.3: Teilgebiete der einzelnen naturwissenschaftlichen Inhaltsbereiche

Inhaltsbereiche	Teilgebiete
Biologie	<p>1. Kennzeichen des Lebendigen und Lebensvorgänge von Organismen</p> <ul style="list-style-type: none"> - Unterschiede zwischen Lebewesen und unbelebten Dingen und was Lebewesen benötigen, um zu leben (z.B. Luft, Nahrung, Wasser, Lebensraum) - Baumerkmale und Verhalten von Tier- und Pflanzengruppen (z.B. Insekten, Vögel, Säugetiere, Reptilien, Blütenpflanzen) - Zusammenhang zwischen Struktur und Funktion (z.B. Knochen, Wurzeln) <p>2. Lebenszyklen, Fortpflanzung und Vererbung</p> <ul style="list-style-type: none"> - Lebenszyklen bekannter Tiere und Pflanzen (z.B. Schmetterlinge) - Vererbung und Fortpflanzungsstrategien (z.B. Samenpflanzen, Brutpflege) <p>3. Organismen, Umwelt und ihre Interaktion</p> <ul style="list-style-type: none"> - Anpasstheit von Lebewesen an ihren Lebensraum durch Körperbau und Verhalten (z.B. tiefe Wurzeln, Färbung) - Reaktionen von Lebewesen auf Bedingungen ihrer Umwelt (z.B. Licht, Temperatur, Gefahr) - Einfluss des Menschen auf die Umwelt (z.B. Verschmutzung von Wasser und Luft) <p>4. Ökosysteme</p> <ul style="list-style-type: none"> - Diversität in den Lebensräumen und bekannte Ökosysteme (z.B. Wald, Teich) - Einfache Nahrungsketten - Wettbewerb in Ökosystemen (z.B. Nahrung, Platz) <p>5. Gesundheit</p> <ul style="list-style-type: none"> - Symptome von Krankheiten, Übertragung ansteckender Krankheiten und die Möglichkeiten, sich vor Ansteckung zu schützen - Maßnahmen der Gesunderhaltung (z.B. Ernährungsweisen, Hygiene)
Physik/Chemie	<p>1. Stoffe (Klassifizierung von Stoffen, Stoffeigenschaften und Veränderungen von Stoffen)</p> <ul style="list-style-type: none"> - Aggregatzustände und ihre spezifischen Merkmale (z.B. Form, Volumen) - Physikalische Eigenschaften von Stoffen und Stoffklassen (z.B. Leitfähigkeit) - Magnetische Anziehung und Abstoßung - Physikalische Veränderungen in Alltagssituationen (z.B. beim Schmelzen, Verdunsten, Lösen) - Chemische Veränderungen in Alltagssituationen (z.B. beim Rosten, Verbrennen, Kochen) <p>2. Energieformen und Energieumwandlung</p> <ul style="list-style-type: none"> - Energiequellen (z.B. Sonne, Wind, Kohle) und deren Nutzung (z.B. für Wärme, Licht) - Licht und Schall in Alltagssituationen (z.B. Lichtquellen, Schatten) - Wärmeübertragung - Elektrizität und einfache Stromkreise <p>3. Kräfte und Bewegungen</p> <ul style="list-style-type: none"> - Bekannte Kräfte (z.B. Erdanziehung, Reibung, Ziehen, Schieben) und deren Einfluss auf die Bewegung von Objekten - Einfache Maschinen, die den Kraftaufwand verringern (z.B. Hebel oder Rampen)
Geografie	<p>1. Charakteristische Landschaftsmerkmale, Ressourcen und Erdgeschichte</p> <ul style="list-style-type: none"> - Charakteristische Landschaftsmerkmale (z.B. Vorkommen von Salz- und Süßwasser) - Natürliche Ressourcen (z.B. Boden, Wasser, Wind) - Erdgeschichte (z.B. Landschaftsbildung, Fossilien) <p>2. Wetter und Klima</p> <ul style="list-style-type: none"> - Wetter und Klima auf der Erde (z.B. Wetterphänomene, Wasserkreislauf, Wettererscheinungen im Laufe eines Tages, Jahres und an verschiedenen Orten) <p>3. Die Erde im Sonnensystem</p> <ul style="list-style-type: none"> - Planeten des Sonnensystems und ihre Bewegungen (z.B. Mondphasen) - Auswirkungen der Bewegungen der Erde (z.B. Tag und Nacht, Jahreszeiten)

erscheinungen über Tage, Jahreszeiten und verschiedene Orte erläutern und die Entstehung von Tag und Nacht oder Jahreszeiten mit der Erdbewegung erklären können. Zusammenfassend zeigt sich eine große Überlappung zwischen den Inhalten und dazugehörigen Themenbereichen aus TIMSS 2019 und den Empfehlungen und Vorgaben aus dem Perspektivrahmen der GDSU beziehungsweise den Bildungsplänen für den Sachunterricht.

4.3.2 Kognitive Anforderungsbereiche

Neben der inhaltlichen Dimension werden in der TIMSS-Rahmenkonzeption auch die drei Anforderungsbereiche *Reproduzieren* von Wissen, *Anwenden* von Wissen sowie *Problemlösen* differenziert. Jedem kognitiven Anforderungsbereich lassen sich verschiedene dahinterliegende Denkprozesse zuordnen, die in Tabelle 4.4 dargestellt sind.

Tabelle 4.4: Denkprozesse der kognitiven Anforderungsbereiche

Reproduzieren	Anwenden	Problemlösen
<ul style="list-style-type: none"> - Erinnern/Wiedererkennen - Beschreiben - Angemessene Beispiele finden 	<ul style="list-style-type: none"> - Vergleichen, Unterschiede feststellen und Klassifizieren - Zusammenhänge herstellen - Anwenden von Modellen - Informationen interpretieren - Erklären 	<ul style="list-style-type: none"> - Analysieren von Problemen - Synthetisieren - Fragen formulieren, Hypothesen aufstellen und Vorhersagen treffen - Experimente planen - Evaluieren - Schlussfolgerungen ziehen - Generalisieren - Begründen

Im Anforderungsbereich *Reproduzieren* geht es im Kern darum, Fakten zu naturwissenschaftlichen Sachverhalten abzurufen, Sachverhalte und Beziehungen, zum Beispiel zwischen Lebewesen oder Materialien, zu beschreiben und angemessene Beispiele zu nennen. Auch die richtige Verwendung von Fachbegriffen und Kenntnisse über naturwissenschaftliche Instrumente gehören dazu. Die hier erforderlichen Wissens Elemente legen den Grundstein für die beiden anderen Anforderungsbereiche. Beim *Anwenden* liegt der Schwerpunkt darauf, erlernte naturwissenschaftliche Konzepte auf bekannte Situationen zu übertragen und Zusammenhänge herzustellen, um Lösungen zu finden und Erklärungen zu generieren. In diesen Anforderungsbereich fällt auch die Nutzung von Repräsentationen wie Modellen oder Diagrammen, um einen Sachverhalt zu erläutern. Beim *Problemlösen* kommt es darauf an, auch unbekannte und komplexere naturwissenschaftliche Probleme zu analysieren. Dieser kognitive Anforderungsbereich umfasst anspruchsvolle naturwissenschaftliche Aktivitäten wie das Planen von Experimenten, das Ziehen von Schlussfolgerungen aus vorhandener Evidenz oder das Evaluieren und Begründen.

Außer den Inhaltsbereichen (Abschnitt 4.3.1) und den kognitiven Anforderungsbereichen werden in der TIMSS-Rahmenkonzeption zusätzlich fünf zentrale naturwissenschaftliche Denk- und Arbeitsweisen (*science practices*) aufgeführt:


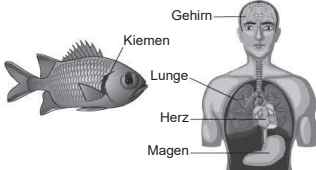

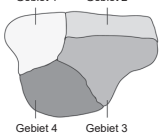
- überprüfbare Fragen entwickeln, die auf Beobachtungen basieren,
- Evidenz generieren, indem systematische Untersuchungen durchgeführt werden, mit denen eine Hypothese bestätigt oder widerlegt werden kann,
- mit Daten arbeiten, was das Darstellen und Beschreiben von Daten sowie Erkennen und Interpretieren von Mustern und Zusammenhängen in den Daten umfasst,
- Forschungsfragen beantworten, indem evidenzbasiert auf die Hypothesen Bezug genommen wird,
- evidenzbasiertes Argumentieren, um Erklärungen und Schlussfolgerungen zu entwickeln, zu begründen und auf neue Situationen anzuwenden.

Diese Denk- und Arbeitsweisen stellen keine eigene Dimension in der Rahmenkonzeption dar, sondern sind in die Inhaltsbereiche und die kognitiven Anforderungen integriert. Sie werden also in den meisten Testaufgaben nicht isoliert erfasst, sondern durch Aufgaben im Kontext der Inhaltsgebiete, die verschiedene kognitive Aktivitäten erfordern. Der Schwerpunkt liegt dabei auf dem Arbeiten mit Daten und dem evidenzbasierten Argumentieren. Hier lassen sich viele Überschneidungen zu den zwei Bereichen *Anwenden* und *Problemlösen* feststellen. Beispielsweise ist das Vergleichen und Feststellen von Unterschieden (Anforderungsbereich *Anwenden*) notwendig, um mit Daten zu arbeiten. Im Rahmen der computerbasierten Testung in TIMSS 2019 wurden spezifische Erkenntnisgewinnungsaufgaben erprobt. In diesen Aufgaben werden alltags- oder laborspezifische Situationen simuliert, in denen die Schülerinnen und Schüler neben dem Inhaltswissen auch Denk- und Arbeitsweisen anwenden müssen, um ein geeignetes Experiment durchzuführen oder ein Problem zu lösen. Diese Aufgaben sind den beiden kognitiven Anforderungsbereichen *Anwenden* und *Problemlösen* zugeordnet. Im aktuellen Studienzyklus gehen diese Aufgaben nicht in die Analysen ein. Möglicherweise werden sie aber zukünftig aufgenommen, um den Bereich der Denk- und Arbeitsweisen umfassender abzubilden.

4.3.3 Beispielaufgaben

Um die Kernbereiche der TIMSS-Rahmenkonzeption zu visualisieren, sind in Abbildung 4.1 Beispielaufgaben für die drei Inhalts- und drei Anforderungsbereiche dargestellt. Implizit sind darin auch die naturwissenschaftlichen Denk- und Arbeitsweisen enthalten. Anders als die Inhaltsbereiche sind die kognitiven Anforderungsbereiche nicht immer eindeutig voneinander abgrenzbar. So sind beispielsweise das *Anwenden* und das *Problemlösen* auch aus theoretischer Perspektive ähnliche kognitive Prozesse. Zudem hängen die kognitiven Prozesse, die zur Bearbeitung einer Aufgabe notwendig sind, stark vom Vorwissen der Schülerinnen und Schüler ab. So kann eine Aufgabe für manche Schülerinnen und Schüler aufgrund von mangelnder Erfahrung das *Anwenden* von Wissen in einem neuen Beispiel oder Kontext erfordern, während es für andere lediglich das *Reproduzieren* von Fakten bedeutet. Der Aufgabenbereich kann also nicht vollständig von der bearbeitenden Person gelöst werden.

Abbildung 4.1: Beispielaufgaben für die Inhaltsbereiche und kognitiven Anforderungsbereiche

	Reproduzieren	Anwenden	Problemlösen																				
Biologie	<p>Auf dem Bild unten siehst du eine Wüste.</p>  <p>Nenne zwei Lebewesen, die auf dem Bild zu sehen sind.</p> <ol style="list-style-type: none"> 1. <input type="text" value="Eidechse"/> 2. <input type="text" value="Kaktus"/> <p>Nenne zwei nicht-belebte Dinge, die auf dem Bild zu sehen sind.</p> <ol style="list-style-type: none"> 1. <input type="text" value="Stein"/> 2. <input type="text" value="Sonne"/> <p>(int. .46/nat. .23)</p> <p>In dieser Aufgabe müssen aus einer Abbildung zwei Lebewesen und zwei unbelebte Dinge identifiziert werden. Grundsätzlich ist hier ein erstes Wissen über Kennzeichen des Lebendigen. Dabei fällt es den Schülerinnen und Schülern oft schwer, Pflanzen als Lebewesen zu erkennen. Der zweite Teil der Aufgabe birgt also die Schwierigkeit, keine der Pflanzen zu benennen.</p>	 <p>Welches Organ des Menschen hat dieselbe Funktion wie die Kiemen von Fischen?</p> <p><input type="radio"/> A Gehirn <input checked="" type="radio"/> B Lunge <input type="radio"/> C Herz <input type="radio"/> D Magen</p> <p>(int. .73/nat. .66)</p> <p>In dieser Aufgabe muss das menschliche Organ identifiziert werden, das dieselbe Funktion wie die Kiemen beim Fisch hat. Zum Lösen der Aufgabe muss Wissen über die Funktionen der Organe verschiedener Lebewesen vorhanden sein und dieses in Beziehung gesetzt werden.</p>	<p>In einem Sommer bemerkt Leo, dass nachts draußen weniger Insekten sind als früher. Er bemerkt auch, dass mehr Fledermäuse da sind.</p> <p>Wie kann die gestiegene Anzahl von Fledermäusen die gesunkene Anzahl von Insekten erklären?</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>Fledermäuse fressen Insekten.</p> </div> <p>(int. .50/nat. .45)</p> <p>Zum Lösen dieser Aufgabe müssen aus den zwei vorgegebenen Beobachtungen – mehr Fledermäuse und weniger Insekten – Schlussfolgerungen gezogen und diese erläutert werden. Hier ist die naturwissenschaftliche Denk- und Arbeitsweise ‚Schlussfolgern‘ implizit enthalten. Das Lösen der Aufgabe erfordert ein Verständnis von Räuber-Beute-Beziehungen.</p>																				
Physik/Chemie	<p>Welchen Zustand haben die folgenden Materialien bei Zimmertemperatur?</p> <p>Klicke bei jedem Material einen Kreis an. In der ersten Reihe wurde die richtige Antwort schon angeklickt.</p> <table border="1"> <thead> <tr> <th>Material</th> <th>fest</th> <th>flüssig</th> <th>gasförmig</th> </tr> </thead> <tbody> <tr> <td>Papier</td> <td><input checked="" type="radio"/> A</td> <td><input type="radio"/> B</td> <td><input type="radio"/> C</td> </tr> <tr> <td>Sauerstoff</td> <td><input type="radio"/> A</td> <td><input type="radio"/> B</td> <td><input checked="" type="radio"/> C</td> </tr> <tr> <td>Öl</td> <td><input type="radio"/> A</td> <td><input checked="" type="radio"/> B</td> <td><input type="radio"/> C</td> </tr> <tr> <td>Kochsalz</td> <td><input checked="" type="radio"/> A</td> <td><input type="radio"/> B</td> <td><input type="radio"/> C</td> </tr> </tbody> </table> <p>(int. .76/nat. .61)</p> <p>In dieser Aufgabe muss der Aggregatzustand verschiedener Alltagsmaterialien genannt werden. Zum Lösen der Aufgabe muss der Fachbegriff Sauerstoff bekannt sein. Öl wird aufgrund der Viskosität oft nicht als Flüssigkeit erkannt, während feste Stoffe wie Kochsalz, die sich gießen lassen, oft den Flüssigkeiten zugeordnet werden.</p>	Material	fest	flüssig	gasförmig	Papier	<input checked="" type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	Sauerstoff	<input type="radio"/> A	<input type="radio"/> B	<input checked="" type="radio"/> C	Öl	<input type="radio"/> A	<input checked="" type="radio"/> B	<input type="radio"/> C	Kochsalz	<input checked="" type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C	<p>Anna stellt eine Tasse aus nassem Ton auf den Tisch. Einige Tage später ist der Ton trocken.</p> <p>Was ist mit dem Wasser im Ton passiert?</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>Das Wasser geht in die Luft.</p> </div> <p>(int. .25/nat. .21)</p> <p>In dieser Aufgabe muss das Wissen über den Prozess der Verdunstung auf einen für viele Grundschülerinnen und Grundschüler weniger bekannten Kontext angewandt werden. Für die Lösung der Aufgabe können die Grundschülerinnen und Grundschüler den Begriff Verdunstung nutzen oder den Vorgang umschreiben.</p>	<p>Karl untersucht, wie man die gleiche Menge Zucker schnell in Wasser auflösen kann. Er bereitet Versuche vor.</p> <p>Klicke bei jedem Versuch das Glas an, in dem sich der Zucker schneller auflösen wird.</p> <p>Versuch 1 unterschiedliche Temperatur</p> <p>Versuch 2 eins umgerührt</p> <p>Versuch 3 unterschiedlich große Würfel</p> <p>Warum ist es wichtig, dass in jedem Glas die gleiche Menge Wasser ist?</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>Weil man nicht weiß, ob die Wassermenge nicht auch wichtig für das Auflösen ist.</p> </div> <p>(int. .24/nat. .22)</p> <p>Im zweiten Teil dieser Doppelaufgabe muss erläutert werden, warum die Wassermenge konstant gehalten wird. Dies setzt ein erstes Wissen über die Denk- und Arbeitsweise ‚kontrolliertes Experimentieren‘ voraus. Als richtig werden kontextgebundene Antworten (siehe Beispiel) oder bereits generalisierte Antworten wie „um einen fairen Test zu machen“ gewertet.</p>
Material	fest	flüssig	gasförmig																				
Papier	<input checked="" type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C																				
Sauerstoff	<input type="radio"/> A	<input type="radio"/> B	<input checked="" type="radio"/> C																				
Öl	<input type="radio"/> A	<input checked="" type="radio"/> B	<input type="radio"/> C																				
Kochsalz	<input checked="" type="radio"/> A	<input type="radio"/> B	<input type="radio"/> C																				
Geografie	<p>Unten stehen fünf Merkmale der Oberfläche der Erde.</p> <p>Ziehe einen Strich von jedem Merkmal zu der passenden Beschreibung. Das erste Merkmal wurde schon für dich zugeordnet.</p> <table border="1"> <thead> <tr> <th>Merkmal</th> <th>Beschreibung</th> </tr> </thead> <tbody> <tr> <td>Ebene</td> <td>steil</td> </tr> <tr> <td>Berg</td> <td>flach</td> </tr> <tr> <td>Meer</td> <td>Salzwasser</td> </tr> <tr> <td>Wüste</td> <td>Süßwasser</td> </tr> <tr> <td>Fluss</td> <td>trocken</td> </tr> </tbody> </table> <p>In dieser Aufgabe* sind fünf allgemeine Merkmale der Oberfläche der Erde aufgelistet, zu denen jeweils eine charakterisierende Beschreibung zugeordnet werden muss. Die Aufgabe setzt Kenntnisse über alltagsnahe Fachbegriffe wie Ebene und Süßwasser voraus und erfordert den Abruf von alltagsnahem Wissen über bekannte Merkmale der Oberfläche der Erde.</p>	Merkmal	Beschreibung	Ebene	steil	Berg	flach	Meer	Salzwasser	Wüste	Süßwasser	Fluss	trocken	<p>Die Abbildung zeigt die Regenmenge, die verschiedene Nutzpflanzen brauchen.</p> <p>Regenmenge, die unterschiedliche Nutzpflanzen brauchen</p>  <p>Ein Bauer will Nutzpflanzen in einer Gegend anbauen, in der es pro Jahr etwa 60 cm Regen gibt. Was würde dort am wahrscheinlichsten wachsen?</p> <p><input type="radio"/> A nur Zwiebeln <input checked="" type="radio"/> B Zwiebeln und Erdnüsse <input type="radio"/> C Baumwolle und Zitrusfrüchte <input type="radio"/> D Bananen, Zitrusfrüchte und Baumwolle</p> <p>(int. .54/nat. .58)</p> <p>Aus dem Balkendiagramm muss abgeleitet werden, welche Pflanzen ein Bauer bei einer vorgegebenen Regenmenge am besten anpflanzen sollte. Es müssen Informationen aus dem Diagramm entnommen werden und diese mit der vorgegebenen Regenmenge in Verbindung gesetzt werden.</p>	<p>Ein anderer Bauer wohnt auf einer tropischen Insel im Ozean und will Bananen anbauen. Unten siehst du eine Abbildung der Insel.</p> <p>Jährliche Regenmenge auf der Insel</p>  <p>Legende</p> <ul style="list-style-type: none"> □ 100 cm oder weniger ▨ 101-140 cm ▩ 141-180 cm ■ Mehr als 180 cm <p>Sieh dir die Abbildung der Insel und die Information in der Abbildung aus dem Teil A dieser Frage an.</p> <p>In welchem Gebiet sollte der Bauer Bananen anbauen?</p> <p><input type="radio"/> A Gebiet 1 <input checked="" type="radio"/> B Gebiet 2 <input type="radio"/> C Gebiet 3 <input type="radio"/> D Gebiet 4</p> <p>(int. .47/nat. .46)</p> <p>Diese Aufgabe ist der zweite Teil der links stehenden Aufgabe. Es muss die Information aus dem Balkendiagramm mit den Informationen aus der Karte und dazugehöriger Legende in Beziehung gesetzt und daraus Schlussfolgerungen gezogen werden. Die Aufgabe setzt den Umgang mit Daten in verschiedenen Repräsentationsformen voraus.</p>								
Merkmal	Beschreibung																						
Ebene	steil																						
Berg	flach																						
Meer	Salzwasser																						
Wüste	Süßwasser																						
Fluss	trocken																						

* = Aufgrund der begrenzten Aufgabenfreigabe ist eine Beispielaufgabe aus TIMSS 2015 (Steffensky, Kleickmann, Kasper & Köller, 2016) abgebildet. Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an. Abdruck und Nutzung der Aufgaben nur mit ausdrücklicher Genehmigung der IEA: <https://www.iea.nl/publications/form/iea-permission-request-form>

4.4 TIMSS-Test

In diesem Abschnitt werden der Aufbau des Tests, der zur Erfassung naturwissenschaftlicher Kompetenzen in TIMSS 2019 eingesetzt wurde (Abschnitt 4.4.1), die Ergebnisse aus den Analysen zur curricularen Validität der Testaufgaben (Abschnitt 4.4.2) sowie die Kompetenzstufen, die zur Illustration der Kompetenzausprägungen der Schülerinnen und Schüler genutzt werden (Abschnitt 4.4.3), beschrieben.

4.4.1 Testentwicklung und -aufbau

Die TIMSS-Testaufgaben für die Naturwissenschaften wurden in einem kooperativen Prozess entwickelt, an dem neben dem *TIMSS & PIRLS International Study Center* am *Boston College* und den nationalen Studienleitungen auch Expertinnen und Experten aus der Naturwissenschaftsdidaktik beteiligt waren (siehe Kapitel 2 in diesem Band). Die in einem iterativen Prozess überarbeiteten Aufgaben wurden in großen nationalen Feldtests erprobt, bevor dann die finalen Testaufgaben für die Haupterhebung ausgewählt wurden. Ziel dieses aufwendigen Prozesses war es, Aufgaben zu erhalten, die sowohl in psychometrischer Hinsicht als auch für die Schülerinnen und Schüler des jeweiligen Staates geeignet sind, naturwissenschaftliche Kompetenzen zu erfassen. Zu den Details und Unterschieden zwischen der computer- und papierbasierten Testung in Deutschland siehe Kapitel 5 in diesem Band.

Das Ziel der TIMSS-Rahmenkonzeption, naturwissenschaftliche Kompetenzen in den Inhaltsbereichen *Biologie*, *Geografie* und *Physik/Chemie* sowie den kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* breit abzudecken, machte die Entwicklung einer großen Zahl an Aufgaben erforderlich. Der finale Aufgabenpool im Bereich Naturwissenschaften bei TIMSS 2019 umfasst 169 Aufgaben² für die vierte Klasse. Von diesen sind 87 *Multiple-Choice*-Aufgaben, bei denen die Schülerinnen und Schüler eine oder mehrere Antworten wählen müssen. Die übrigen 82 Testaufgaben sind sogenannte *Constructed-Response-Items*, die eine kurze schriftliche Antwort erfordern oder bei denen die Schülerinnen und Schüler Begriffe oder Bilder via *drag-&-drop* zu- oder anordnen müssen. Für die Trendschätzungen wurden dabei 94 Aufgaben (56%) aus den vorherigen Studienzyklen übernommen. Die verbleibenden 74 Aufgaben (44%) wurden für TIMSS 2019 neu entwickelt.

Wie oben beschrieben lassen sich entsprechend der TIMSS-Rahmenkonzeption die einzelnen Aufgaben jeweils einem Inhaltsbereich und einem kognitiven Anforderungsbereich zuordnen. Dadurch können zusätzlich zur Erfassung der Kompetenz auf der Gesamtskala Naturwissenschaften auch Kompetenzen im Hinblick auf die einzelnen Inhalts- und Anforderungsbereiche untersucht werden. In Tabelle 4.5 ist die Verteilung der Aufgaben auf die Inhaltsbereiche *Biologie*, *Physik/Chemie* und *Geografie* dargestellt. Die drei Bereiche sind nicht gleichmäßig berücksichtigt. Es lassen sich 43 Prozent der Aufgaben dem Inhaltsbereich *Biologie* zuordnen, 36 Prozent dem Bereich *Physik/Chemie* und lediglich 21 Prozent dem Inhaltsbereich *Geografie*. Diese Schwerpunktsetzung auf die *Biologie* und – etwas weniger ausgeprägt – auf die *Physik/Chemie* fand

2 In den Naturwissenschaften wurde für Deutschland zusätzlich ein Item (aus den Bereichen Biologie und Problemlösen) aufgrund ungenügender Passung von englischsprachigem Original und deutscher Übersetzung aus der Skalierung ausgeschlossen.

sich auch in den vorangegangenen Studienzyklen wieder. Sie basiert auf der Rahmenkonzeption von TIMSS, die sich an den Curricula und Standards der teilnehmenden Staaten orientiert (Mullis, Martin, Goh & Cotter, 2016). In diesem Themenkanon sind geografische Inhalte nicht so stark repräsentiert wie biologische oder physikalisch-chemische. Analysen von Schulbüchern aus Deutschland deuten darauf hin, dass im Sachunterricht geografische Inhalte vergleichsweise stark repräsentiert sind (Blaseio, 2009). Dafür sprechen auch die Befunde der Lehrkräftebefragung (siehe Tabelle 4.1). Allerdings legen viele der Lehrpläne einen Schwerpunkt auf regionalgeografische und weniger auf naturgeografische Themen.

Tabelle 4.5: Verteilung der Testaufgaben auf die naturwissenschaftlichen Inhaltsbereiche

Naturwissenschaftliche Inhaltsbereiche	2007		2011		2015		2019	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Biologie	74	43	78	45	74	44	73	43
Physik/Chemie	64	37	63	36	61	36	61	36
Geografie	36	21	34	20	33	20	35	21
Gesamt	174	100	175	100	168	100	169	100

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

In Tabelle 4.6 ist die Verteilung der Aufgaben auf die kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen* dargestellt. Wie in den vergangenen Studienzyklen wird bei circa 40 Prozent der Aufgaben Faktenwissen abgefragt, während circa 60 Prozent der Aufgaben kognitiv anspruchsvollere Anforderungen beinhalten. Trotz dieser angenommenen Hierarchie in den kognitiven Anforderungen bedeutet dies nicht, dass die Aufgaben im Anforderungsbereich *Reproduzieren* immer leichter sind als die in den anderen Anforderungsbereichen.

Tabelle 4.6: Verteilung der Testaufgaben auf die kognitiven Anforderungsbereiche

Kognitive Anforderungsbereiche	2007		2011		2015		2019	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Reproduzieren	77	44	72	41	67	40	69	41
Anwenden	63	36	71	41	66	39	64	38
Problemlösen	34	20	32	18	35	21	36	21
Gesamt	174	100	175	100	168	100	169	100

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

4.4.2 Curriculare Validität

Der Themenkanon, der den TIMSS-Aufgaben zugrunde liegt, und auch die damit einhergehenden inhaltlichen Schwerpunktsetzungen orientieren sich an den Curricula und Standards der Staaten, die an TIMSS teilnehmen. Ähnlich wie in der Mathematik gibt es in den Naturwissenschaften einen relativ großen nationalen wie internationalen Konsens über zentrale Konzepte und Themengebiete wie Kennzeichen von Lebewesen, Struktur und Funktion, Ökosysteme, Stoffe und ihre Eigenschaften, Energie, Kräfte oder Klima, Landschaft und ihre Merkmale, die sich zum Beispiel an den *Basiskonzepten* (KMK, 2004), *Big Ideas* (Harlen, 2015) oder *Core Concepts* und *Crosscutting Concepts* (National Research Council, 2012) ablesen lassen. Gleichwohl ist nicht jede einzelne Facette eines Themenbereiches in allen Curricula der teilnehmenden Staaten abgebildet. Unterschiede im Umfang und in Gewichtungen ergeben sich zum Teil auch aus der unterschiedlichen Organisation des naturwissenschaftlichen Unterrichts in der Grundschule. So gibt es in den meisten Staaten Naturwissenschaften als ein Fach, das nicht wie im Sachunterricht noch andere Bereiche integriert.

Zur Überprüfung der Passung wurden die in den Aufgaben thematisierten Inhalte mit den curricularen Vorgaben der 16 Länder der Bundesrepublik Deutschland abgeglichen. Eine Schwierigkeit für die Ermittlung der sogenannten curricularen Validität stellen dabei die unterschiedlichen curricularen Vorgaben für den Sachunterricht in Deutschland dar, denn anders als im Fach Mathematik gibt es für das Fach Sachunterricht keine länderübergreifenden Bildungsstandards. Aus diesen Gründen wurde für die Analyse der curricularen Validität in erster Linie der Lehrplan aus Nordrhein-Westfalen herangezogen (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008).

Nordrhein-Westfalen ist das bevölkerungsreichste Land der Bundesrepublik Deutschland und dementsprechend machen Schülerinnen und Schüler aus diesem Land der Bundesrepublik Deutschland einen erheblichen Teil der TIMSS-Stichprobe aus, sodass eine Orientierung an diesen Vorgaben sinnvoll erscheint. Zudem ist der Lehrplan vergleichbar mit denen der anderen Länder in Deutschland. Zusätzlich wurden der (länderübergreifende) Perspektivrahmen Sachunterricht (GDSU, 2013) sowie die für den Erhebungsraum aktuellen Lehrpläne der Länder Bayern (Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2014), Baden-Württemberg (Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2016), Niedersachsen (Niedersächsisches Kultusministerium, 2017) und Sachsen³ (Sächsisches Staatsministerium für Kultus, 2004/2009) berücksichtigt. Der Lehrplan aus Nordrhein-Westfalen wie auch die Lehrpläne der anderen Länder berücksichtigen neben naturwissenschaftlichen Inhalten auch naturwissenschaftliche Denk- und Arbeitsweisen als bedeutsame Zielbereiche. In den konkreten Kompetenzerwartungen für das Ende von Klasse zwei und das Ende von Klasse vier, die in der Regel in den Lehrplänen formuliert sind, werden Inhalte und (fachspezifische) Denk- und Arbeitsweisen integriert formuliert. Zentrale Kriterien für die curriculare Passung waren die angesprochenen Inhaltsgebiete und Konzepte sowie die verwendeten Fachbegriffe, die sich in den verbindlichen Inhalten und den Kompetenzerwartungen wiederfinden. Dieses Vorgehen wurde auch in den bisherigen Studienzyklen genutzt.

Die Curriculumvalidität, also die Frage, inwieweit die Inhalte der TIMSS-Aufgaben auch tatsächlich im Unterricht behandelt wurden, lässt sich mit die-

³ Der aktuelle Bildungsplan trat erst am 1.8.2019 in Kraft und wurde deswegen noch nicht berücksichtigt.

ser Analyse nicht einschätzen. So ist generell unklar, welche Steuerungsfunktion die curricularen Vorgaben haben. Im mehrperspektivischen Sachunterricht kommt hinzu, dass Lehrpersonen möglicherweise auch individuelle Schwerpunkte auf einzelne Perspektiven in ihrem Unterricht setzen.

In Tabelle 4.7 sind die Anteile der curricular nicht validen Testaufgaben aufgeschlüsselt nach den drei Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geografie*. Insgesamt sind nur 7 Prozent der Aufgaben als curricular nicht valide eingeschätzt worden. Diese verteilen sich relativ gleichmäßig auf die drei Inhaltsbereiche mit 6 bis 8 Prozent.

Tabelle 4.7: Anteile curricular nicht valider Testaufgaben in Naturwissenschaften nach Inhaltsbereichen

Naturwissen- schaftliche Inhaltsbereiche	2007			2011			2015			2019		
	Gesamt	nicht valide		Gesamt	nicht valide		Gesamt	nicht valide		Gesamt	nicht valide	
		n	%		n	%		n	%		n	%
Biologie	74	12	16	78	21	27	74	6	8	73	6	8
Physik/Chemie	64	3	5	63	5	8	61	5	8	61	4	7
Geografie	36	15	42	34	14	42	33	5	15	35	2	6
Gesamt	174	30	17	175	40	23	168	16	10	169	12	7

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Die sechs Aufgaben aus dem Inhaltsbereich der *Biologie* lassen sich verschiedenen Teilgebieten (siehe Tabelle 4.3) zuordnen. In erster Linie sind es Aufgaben, in denen Beispiele für Lebensvorgänge und die Anpasstheit von Lebewesen an einen Lebensraum thematisiert werden, die für viele Schülerinnen und Schüler aus Deutschland nicht unbedingt vertraut sind, zum Beispiel Nahrungsaufnahme bei Schlangen, Wasserspeicherung von Kakteen. Weiterhin sind Aufgaben als curricular nicht valide eingeschätzt, in denen es um Fotosynthese geht. Im Inhaltsbereich *Physik/Chemie* stammen die curricular nicht validen Aufgaben überwiegend aus dem Teilgebiet *Kräfte und Bewegungen*, ein Bereich, der im Sachunterricht in der Regel unter der technischen Perspektive, beispielsweise im Hinblick auf einfache Kraftwandler, auf Stabilität beim Bauen, auf der Phänomenebene thematisiert wird. Fachbegriffe wie Erdanziehungskraft oder auch die unterschiedlichen Richtungen von Kräften und deren detailliertes Zusammenspiel werden nur in Ausnahmen erarbeitet. Im Inhaltsgebiet *Geografie* sind ein Teil der Aufgaben als curricular nicht valide eingeschätzt, die dem Teilgebiet *die Erde im Sonnensystem* zugeordnet sind. Dieses Thema ist so nicht vorgesehen, auch wenn man durchaus didaktische Materialien dazu findet und es im (alten) Lehrplan in Sachsen als Wahlpflichtthema aufgeführt wird. Eine Facette dieses Themenbereichs *Auswirkungen der Bewegungen der Erde*, z. B. *Tag und Nacht*, *Jahreszeiten* findet sich aber in der naturwissenschaftlichen Perspektive, wenn es um die Veränderungen in der Natur geht. So lautet eine Kompetenzerwartung im nordrhein-westfälischen Lehrplan: „Die Schülerinnen und Schüler beschreiben Veränderungen in der Natur und stellen Entwicklungsphasen dar (z. B. *Wasserkreislauf*, *Jahreszeiten*)“ (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008, S. 43). An dieser Kompetenzerwartung erkennt man auch die gewisse Schwierigkeit bei der

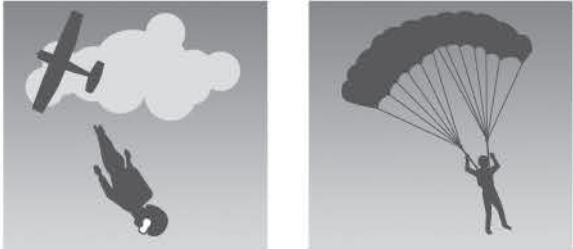
Einschätzung der Passung von Aufgaben und curricularen Vorgaben, da es unklar bleibt, welches Niveau hier angestrebt wird.

Die meisten der Aufgaben, die als nicht curricular valide eingeschätzt wurden, stammen aus dem Anforderungsbereich *Reproduzieren*. In vielen Fällen ist dies darauf zurückzuführen, dass die Schülerinnen und Schüler zur Lösung der Aufgaben über Faktenwissen verfügen müssen, das im Lehrplan Nordrhein-Westfalens beziehungsweise den weiteren Dokumenten nicht vorgesehen ist, zum Beispiel Wissen über spezifische Tiere und Pflanzen oder spezifische Fachbegriffe.

In Abbildung 4.2 ist ein Beispiel einer Aufgabe dargestellt, die als nicht curricular valide eingeschätzt wurde. Hier müssen die Schülerinnen und Schüler die Kraft identifizieren, die den Flug des Fallschirms erklärt. Auch wenn es sicher viele Schülerinnen und Schüler gibt, die den Begriff Erdanziehungskraft aus dem Alltag kennen, ist er kein regulärer Inhalt der Lehrpläne der Grundschule.

Abbildung 4.2: Beispiel für eine curricular nicht valide Aufgabe

Ein Fallschirmspringer springt aus einem Flugzeug und gleitet mit einem Fallschirm zur Erde.



Woran liegt es, dass der Fallschirmspringer Richtung Erde fällt?

- ☐ A an der Luft um die Erde
- ☐ B an der Magnetkraft der Erde
- ☒ C an der Erdanziehungskraft
- ☐ D an der Erddrehung

(int. .67/nat. .62) SE71140

Aufgabenbeispiel zu Physik/Chemie
Teilgebiet: Kräfte und Bewegungen

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

Abdruck und Nutzung der Aufgaben nur mit ausdrücklicher Genehmigung der IEA: <https://www.iea.nl/publications/form/iea-permission-request-form>

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Zusammenfassend kann festgehalten werden, dass der Anteil der in TIMSS 2019 eingesetzten naturwissenschaftlichen Aufgaben, die als curricular valide eingeschätzt wurden, mit circa 90 Prozent sehr hoch ist. Der Anspruch der TIMSS-Rahmenkonzeption, Kompetenzen zu erfassen, die in der Schule aufgrund von curricularen Vorgaben vermittelt werden, wurde also in Deutschland weitestgehend erfüllt.

Im Vergleich zu TIMSS 2015 fanden sich für die Inhaltsbereiche *Biologie* und *Physik/Chemie* ähnliche Werte (siehe Tabelle 4.7), lediglich im Bereich *Geografie* wurden zwei Aufgaben mehr als curricular nicht valide eingeschätzt. Der gesamte Anteil der nicht validen Aufgaben ist im Vergleich zu TIMSS 2015 ungefähr gleich geblieben. In den Studienzyklen davor war er etwas größer. Das hängt vermutlich mit dem leicht höheren Anteil an Aufgaben aus dem Anforderungsbereich *Reproduzieren* in TIMSS 2007 und TIMSS 2011 zusammen. Zudem haben möglicherweise die Weiterentwicklungen der curricularen Vorgaben in Deutschland in den letzten zehn Jahren sowie die Anpassungen der TIMSS-Rahmenkonzeptionen an die Curricula der verschiedenen Teilnehmerstaaten hier einen Einfluss.

Um zu prüfen, ob die – wenn auch geringen – Einschränkungen hinsichtlich der curricularen Validität, einen Einfluss auf die Ergebnisse der Schülerinnen und Schüler im Bereich Naturwissenschaften in Deutschland haben, wurden bei TIMSS 2011 und TIMSS 2015 auch Analysen durchgeführt, bei denen die als nicht curricular valide eingeschätzten Aufgaben ausgeschlossen wurden. In TIMSS 2015, wo 10 Prozent der Aufgaben als nicht curricular valide eingeschätzt wurden, zeigten sich kaum Unterschiede in den Ergebnissen zwischen Analysen mit allen und Analysen mit lediglich den curricular validen Aufgaben. So erhöhte sich in den Analysen mit ausschließlich den Aufgaben, die als curricular valide eingeschätzt wurden, der Mittelwert der Schülerinnen und Schüler aus Deutschland um 3 Punkte und auch der Rangplatz von Deutschland verschob sich um drei Plätze nach oben (Steffensky et al., 2016). Auch in TIMSS 2011 zeigen sich nur geringfügige Unterschiede von 2 Punkten in den mittleren Leistungen beziehungsweise drei Rangplätzen beim Vergleich der Ergebnisse mit allen Aufgaben und denen auf der Grundlage der als curricular valide eingeschätzten Aufgaben (Kleickmann et al., 2012), obwohl der Anteil der nicht curricular validen Aufgaben mit 23 Prozent deutlich höher eingeschätzt wurde. Angesichts der geringen Unterschiede in den vorangegangenen Zyklen wurde in TIMSS 2019 darauf verzichtet, die relativ aufwendigen Neuskalierungen und Analysen national nochmals durchzuführen.

4.4.3 Kompetenzstufen in den Naturwissenschaften

Bei TIMSS 2019 wird wie in den vorherigen Studienzyklen die Verteilung der Kompetenzwerte auf fünf Kompetenzstufen analysiert, die die inhaltlichen und kognitiven Leistungsniveaus der Schülerinnen und Schüler widerspiegeln (für Details zur Skalierungsmethode siehe Wendt, Bos, Kasper et al., 2016). Die Analyse der inhaltlichen und kognitiven Anforderungen, die zur erfolgreichen Bearbeitung der Aufgaben einer Kompetenzstufe erforderlich sind, lassen eine stärker inhaltliche Beschreibung und Interpretation der erreichten Kompetenzwerte zu.

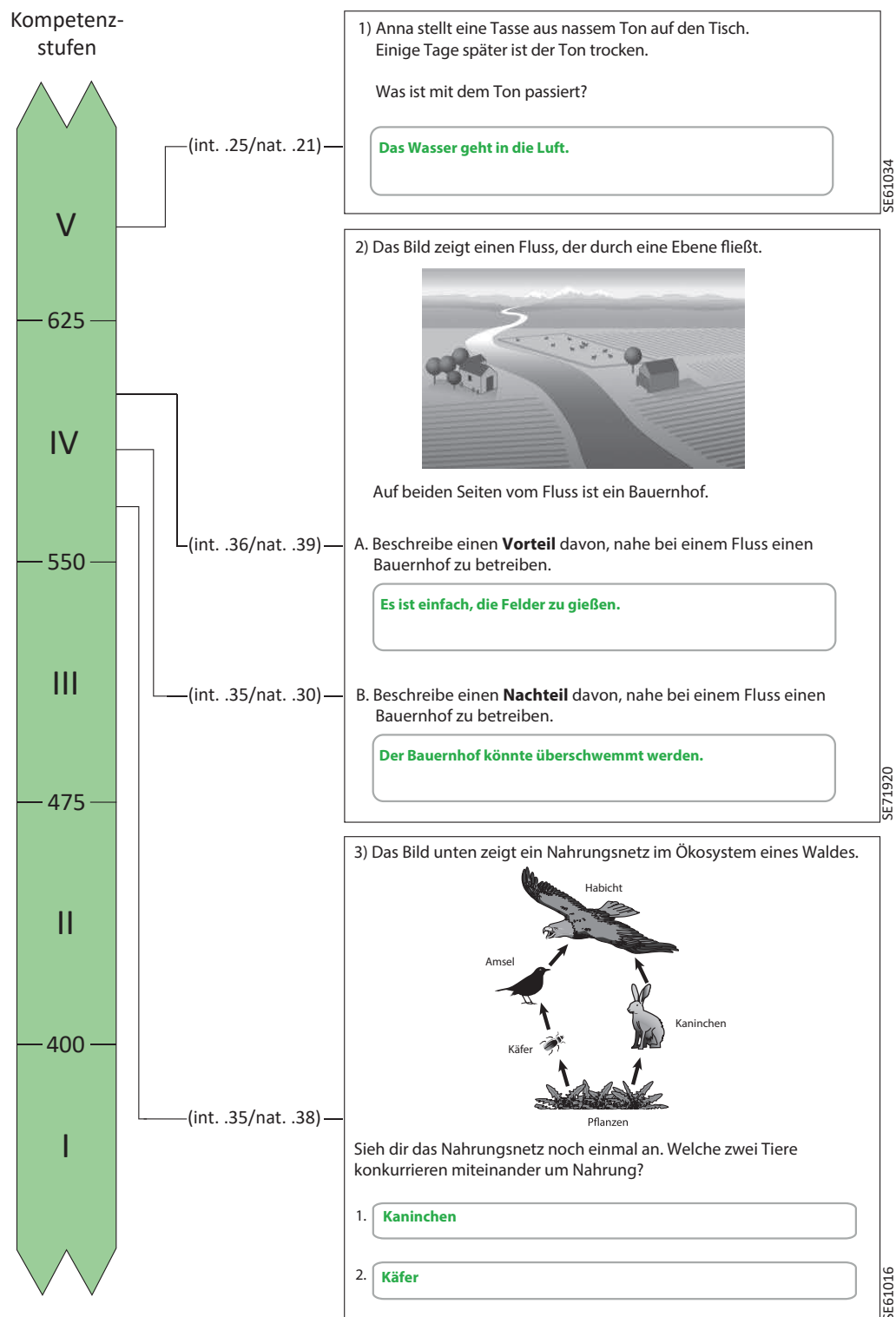
Wie in Mathematik gibt es vier Bezugspunkte, sogenannte *Benchmarks*, die die Kompetenzskala in fünf Abschnitte einteilt, die den fünf Kompetenzstufen entsprechen. Diese werden von der internationalen TIMSS-Studienleitung festgelegt und sind identisch zu den *Benchmarks* der bisherigen Studienzyklen, sodass Vergleiche über die Verteilung auf die Kompetenzstufen über die Zeit möglich sind. In Tabelle 4.8 ist beschrieben, welches Wissen und Verständnis für die fünf Kompetenzstufen charakteristisch ist. Die Kompetenzstufen sind hierarchisch geordnet, das heißt Schülerinnen und Schüler, die sich beispielsweise auf Kompetenzstufe IV befinden, lösen Testaufgaben, die typisch für die-

se Stufe sind, mit einer höheren Wahrscheinlichkeit als Testaufgaben der Stufe V. Leistungsschwache Schülerinnen und Schüler befinden sich auf den unteren Kompetenzstufen I (< 400 Punkte) und II (400–474 Punkte). Sie verfügen lediglich über ein rudimentäres Wissen einzelner isolierter Fakten in den Naturwissenschaften. Schülerinnen und Schüler in der mittleren Kompetenzstufe III (475–549 Punkte) müssen ihr Basiswissen nicht nur reproduzieren, sondern in einfachen Kontexten anwenden können. Auf den obersten Kompetenzstufen IV (550–624 Punkte) und V (≥ 625 Punkte) befinden sich leistungsstarke Schülerinnen und Schüler. Sie können Aufgaben mit mehreren Zwischenschritten lösen, Wissen in komplexeren Kontexten sowie naturwissenschaftliche Denk- und Arbeitsweisen anwenden.

Tabelle 4.8: Beschreibung der fünf Kompetenzstufen für die Gesamtskala Naturwissenschaften

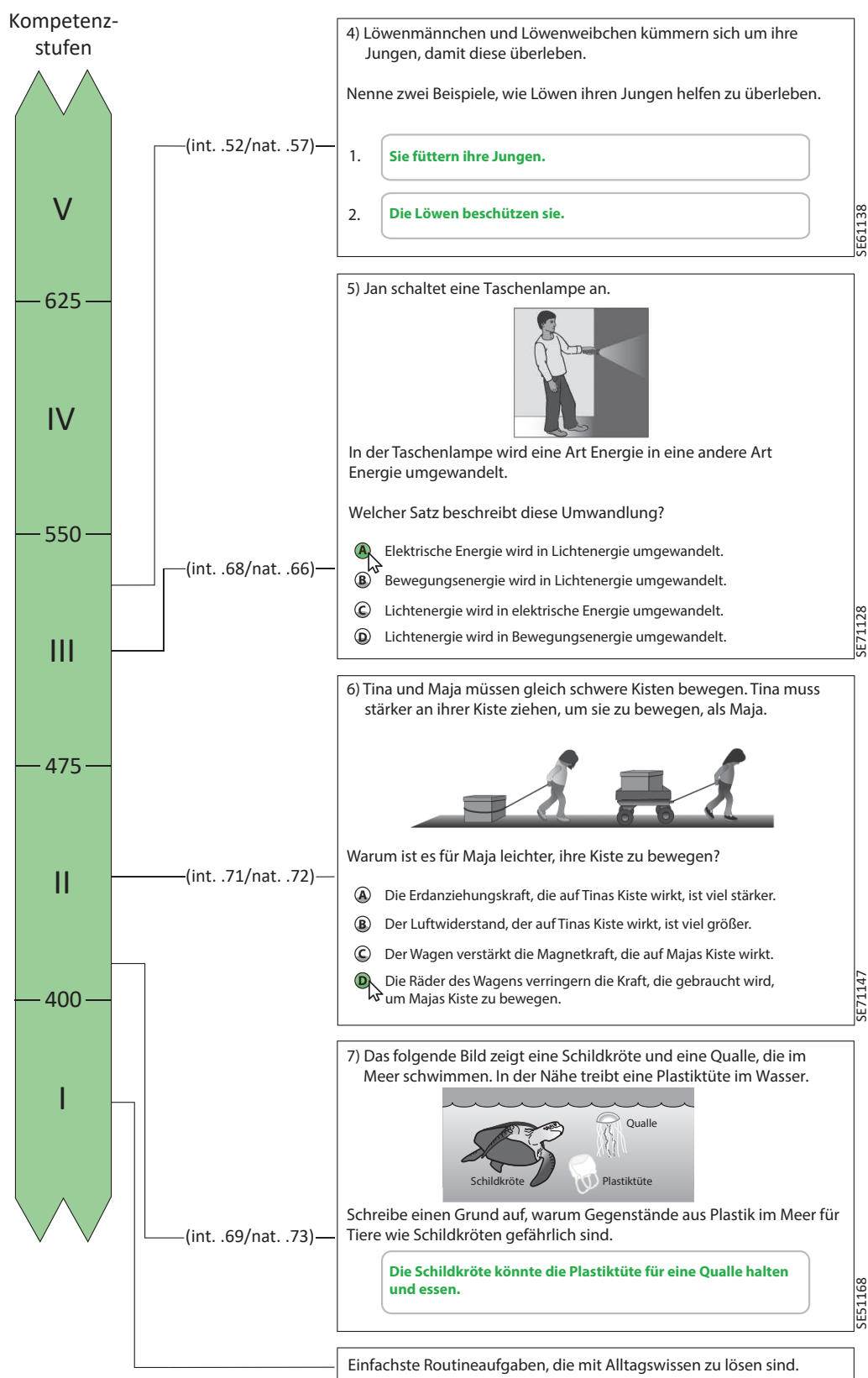
Kompetenzstufe V (ab 625): fortgeschritten
<p><i>Die Schülerinnen und Schüler verfügen über ein grundlegendes Verständnis in den drei Inhaltsbereichen Biologie, Physik/Chemie und Geografie sowie ein Verständnis der naturwissenschaftlichen Denk- und Arbeitsweisen.</i></p> <p>Sie sind in der Lage, ihr Wissen über Merkmale und Lebensvorgänge einer Vielzahl von Lebewesen, Beziehungen in Ökosystemen und das Interagieren von Organismen mit Lebensräumen darzustellen. Sie zeigen weiterhin ein grundlegendes Verständnis von Stoffeigenschaften, den Aggregatzuständen sowie physikalischen und chemischen Veränderungen von Stoffen. Sie können dabei auch einschätzen, inwiefern Vorgehensweisen geeignet sind, um Stoffe zu trennen. Auch im Bereich der Geografie verfügen sie über ein grundlegendes Verständnis von Landschaftsmerkmalen, abgelaufenen und laufenden naturgeografischen Prozessen sowie die Erdrevolution und -rotation. Die Schülerinnen und Schüler weisen zudem elementare Fähigkeiten im Bereich der naturwissenschaftlichen Denk- und Arbeitsweisen auf. So können sie einfache kontrollierte Experimente planen und Schlüsse aus Beschreibungen und Diagrammen ziehen.</p>
Kompetenzstufe IV (550–624): hoch
<p><i>Die Schülerinnen und Schüler verfügen über Wissen in den drei Inhaltsbereichen Biologie, Physik/Chemie und Geografie und können dieses anwenden.</i></p> <p>Sie verfügen über Wissen über Merkmale und Lebenszyklen von Pflanzen und Tieren, Ökosystemen und Interaktionen zwischen Lebewesen und ihrer Umwelt und können dieses auch anwenden. Weiterhin zeigen sie Wissen über die Aggregatzustände, Eigenschaften von Stoffen und Energietransfer in praktischen Kontexten und können dies auch anwenden. Zudem weisen sie ein erstes Verständnis über Kräfte und Bewegungen auf. Die Schülerinnen und Schüler können ihr Wissen über die Struktur, Vorgänge und Ressourcen der Erde und Erdgeschichte anwenden und sie verfügen über grundlegendes Wissen über das Sonnensystem. Sie zeigen elementare Kenntnisse und Fähigkeiten bezüglich naturwissenschaftlicher Denk- und Arbeitsweisen. So können sie einfache Schlussfolgerungen aus Modellen und Diagrammen ziehen.</p>
Kompetenzstufe III (475–549): durchschnittlich
<p><i>Die Schülerinnen und Schüler verfügen über ein Basiswissen in den drei Inhaltsbereichen Biologie, Physik/Chemie und Geografie.</i></p> <p>Sie haben ein Basiswissen über die Entwicklung von Pflanzen und Tieren, z.B. was Pflanzen zum Wachsen brauchen, sowie über ausgewählte Eigenschaften von Materialien. Sie kennen ausgewählte Fakten zum Thema Strom, z.B. den Aufbau eines Stromkreises, und können grundlegendes Wissen im Bereich Kräfte und Bewegungen anwenden. Hinsichtlich der Charakteristika der Erde sowie spezifischer Landschaftsmerkmale zeigen sie ein erstes Verständnis.</p>
Kompetenzstufe II (400–474): niedrig
<p><i>Die Schülerinnen und Schüler verfügen über ein Basiswissen in den Bereichen Biologie und Physik/Chemie.</i></p> <p>Sie identifizieren Tiere mit einer Wirbelsäule, Materialien, die Wärme gut leiten und erkennen natürliche Ressourcen wie Wasser und Boden.</p>
Kompetenzstufe I (unter 400): rudimentär
<p><i>Die Schülerinnen und Schüler verfügen über rudimentäres schulisches Anfangswissen.</i></p> <p>Selbst einfache Aufgaben werden nur gelegentlich oder ansatzweise gelöst.</p>

Die folgenden Abbildungen 4.3 und 4.4 zeigen Beispiele für Aufgaben, um die Kompetenzstufen zu illustrieren. Die in Klammern gezeigten Werte entsprechen den internationalen und nationalen Lösungshäufigkeiten. Beispielsweise wurde die erste Aufgabe von 25 Prozent der Schülerinnen und Schüler im internationalen Durchschnitt und von 21 Prozent der Schülerinnen und Schüler in

Abbildung 4.3: Kompetenzstufen und Beispielaufgaben (Gesamtskala Naturwissenschaften) I

Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an.
 Abdruck und Nutzung der Aufgaben nur mit ausdrücklicher Genehmigung der IEA: <https://www.iea.nl/publications/form/iea-permission-request-form>

Abbildung 4.4: Kompetenzstufen und Beispielaufgaben (Gesamtskala Naturwissenschaften) II



Die Werte in Klammern geben die relativen internationalen und nationalen Lösungshäufigkeiten an.
Abdruck und Nutzung der Aufgaben nur mit ausdrücklicher Genehmigung der IEA: <http://www.iea.nl/permissions.html>

Deutschland gelöst. Die Zuordnung einer Aufgabe zu den Kompetenzstufen erfolgt anhand der internationalen Lösungshäufigkeiten. Im Folgenden werden für die sieben Aufgaben die Kompetenzen beschrieben, die Schülerinnen und Schüler zur Lösung der Aufgaben benötigen.

Beispiele für Kompetenzstufe V (fortgeschritten)

In der Aufgabe 1 sollen die Schülerinnen und Schüler erläutern, was beim Trocknen von Ton passiert. Das Beispiel ist zwar (relativ) alltagsnah, aber es ist in der Regel kein typisches Unterrichtsbeispiel, an dem der Prozess der Verdunstung thematisiert wird. Dieser wird häufig am Beispiel des Wasserkreislaufs erarbeitet. Die Schülerinnen und Schüler müssen also ihr Wissen über die Verdunstung von Wasser auf ein (in der Regel) neues Beispiel anwenden. Sie können dabei den Fachbegriff Verdunsten nutzen oder den Vorgang umschreiben („Das Wasser geht in die Luft“).

Beispiel für Kompetenzstufe IV (hoch)

Aufgabe 2 stammt aus dem Inhaltsbereich *Geografie*. Die Schülerinnen und Schüler können im ersten Teil verschiedene Vorteile anführen, die auf die Bedeutung von Wasser für Tiere und Pflanzen verweisen oder aber auch auf die Möglichkeit, das bewegte Wasser mithilfe eines Wasserrads zur Stromerzeugung zu nutzen. Dabei wenden sie naturwissenschaftliches Wissen an. Als richtig werden auch Antworten gewertet, die auf die Qualität des Bodens nahe dem Wasser eingehen. Im zweiten Teil der Aufgabe soll ein Nachteil genannt werden. Mögliche Nachteile sind die Überflutungsgefahr, aber auch die Gefahr, dass das Wasser aus dem Fluss mit Schadstoffen belastet sein könnte und dann auf die Felder oder zu den Tieren gelangt oder umgekehrt der Dung der Tiere in das Wasser gelangt und somit den Fluss belastet. Dieser Teil der Aufgabe fällt den Schülerinnen und Schülern aus Deutschland etwas schwerer, möglicherweise, weil diese Aspekte im geografiebezogenen Sachunterricht seltener im Mittelpunkt stehen.

In der Aufgabe 3 ist ein einfaches Nahrungsnetz dargestellt. Aus der Abbildung müssen die Schülerinnen und Schüler schlussfolgern, welche beiden Tiere um Nahrung konkurrieren. Die Aufgabe erfordert ein erstes Verständnis über Nahrungsbeziehungen in einem Ökosystem (hier Wald).

Beispiele für Kompetenzstufe III (durchschnittlich)

Aufgabe 4 ist eine eher schwierige Aufgabe aus dem Anforderungsbereich *Reproduzieren*. Die Schülerinnen und Schüler müssen zwei unterschiedliche Beispiele nennen, wie Löwen Brutpflege betreiben. Als richtig werden Antworten gewertet, in denen auf das Füttern oder den Schutz vor Feinden verwiesen wird oder wie man jagt oder Wasserstellen findet. Auch hier sind konkrete Antworten („beschützen vor Hyänen“) oder generalisierte Antworten möglich („Die Löwen beschützen sie“).

Auch Aufgabe 5 ist eine eher schwierige Aufgabe aus dem Anforderungsbereich *Reproduzieren*. Hier müssen die Schülerinnen und Schüler Wissen über die Umwandlung verschiedener Energieformen in einem alltagsnahen Kontext (Taschenlampe) haben.

Beispiele für Kompetenzstufe II (niedrig)

In Aufgabe 6 müssen die Schülerinnen und Schüler aus vorgegebenen Antworten die wählen, die erklärt, warum sich ein schwerer Gegenstand leichter mit einem

Karren ziehen lässt. Selbst wenn die Schülerinnen und Schüler noch keine differenziertere Vorstellung von Kräften haben, können sie hier Alltagserfahrungen anwenden, um die Aufgabe zu lösen.

In der Aufgabe 7 müssen die Schülerinnen und Schüler einen Grund angeben, warum Gegenstände aus Plastik für Meerestiere wie eine Schildkröte gefährlich sind. Als richtig werden Antworten gewertet, die auf die Gefahr des Erstickens, des Fressens oder der eingeschränkten Bewegungsfreiheit verweisen. Bei dieser Aufgabe fällt auf, dass die Schülerinnen und Schüler in Deutschland einen etwas höheren Mittelwert haben im Vergleich zum internationalen Mittelwert. Dies liegt vermutlich daran, dass das Thema Plastik im Meer in Deutschland in den letzten Jahren viel Aufmerksamkeit erfahren hat und zum Beispiel im Zusammenhang mit Bildung für nachhaltige Entwicklung auch im Sachunterricht häufig thematisiert wurde.

4.5 Ergebnisse zu den naturwissenschaftlichen Leistungen in Deutschland am Ende der vierten Jahrgangsstufe

Die folgenden Ergebnisse gliedern sich nach den in der Einleitung aufgeführten übergeordneten Fragestellungen. Die hier dargestellten Daten beruhen auf den international berichteten originalen Schätzungen in der computerbasierten Stichprobe (siehe Kapitel 5 in diesem Band). Im Abschnitt 4.5.1 werden zunächst die Ergebnisse der Gesamtskala Naturwissenschaften im internationalen Vergleich dargestellt. Anschließend wird die Verteilung auf die fünf Kompetenzstufen berichtet, wobei der Schwerpunkt hier auf den leistungsschwachen und den leistungsstarken Schülerinnen und Schülern liegt (Abschnitt 4.5.2). Die Ergebnisse zu den drei Inhaltsbereichen und den drei kognitiven Anforderungsbereichen werden in den Abschnitten 4.5.3 beziehungsweise 4.5.4 skizziert. Im Abschnitt 4.5.5 werden abschließend die Einstellungen der Schülerinnen und Schüler zum Sachunterricht berichtet. In allen Abschnitten ist der Vergleich mit den entsprechenden Daten aus den vorangegangenen Studienzyklen integriert.

Im Jahr 2019 nahmen 58 Staaten an der TIMS-Studie teil, wovon 29 OECD- und 24 EU-Staaten sind. An der Studie waren auch drei Staaten beteiligt, die mit Schülerinnen und Schülern aus der Jahrgangsstufe 5 teilnahmen, sowie sechs sogenannte *Benchmark*-Teilnehmer, bei denen es sich um bestimmte Regionen aus Kanada, der Russischen Föderation, Spanien sowie den Vereinigten Arabischen Emiraten handelt.

Zur Einordnung der Ergebnisse zur naturwissenschaftlichen Kompetenz der Schülerinnen und Schüler in Deutschland werden im Folgenden die Ergebnisse aus Deutschland mit einzelnen anderen Staaten sowie mit drei *Vergleichsgruppen* (VG) verglichen. Die erste Vergleichsgruppe umfasst die Mittelwerte aller Teilnehmerstaaten (internationaler Mittelwert, ohne *Benchmark*-Teilnehmer), die zweite Vergleichsgruppe die teilnehmenden OECD-Staaten (VG_{OECD}) und die dritte Vergleichsgruppe die teilnehmenden EU-Staaten (VG_{EU}). In den Trendanalysen wird auf diese Vergleichsgruppen verzichtet, da sie in den vier Studienzyklen unterschiedlich zusammengesetzt waren. Um eine internationale Vergleichbarkeit der Ergebnisse insgesamt zu gewährleisten, ist die Einhaltung strenger methodischer Standards, die von der internationalen Studienleitung vorgegeben werden, grundlegend. Nationale Untersuchungspopulationen und -stichproben, die spezifische Besonderheiten aufweisen, sind in den Abbildungen dieses Abschnitts durch die Formatierung der Bezeichnungen der Teilnehmerstaaten so-

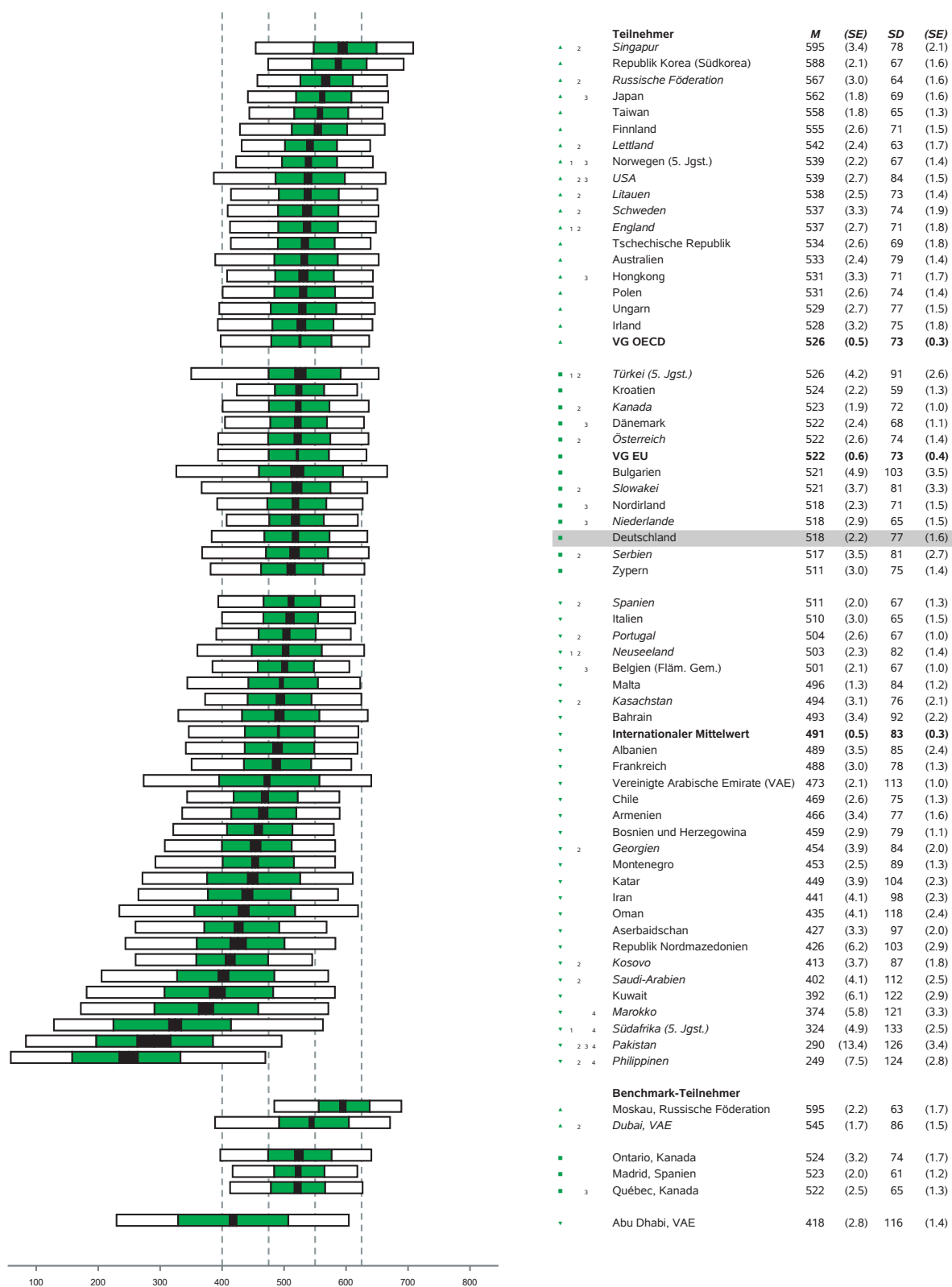
wie die beigegebenen Fußnoten gekennzeichnet (für eine ausführliche Erläuterung der Bedeutung der einzelnen Fußnoten siehe Kapitel 2 und Anhang in diesem Band). Teilnehmerstaaten, die aus unterschiedlichen Gründen die methodischen Standards nicht eingehalten haben, sind kursiv gesetzt. Dies trifft auf 25 der 64 Teilnehmerstaaten und -regionen zu. Für die anderen 39 Teilnehmerstaaten, unter anderem Deutschland, ist eine internationale Vergleichbarkeit der Ergebnisse uneingeschränkt gegeben. Dagegen sind Schlüsse auf die jeweilige Population aller Viertklässlerinnen und Viertklässler bei den kursiv gesetzten Teilnehmerstaaten mit teils erheblicher Unsicherheit verbunden. Auch für den Vergleich der Ergebnisse Deutschlands mit kursiv gesetzten Teilnehmerstaaten und Regionen, sind unter Berücksichtigung der Tabellen im Anhang Besonderheiten der Stichproben zu prüfen, inwiefern ein differentes Abschneiden durch Unterschiede in den Teilnahmebedingungen bedingt sein könnte.

4.5.1 Kompetenzen im internationalen Vergleich

Im Folgenden wird auf die Ergebnisse der Gesamtskala Naturwissenschaften im internationalen Vergleich eingegangen. Dabei werden die Leistungsmittelwerte im internationalen Vergleich, die Streuung der Werte sowie der Vergleich mit TIMSS 2007, 2011 und 2015 dargestellt.

Unterschiede im Niveau der naturwissenschaftlichen Kompetenz im internationalen Vergleich: In Abbildung 4.5 sind die Mittelwerte (M) und Standardabweichungen (SD) mit ihren jeweiligen Standardfehlern (SE) für die naturwissenschaftliche Kompetenz für jeden teilnehmenden Staat und *Benchmark-Teilnehmer* angegeben. Die Standardabweichung ist ein Maß der Streuung der Werte innerhalb eines Staates, also der Homogenität der Kompetenzverteilung. Je größer die Standardabweichung, desto heterogener ist die Kompetenzverteilung. Die Standardfehler geben Auskunft über die Präzision der Mittelwerts- und Standardabweichungsschätzung. Je kleiner die Standardfehler sind, desto präziser ist diese Auskunft. Zusätzlich zu den Mittelwerten einzelner Staaten sind die Mittelwerte der oben beschriebenen Vergleichsgruppen (alle Teilnehmerstaaten, OECD-Staaten und EU-Staaten) angegeben. Die Staaten sind entsprechend ihres Abschneidens in TIMSS 2019 angeordnet.

Die höchsten mittleren Kompetenzwerte zeigen Schülerinnen und Schüler aus Singapur (595 Punkte) und der Republik Korea (588 Punkte). Diesen Staaten folgen mit etwas Abstand die Russische Föderation, Japan und Taiwan mit 567, 562 beziehungsweise 558 Punkten. Der internationale Mittelwert liegt bei 491 Punkten, 38 der Teilnehmerstaaten erreichen Kompetenzwerte über dem Mittelwert, 20 Staaten liegen unter dem Mittelwert. Deutlich schwächer mit Mittelwerten unter 350 Punkten schneiden Südafrika (Jahrgangsstufe 5, 324 Punkte), Pakistan (290 Punkte) und die Philippinen (249 Punkte) ab. In Deutschland liegt der mittlere Kompetenzwert der Schülerinnen und Schüler bei 518 Punkten, also 27 Punkte über dem internationalen Mittelwert. Dieser Wert unterscheidet sich nicht signifikant vom Mittelwert der EU-Staaten (522 Punkte). Staaten, die vergleichbar mit Deutschland abschneiden, in denen sich die Mittelwerte also statistisch nicht unterscheiden, sind die Türkei, Kroatien, Kanada, Dänemark, Österreich, Bulgarien, die Slowakei, Nordirland, die Niederlande, Serbien und Zypern. In Abbildung 4.5 stehen diese Staaten in einem Block mit Deutschland und sind mit einem grünen Quadrat gekennzeichnet. Der mittlere

Abbildung 4.5: Testleistung der Schülerinnen und Schüler im internationalen Vergleich – Gesamtskala Naturwissenschaften▲ = Mittelwert im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)▼ = Mittelwert im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)■ = Unterschied zu Deutschland nicht statistisch signifikant ($p > .05$)

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

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

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

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

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

Kompetenzwert von Deutschland liegt (gerade) signifikant unter dem Mittelwert der zweiten wichtigen Vergleichsgruppe, den OECD-Staaten (526 Punkte), auch wenn dieser Unterschied eher gering ist. Eine große Differenz zeigt sich zu den Spitzenreitern Singapur und der Republik Korea von 77 beziehungsweise 70 Punkten. Dieser Unterschied entspricht in etwa dem einer Kompetenzstufe.

Streuung der naturwissenschaftlichen Kompetenz im internationalen Vergleich: Während die Mittelwerte über das im Durchschnitt erreichte Niveau der naturwissenschaftlichen Kompetenz Auskunft geben, veranschaulichen Standardabweichung und Perzentilbänder (siehe Abbildung 4.5) die Streuung und Verteilung der Kompetenz. Die Perzentilwerte teilen die Bandbreite der Kompetenzwerte in definierte Prozentbereiche ein. Der Kompetenzwert, der dem 75. Perzentil entspricht, sagt aus, dass 75 Prozent der Schülerinnen und Schüler einen niedrigeren oder identischen Kompetenzwert und entsprechend 25 Prozent einen höheren Kompetenzwert erzielt haben. Die Breite des Perzentilbandes vom 5. bis zum 95. Perzentil repräsentiert die Streuung der Kompetenzwerte, über die die mittleren 90 Prozent der Schülerinnen und Schüler verfügen, also die, die nicht zu den jeweils 5 Prozent der Schülerinnen und Schüler mit den niedrigsten und höchsten Kompetenzwerten gehören. Entsprechend gibt die Breite des Perzentilbandes vom 25. bis zum 75. Perzentil die Bandbreite der mittleren Kompetenzwerte der 50 Prozent der Schülerinnen und Schüler an, die nicht zu den jeweils 25 Prozent der Schülerinnen und Schüler mit den niedrigsten und höchsten Kompetenzwerten gehören. Abbildung 4.5 zeigt, dass die Streuung der Kompetenzwerte in allen Staaten beachtlich ist, es aber trotzdem zwischen den Staaten bemerkenswerte Unterschiede gibt. Sehr große Streuungen finden sich vor allem in den Staaten, die ein niedriges Niveau der naturwissenschaftlichen Kompetenz haben. Beispielsweise betragen die Kompetenzunterschiede zwischen dem 5. und dem 95. Perzentil in Marokko 399 Punkte ($SD = 121$) und in Pakistan 413 Punkte ($SD = 126$) (Mullis, Martin, Foy, Kelly & Fishbein, 2020). Die größte Streuung weist Südafrika auf ($SD = 133$). Hier beträgt der Unterschied 434 Punkte. Eine relativ große Streuung zeigt sich auch in Bulgarien (340 Punkte, $SD = 103$), einem der Staaten, der eine ähnliche mittlere Kompetenz (521 Punkte) aufweist wie Deutschland.

In der Russischen Föderation, Taiwan und Finnland ist die Streuung zwischen dem 5. und dem 95. Perzentil mit 209, 215 beziehungsweise 233 Punkten ($SD = 64$, 65 beziehungsweise 71) gering (siehe Tabellen 4.9.1 und 4.9.2). Diesen Staaten gelingt es besser, die naturwissenschaftliche Kompetenz ihrer Grundschülerinnen und Grundschüler in der Breite zu fördern, aber gleichzeitig die leistungsstärkste Gruppe nicht zu vernachlässigen. Sehr geringe Streuungen zeigen sich in Kroatien und Dänemark mit 194 und 224 Punkten ($SD = 59$ und 68), allerdings sind hier die Kompetenzwerte, die die 5 Prozent der leistungsstärksten Schülerinnen und Schüler erreichen, niedriger im Vergleich zu den vorher genannten Staaten (Tabelle 4.9.1 und 4.9.2). Für die beiden Staaten auf den höchsten Rangplätzen zeigt sich wie in den Jahren davor für die Republik Korea eine eher geringe Streuung (219 Punkte), während die Streuung für Singapur mit 254 deutlich größer ist ($SD = 67$ bzw. 78).

In Deutschland erreichen die 5 Prozent der leistungsschwächsten Schülerinnen und Schüler maximal einen Kompetenzwert von 383 Punkten und die 5 Prozent der leistungsstärksten minimal einen Kompetenzwert von 634 Punkten. Der Unterschied liegt entsprechend bei 251 Punkten. Die Standardabweichung liegt bei 77 Punkten. Dies bedeutet, dass circa zwei Drittel der Schülerinnen und

Schüler Kompetenzwerte zwischen 441 ($518 - 77$) und 595 ($518 + 77$) Punkten erreichen. Damit sind die Standardabweichungen ähnlich wie die der EU- und OECD-Vergleichsgruppen ($SD =$ beide 73).

Leistungsstärkste und leistungsschwächste Schülerinnen und Schüler: Die Perzentilbänder geben auch Auskunft darüber, wie gut die Förderung der leistungsstärksten und -schwächsten Schülerinnen und Schüler gelingt. So kennzeichnet der Punktwert des 95. Perzentils den Kompetenzwert, bei dem die Gruppe der 5 Prozent leistungsstärksten Schülerinnen und Schüler eines Staates beginnt. Eine besonders gelungene Spitzenförderung zeigt sich in Singapur (hier beginnt die Spitzengruppe bei 713 Punkten) und der Republik Korea (693 Punkte). Weitere kompetenzstarke Staaten wie die Russische Föderation, Finnland, Japan und die USA zeigen Werte zwischen 662 (Finnland) und 668 (Japan) Punkten (Mullis, Martin, Foy, Kelly & Fishbein, 2020). Die Spitzengruppe der Schülerinnen und Schüler in Deutschland beginnt bei 634 Punkten und erreicht damit ein deutlich niedrigeres Niveau. Entsprechend dürfte bei der Förderung hochkompetenter Schülerinnen und Schüler in Deutschland das Potenzial noch nicht ausgeschöpft sein. Auch bei der Betrachtung der 5 Prozent leistungsschwächsten Schülerinnen und Schüler zeigen sich Unterschiede zwischen den Teilnehmerstaaten. So erreichen diese in den sehr kompetenzstarken Staaten maximale Werte zwischen 454 und 457, in der Republik Korea sogar 474 Punkte, während es in Deutschland lediglich 338 Punkte sind. Das deutet darauf hin, dass es in Deutschland weniger gut gelingt, Schülerinnen und Schüler mit ungünstigen Voraussetzungen auf geeignete Weise zu fördern.

Unterschiede in der naturwissenschaftlichen Kompetenz zwischen TIMSS 2007, 2011, 2015 und 2019: Durch die Teilnahme von Deutschland an vier Studienzyklen ist es möglich, Tendaussagen über den Verlauf von zwölf Jahren darzustellen. In Abbildung 4.6 sind die nationalen Mittelwerte und die zugehörigen Standardfehler der naturwissenschaftlichen Kompetenz für die Staaten gegenübergestellt, die seit 2007 kontinuierlich an TIMSS teilgenommen haben. Eine Ausnahme stellt Österreich dar, das 2015 nicht an TIMSS teilgenommen hat. Da Österreich aber vergleichbare Entwicklungen im Bildungssystem wie Deutschland zu verzeichnen hat, zum Beispiel die Einführung von Bildungsstandards im Jahr 2008 für Deutsch und Mathematik am Ende der Grundschulzeit, eignet es sich zum Vergleich und ist deshalb hier mit aufgenommen. Im rechten Teil der Abbildung sind die Differenzbalken für den Vergleich von TIMSS 2007 und 2019 dargestellt.

Während die mittlere Kompetenz in den Naturwissenschaften in Deutschland zwischen 2007 und 2015 stabil war (528 Punkte), zeigt sich im aktuellen Zyklus eine signifikante Abnahme der mittleren Kompetenz um 10 Punkte (Abbildung 4.6). Diese Veränderung lässt sich nicht durch die computerbasierte Testform erklären, die im Vergleich zur papierbasierten Testung für die Schülerinnen und Schüler etwas schwieriger ist (siehe Kapitel 5 in diesem Band). So zeigt sich auch in der Stichprobe, die papierbasiert getestet wurde, eine zwar etwas geringere, aber auch negative Entwicklung der mittleren Kompetenz auf 522 Punkte. Ob dies den Beginn eines Negativtrends darstellt, lässt sich aktuell nicht einschätzen, generell ist diese Entwicklung zwar ungünstig, aber zunächst als gering einzuschätzen.

Die Trends in den anderen teilnehmenden Staaten sind unterschiedlich. So zeigt sich in einer Reihe von Staaten eine vergleichbare Abnahme der mitt-

Tabelle 4.9.1: Testleistungen nach Perzentilen im Vergleich: TIMSS 2007, 2011, 2015 und 2019 – Gesamtskala Naturwissenschaften

Fußnoten														
2007	2011	2015	2019 T	Teilnehmer ^a	TIMSS 2007 Perzentile				TIMSS 2011 Perzentile					
					5 (SE)	25 (SE)	75 (SE)	95 (SE)	5 (SE)	25 (SE)	75 (SE)	95 (SE)		
	2	2	2	8 Singapur	418 (7.0) +	531 (6.0) +	652 (4.4)	727 (3.9) –	427 (6.7) +	531 (5.8) +	644 (4.1)	713 (4.0)		
	2		2	8 Russische Föderation	407 (7.6) +	495 (6.6) +	601 (4.2) +	672 (8.8)	430 (5.2) +	505 (3.6) +	603 (2.9) +	667 (4.8)		
			3	8 Japan	428 (7.0)	505 (3.1) +	595 (1.4) +	655 (3.2) +	449 (4.1)	519 (2.6)	601 (1.9) +	658 (2.8) +		
				Taiwan	423 (6.5) +	508 (2.0) +	609 (2.4)	679 (3.0) –	420 (6.5) +	506 (4.3) +	603 (2.6)	664 (4.2)		
2 3	2 3	2 3	2 3	USA	392 (4.9)	484 (3.2)	597 (3.4)	668 (3.2)	406 (3.9) –	494 (2.4)	599 (2.5)	666 (2.3)		
2	2	2	2	8 Litauen	401 (5.2)	473 (2.8) +	559 (3.2) +	615 (4.2) +	397 (6.0) +	471 (2.9) +	561 (2.0) +	620 (5.0) +		
		2	2	8 Schweden	400 (3.3)	478 (4.0) +	575 (3.0) +	642 (4.0)	403 (6.2)	486 (3.1)	586 (2.6)	648 (4.2)		
1 3	1 3	1	1 2	8 England	403 (6.4)	491 (6.0)	596 (3.6)	666 (6.3) –	384 (6.3) +	476 (5.1) +	586 (3.5)	653 (3.0)		
	2			Tschechische Republik	386 (8.9) +	467 (4.4) +	567 (3.8) +	635 (4.3)	412 (9.3)	491 (2.3)	586 (3.8)	648 (3.8)		
				Australien	384 (9.0)	478 (4.4)	583 (2.6)	651 (4.8)	371 (9.0)	466 (4.3) +	571 (3.0) +	638 (5.3) +		
2 3	2		3	8 Hongkong	437 (4.6) –	511 (3.7) –	601 (3.6) –	659 (4.2) –	406 (16.3)	493 (3.3)	585 (2.7)	644 (5.1)		
				Ungarn	383 (9.4)	485 (3.8)	595 (3.4) –	661 (4.7) –	377 (8.7)	484 (6.4)	594 (4.0)	662 (4.2) –		
3	2 3	2 3	3	8 Dänemark	383 (6.5) +	468 (2.6) +	570 (5.1)	636 (4.7)	401 (7.0)	483 (3.7)	578 (3.1) –	640 (4.3) –		
	2		2	8 Österreich	388 (4.4)	477 (3.3)	580 (2.5)	644 (3.5)	408 (4.8)	485 (4.2) –	581 (2.0)	640 (2.5)		
			2	Slowakei	376 (13.9)	476 (4.7)	584 (2.8) –	652 (6.9) –	390 (13.9)	486 (4.3)	586 (2.7) –	648 (5.4) –		
3	3	3	3	8 Niederlande	421 (6.3)	484 (3.6)	565 (3.9)	617 (5.0)	439 (8.2) –	497 (2.2) –	568 (1.8)	613 (2.2)		
				Deutschland	393 (7.6)	479 (4.2) –	582 (1.6) –	647 (6.0)	406 (7.1) –	482 (3.6) –	577 (2.8)	636 (4.9)		
2	3	2 3		8 Italien	395 (5.0)	484 (5.2) –	590 (2.9) –	664 (4.4) –	397 (8.3)	477 (3.4) –	573 (3.1) –	641 (3.3) –		
1 2	1 3 6 1	3	1 2	8 Neuseeland	344 (4.9) +	447 (3.7)	568 (2.9)	643 (2.6) –	345 (6.9)	442 (3.2)	558 (2.3)	626 (2.8)		
2	2	2	2	Georgien	273 (8.6) +	361 (5.2) +	477 (4.4) +	552 (7.9) +	299 (6.5)	401 (7.2)	516 (3.7)	585 (2.9)		
				8 Iran	267 (7.1)	371 (6.5)	506 (5.3)	587 (3.3)	274 (6.7)	390 (5.6)	523 (3.7) –	604 (6.0) –		
				Benchmark-Teilnehmer										
2 3 6	2	2	2	8 Dubai, VAE	267 (7.6) +	390 (5.2) +	537 (3.3) +	620 (3.3) +	260 (5.7) +	386 (3.4) +	544 (2.9) +	631 (2.9) +		
2	2			8 Ontario, Kanada	396 (8.1)	487 (3.0) –	590 (4.7) –	657 (5.2) –	393 (5.8)	479 (3.3)	581 (3.7)	646 (3.3)		
2		2 3	3	8 Québec, Kanada	405 (6.8)	474 (3.6)	563 (3.4)	623 (9.2)	417 (6.2)	478 (3.5)	556 (4.6)	611 (3.2) +		

+ = Mittelwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)– = Mittelwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

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

6 = Abweichender Testzeitpunkt

8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

leren Kompetenz um circa 10 Punkte im Vergleich zu TIMSS 2015 wie in Deutschland. Dazu gehört zum Beispiel die Flämische Gemeinschaft in Belgien, Bulgarien, Kroatien, Ungarn und Polen (Mullis et al., 2020). Relativ stabil bleiben die Leistungswerte zwischen 2015 und 2019 unter anderem in Schweden, der Slowakei, der Tschechischen Republik, den Niederlanden, Frankreich, England und in Neuseeland. Auffällige positive Unterschiede zu TIMSS 2015 zeigen sich vor allem in Staaten, die bislang schwächer, also unter dem internationalen Mittelwert abgeschnitten haben. Beispiele solcher Staaten sind Zypern, Bahrain, Armenien, der Iran und die Vereinigten Arabischen Emirate (Mullis et al., 2020). Die mittleren Kompetenzwerte der meisten der 44 Staaten, die an TIMSS 2015 und 2019 teilgenommen haben, bleiben aber stabil. Dies gilt auch für den Vergleich mit TIMSS 2007, von den 21 Staaten, die an beiden Studien teilgenommen haben, haben sechs höhere durchschnittliche, drei geringere und die anderen gleiche Kompetenzwerte erzielt (Abbildung 4.6).

Unterschiede in der Streuung zwischen TIMSS 2007, 2011, 2015 und 2019: Neben Unterschieden in den Mittelwerten ist auch der Vergleich der Heterogenität der Leistungen in den verschiedenen Studienzyklen von Interesse. In Tabelle 4.9.1 und 4.9.2 sind neben den Angaben zu den Perzentilen für die Kompetenzen

Tabelle 4.9.2: Testleistungen nach Perzentilen im Vergleich: TIMSS 2007, 2011, 2015 und 2019 – Gesamtskala Naturwissenschaften

Fußnoten														
2007	2011	2015	2019 T	Teilnehmer ^A	TIMSS 2015 Perzentile					TIMSS 2019 Perzentile				
					5 (SE)	25 (SE)	75 (SE)	95 (SE)		5 (SE)	25 (SE)	75 (SE)	95 (SE)	
2	2	2	2	»Singapur	434 (7.1) +	540 (5.2)	650 (3.5)	716 (4.9)		454 (6.3) ▲	548 (4.5) ▲	649 (3.3) ▲	708 (3.7) ▲	
2			2	»Russische Föderation	449 (6.8)	524 (3.6)	615 (3.4)	674 (2.8)		457 (7.3) ▲	526 (3.5) ▲	611 (2.9) ▲	666 (4.6) ▲	
			3	»Japan	459 (3.9) –	528 (2.4) –	613 (1.9)	671 (4.2)		442 (4.5) ▲	519 (2.7) ▲	609 (2.1) ▲	668 (3.4) ▲	
				Taiwan	435 (5.5)	513 (2.3)	602 (2.4)	661 (4.1)		444 (4.7) ▲	516 (2.7) ▲	604 (1.9) ▲	659 (4.1) ▲	
2 3	2 3	2 3	2 3	USA	404 (4.8) –	495 (3.2)	602 (2.1)	669 (3.4)		387 (5.6) ■	486 (3.6) ▲	598 (2.2) ▲	664 (2.6) ▲	
2	2		2	»Litauen	406 (5.4)	483 (3.9)	576 (3.0) +	634 (3.9) +		414 (6.0) ▲	492 (3.4) ▲	588 (2.2) ▲	651 (4.7) ▲	
			2	Schweden	410 (8.7)	496 (4.9)	591 (3.4)	652 (4.4)		409 (7.2) ▲	490 (4.6) ▲	588 (3.8) ▲	652 (3.8) ▲	
1 3	1 3	1	1 2	»England	417 (5.1)	490 (3.2)	583 (3.0)	648 (4.0)		413 (6.8) ▲	491 (3.3) ▲	587 (3.3) ▲	648 (4.8) ▲	
			2	Tschechische Republik	414 (6.1)	490 (3.5)	583 (2.7)	644 (3.4)		414 (5.1) ▲	490 (3.4) ▲	582 (2.2) ▲	640 (3.2) ■	
				Australien	389 (7.5)	476 (3.8)	576 (3.1) +	640 (4.8) +		389 (6.0) ■	484 (2.8) ▲	587 (2.4) ▲	653 (4.3) ▲	
2 3	2		3	»Hongkong	437 (5.7) –	512 (4.0) –	604 (3.6) –	668 (4.5) –		408 (6.5) ▲	486 (4.0) ▲	580 (3.2) ■	643 (4.1) ■	
				Ungarn	393 (10.2)	493 (5.6) –	598 (2.6) –	663 (3.6) –		395 (5.6) ■	479 (3.9) ▲	584 (3.2) ▲	647 (3.2) ▲	
3	2 3	2 3	3	»Dänemark	406 (4.8)	483 (3.3)	575 (2.7)	633 (3.1)		405 (6.6) ▲	478 (3.1) ▲	569 (2.7) ■	629 (3.4) ■	
	2		2	»Österreich						394 (6.0) ■	474 (3.2) ■	575 (3.5) ■	636 (4.2) ■	
			2	Slowakei	362 (7.3)	473 (4.4)	578 (3.2)	645 (3.6)		367 (13.8) ■	479 (4.4) ■	575 (2.6) ■	634 (4.1) ■	
3	3	3	3	»Niederlande	414 (5.6)	477 (2.8)	559 (2.8)	610 (4.1)		407 (5.4) ▲	476 (3.9) ■	564 (2.8) ▼	619 (5.7) ▼	
				Deutschland	409 (5.3) –	483 (3.0) –	577 (2.6)	638 (3.6)		383 (4.1) ■	468 (3.4) ■	573 (2.7) ■	634 (4.8) ■	
2	3	2 3		»Italien	399 (5.5)	474 (4.4)	562 (2.5)	618 (2.8)		400 (5.7) ▲	467 (3.1) ■	555 (3.7) ▼	615 (7.0) ▼	
1 2	1 3 6 1	3	1 2	»Neuseeland	351 (6.6)	451 (4.4)	566 (2.4)	633 (3.5)		360 (4.0) ▼	448 (3.2) ▼	561 (2.5) ▼	629 (3.7) ■	
2	2	2	2	Georgien	296 (9.4)	397 (5.1)	511 (3.6)	584 (7.2)		307 (10.0) ▼	400 (5.0) ▼	512 (3.6) ▼	582 (5.1) ▼	
				»Iran	238 (10.9)	354 (7.5) +	496 (4.1) +	571 (4.3) +		265 (10.4) ▼	377 (6.0) ▼	511 (3.9) ▼	587 (4.4) ▼	
Benchmark-Teilnehmer														
2 3 6	2	2	2	»Dubai, VAE	324 (4.7) +	455 (3.6) +	591 (2.2) +	670 (2.5)		389 (5.6) ■	492 (2.8) ▲	605 (1.9) ▲	671 (2.9) ▲	
2	2			»Ontario, Kanada	405 (4.9)	484 (2.8)	580 (2.8)	643 (3.7)		397 (6.3) ■	474 (4.7) ■	577 (3.8) ■	641 (5.3) ■	
2		2 3	3	»Québec, Kanada	415 (6.8)	481 (4.7)	569 (5.1)	629 (4.7)		413 (5.3) ▲	479 (4.4) ■	566 (3.2) ■	627 (4.7) ■	

▲ = Punktwerte in den Perzentilen im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)▼ = Punktwerte in den Perzentilen im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)■ = Unterschied zu Punktwerten in den Perzentilen im Vergleich zu Deutschland nicht statistisch signifikant ($p > .05$)+ = Mittelwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)– = Mittelwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

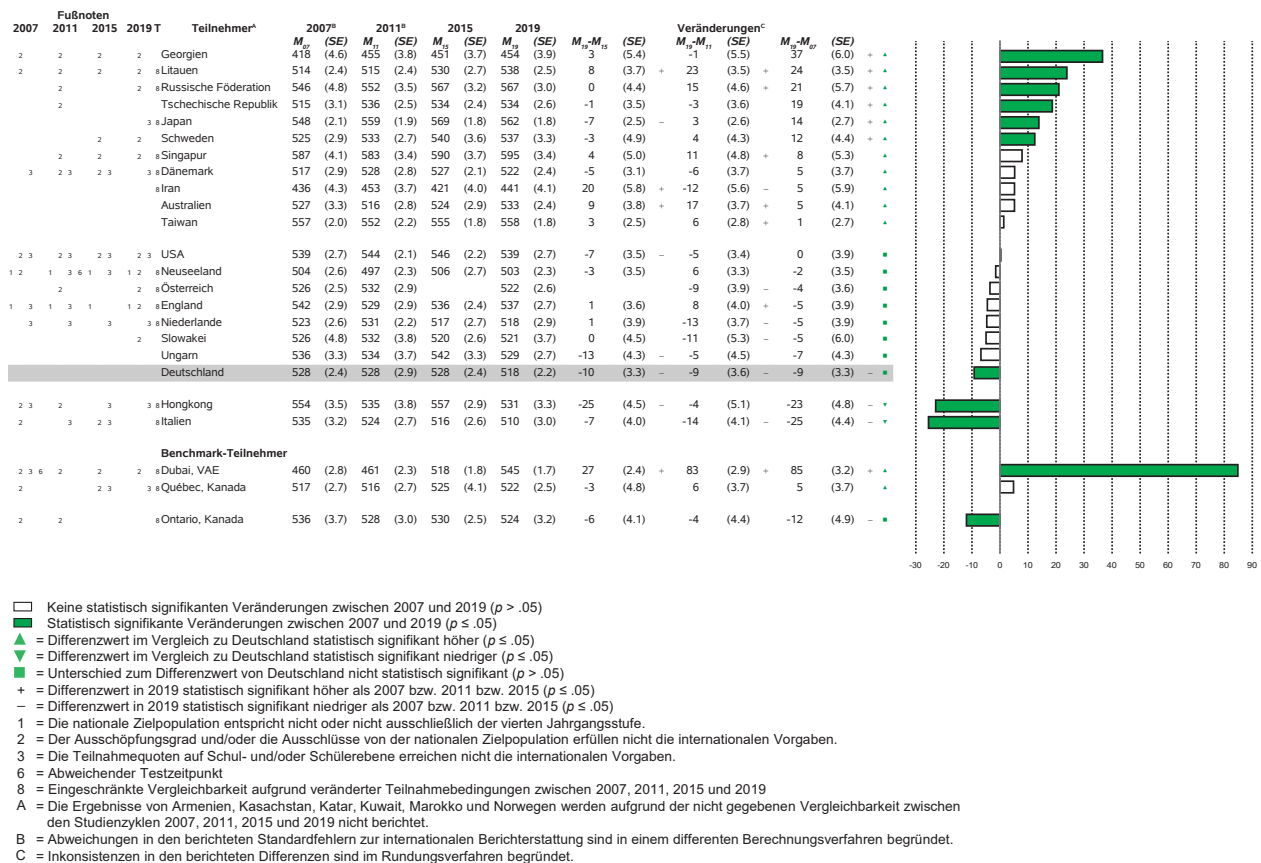
3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

6 = Abweichender Testzeitpunkt

8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

in TIMSS 2019, die oben bereits beschrieben wurden, auch die entsprechenden Werte für die vorherigen Studienzyklen dargestellt. Aufgeführt sind mit Ausnahme von Österreich nur die Staaten, die seit 2007 an TIMSS teilgenommen haben. Die Staaten sind in der Rangfolge des mittleren Kompetenzwertes in 2019 geordnet. Für Deutschland zeigen die Ergebnisse, dass die Streuung der Leistungen, über die die mittleren 90 Prozent der Schülerinnen und Schüler verfügen, im Vergleich zu TIMSS 2015 (229) und TIMSS 2011 (230 Punkte) mit 251 Punkten zugenommen hat und ähnlich ist wie in TIMSS 2007, wo die Streuung 254 Punkte betrug. Die Unterschiede ergeben sich vor allem durch die breitere Streuung im unteren Leistungsbereich. Betrachtet man das 5. Perzentil, also den Kompetenzwert, den die 5 Prozent der leistungsschwächsten Schülerinnen und Schüler maximal erreichen, so liegt der Wert 2019 bei 383 Punkten, 2015 bei 409 Punkten, 2011 bei 406 Punkten und 2007 bei 393 Punkten. Der aktuelle Wert ist also beachtlich niedriger. Im oberen Leistungsspektrum sind die Werte zumindest im Vergleich zu den letzten beiden Studienzyklen sehr ähnlich. Hier liegt der Kompetenzwert, den die 5 Prozent der leistungsstärksten Schülerinnen und Schüler mindestens erreichen, in TIMSS 2019 bei 634 Punkten im Vergleich zu 638 (TIMSS 2015), 636 (TIMSS 2011) und 647 (TIMSS 2007) Punkten.

Abbildung 4.6: Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 – Gesamtskala Naturwissenschaften

Zusammengefasst zeigt sich also, dass die Leistungen etwas weniger homogen sind und die Streuung sich eher im unteren Leistungsbereich vergrößert hat. In den meisten anderen Teilnehmerstaaten ist die Streuung der Leistungen relativ stabil. Ähnliche Veränderungen wie in Deutschland zeigen sich in Österreich, der Flämischen Gemeinschaft in Belgien, Portugal und den Niederlanden.

4.5.2 Kompetenzstufen

In diesem Abschnitt wird zunächst erläutert, wie sich die Schülerinnen und Schüler auf die Kompetenzstufen verteilen und ob sich Unterschiede zwischen den Verteilungen von TIMSS 2007 bis TIMSS 2019 finden. Die gezielte Betrachtung der Verteilung ist insbesondere interessant im Hinblick auf die oberen und unteren Kompetenzstufen und liefert zu dem Blick auf die Perzentile weitere stärker inhaltliche Informationen. Die Erläuterungen zur inhaltlichen Bedeutung der fünf Kompetenzstufen, die von einem sporadischen Alltagswissen bis zu einem grundlegenden Verständnis von Inhalten sowie Denk- und Arbeitsweisen reichen, finden sich im Abschnitt 4.3.

Abbildung 4.7 zeigt die Verteilung der Schülerinnen und Schüler auf die fünf Kompetenzstufen für alle teilnehmenden Staaten. Die Kompetenzausprägungen, die auf den einzelnen Stufen erforderlich sind, bauen aufeinander auf. Schülerinnen und Schüler der Kompetenzstufe IV verfügen dementsprechend auch

über die Fähigkeiten und Fertigkeiten der Stufen I bis III. Entsprechend sind auf der linken Seite der Abbildung 4.7 die kumulierten prozentualen Anteile der Kompetenzstufen dargestellt. Die Staaten sind nach ihrem Mittelwert der naturwissenschaftlichen Kompetenz in einer Rangreihe angeordnet.

In allen teilnehmenden Staaten erreichen im Mittel 17 Prozent der Schülerinnen und Schüler lediglich die Kompetenzstufe I und befinden sich damit am untersten Ende des Leistungsspektrums. Diese Gruppe von Schülerinnen und Schülern verfügt allenfalls über sporadisches Alltagswissen. Am anderen Ende des Leistungsspektrums auf der Kompetenzstufe V befinden sich im Mittel 7 Prozent der Schülerinnen und Schüler. Diese verfügen bereits über ein vertieftes Verständnis von Konzepten in allen Inhaltsbereichen und von Denk- und Arbeitsweisen. Die Unterschiede in der Verteilung auf die höchste und niedrigste Kompetenzstufe zwischen den teilnehmenden Staaten sind enorm. So erreichen 38 beziehungsweise 29 Prozent der Schülerinnen und Schüler in Singapur beziehungsweise der Republik Korea die Kompetenzstufe V, in der Russischen Föderation und Japan sind es 17 bis 18 Prozent, während in anderen Staaten wie den Philippinen keine Schülerinnen und Schüler diese Kompetenzstufe erreichen. Entsprechendes zeigt sich auch hinsichtlich der untersten Kompetenzstufe. So befinden sich 87 Prozent der Schülerinnen und Schüler aus den Philippinen auf der Kompetenzstufe I, während in den sehr leistungsstarken Staaten hier nur 1 bis 2 Prozent eingeordnet werden (Mullis et al., 2020).

Die Verteilung in Deutschland ist ähnlich wie in den teilnehmenden EU- und OECD-Staaten. Die niedrigste und die höchste Kompetenzstufe (Kompetenzstufe I beziehungsweise V) sind mit jeweils circa 7 Prozent der Schülerinnen und Schüler besetzt. In den beiden Vergleichsgruppen EU- und OECD-Staaten sind es 6 Prozent auf der Kompetenzstufe I und 7 beziehungsweise 9 Prozent auf der Kompetenzstufe V. Fasst man die beiden unteren Kompetenzstufen I und II zusammen, ergeben sich für Deutschland 28 Prozent (zum Vergleich: EU-Staaten 25 % und OECD-Staaten 24 %). Dieser relativ großen Gruppe (sehr) leistungsschwacher Schülerinnen und Schüler fehlen zentrale Grundlagen für den anschließenden Biologie-, Chemie- und Physikunterricht sowie die naturwissenschaftlichen Anteile des Geografieunterrichts der weiterführenden Schulen. Anderen Staaten wie beispielsweise Lettland, Finnland und der Republik Korea gelingt es besser, am unteren Ende des Leistungsspektrums zu fördern und die Anzahl der Schülerinnen und Schüler auf den beiden unteren Kompetenzstufen deutlich zu verringern. So finden sich hier lediglich 1 bis 3 Prozent der Schülerinnen und Schüler auf der untersten Kompetenzstufe und 15, 13 oder 5 Prozent auf den beiden untersten Kompetenzstufen. Die zweithöchste Kompetenzstufe IV, in der sich Schülerinnen und Schüler befinden, die bereits ein in erster Linie inhaltspezifisches anwendbares Wissen entwickelt haben, umfasst 30 Prozent der Schülerinnen und Schüler in Deutschland. Damit ist der Anteil etwas höher als der internationale Durchschnitt (24 %) und vergleichbar mit dem der anderen EU- und OECD-Staaten. Die Gruppe der Schülerinnen und Schüler auf den beiden höchsten Kompetenzstufen macht in Deutschland wie auch in den EU-Staaten gemeinsam einen Anteil von 37 Prozent aus (OECD-Staaten: 40 %). Die Gruppe verfügt über (sehr) gute Voraussetzungen für die erfolgreiche Teilnahme am naturwissenschaftlichen Unterricht in den weiterführenden Schulen. Es zeigt sich allerdings auch hier, dass es anderen Staaten besser gelingt, leistungsstarke Schülerinnen und Schüler zu fördern. So umfasst die Gruppe von Schülerinnen und Schülern auf den beiden obersten Kompetenzstufen beispielsweise in Singapur 74 Prozent, in Finnland 56 Prozent und in den USA 48 Prozent.

Tabelle 4.10.1: Prozentuale Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007, 2011, 2015 und 2019

Fußnoten					TIMSS 2007 Kompetenzstufen										TIMSS 2011 Kompetenzstufen																
2007	2011	2015	2019	T	Teilnehmer ^A	I		II		III		IV		V		I		II		III		IV		V							
						%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)						
	2	2		2	»Singapur	3.8	(0.5)	–	7.9	(0.8)	–	20.1	(1.2)	32.2	(1.4)	+ 35.9	(1.9)	3.2	(0.4)	–	7.9	(0.8)	–	21.1	(1.1)	34.6	(1.1)	33.2	(1.7)		
	2			2	»Russische Föderation	4.3	(0.9)	–	13.9	(1.3)	–	32.3	(1.3)	33.7	(1.3)	+ 15.8	(1.9)	2.1	(0.4)	–	12.2	(0.9)	–	33.5	(1.4)	–	36.3	(1.4)	+ 15.8	(1.4)	
				3	»Japan	2.7	(0.4)	11.3	(0.9)	–	35.2	(0.9)	–	38.4	(0.9)	+ 12.4	(1.0)	+ 1.3	(0.2)	8.4	(0.7)	32.8	(1.2)	43.2	(1.2)	14.3	(1.0)	+			
					Taiwan	3.0	(0.4)	–	11.0	(0.6)	30.9	(1.1)	36.4	(1.3)	+ 18.7	(1.0)	–	3.1	(0.4)	–	11.6	(0.9)	31.8	(1.0)	38.3	(1.2)	+ 15.2	(0.9)			
2 3	2 3	2 3	2 3	2 3	»USA	6.0	(0.6)	16.1	(0.7)	31.2	(0.9)	31.8	(0.9)	15.0	(0.9)	4.4	(0.4)	+ 14.1	(0.6)	32.2	(0.7)	34.6	(0.8)	14.7	(0.8)						
2	2	2		2	»Litauen	4.9	(0.6)	21.2	(1.2)	–	43.6	(1.3)	–	27.0	(1.3)	+ 3.3	(0.4)	+ 5.3	(0.6)	–	21.4	(1.0)	–	42.4	(1.2)	–	26.5	(1.4)	+ 4.3	(0.5)	+
				2	Schweden	5.0	(0.6)	19.0	(1.1)	–	38.5	(1.0)	–	29.4	(1.3)	+ 8.0	(0.6)	+ 4.7	(0.5)	16.2	(0.9)	35.5	(1.0)	33.5	(1.0)	10.1	(1.0)				
1 3	1 3	1 3	1	1 2	»England	4.7	(0.6)	14.7	(0.8)	32.7	(1.2)	+ 33.5	(1.1)	14.2	(1.2)	–	6.8	(0.7)	–	17.7	(0.9)	–	33.2	(1.3)	+ 31.3	(1.3)	11.1	(0.9)			
				2	Tschechische Republik	7.0	(0.8)	–	21.5	(1.1)	–	38.5	(1.5)	26.1	(1.7)	+ 7.0	(0.7)	3.5	(0.7)	15.5	(1.1)	36.7	(1.4)	34.3	(1.4)	10.0	(0.9)				
					Australien	6.7	(0.8)	16.9	(1.1)	35.5	(1.3)	30.6	(2.1)	10.2	(0.7)	8.8	(1.0)	–	19.6	(1.0)	–	36.3	(1.1)	27.9	(1.2)	+ 7.4	(0.7)	+			
				3	»Hongkong	2.0	(0.4)	+ 10.1	(1.0)	+ 33.1	(1.4)	+ 40.7	(1.5)	–	14.1	(1.4)	–	4.5	(1.2)	13.9	(1.1)	37.0	(1.6)	35.5	(1.7)	9.1	(0.9)				
					Ungarn	6.6	(0.8)	15.1	(1.3)	31.1	(1.5)	33.7	(1.5)	13.4	(1.0)	–	7.4	(0.9)	14.8	(1.1)	+ 31.6	(1.2)	33.1	(1.6)	13.2	(0.9)	–				
	3	2 3	2 3	3	»Dänemark	7.1	(0.8)	–	20.8	(1.0)	37.6	(1.6)	27.7	(1.1)	6.9	(0.8)	4.8	(0.7)	17.1	(1.1)	38.6	(1.2)	31.5	(1.3)	8.0	(0.8)	–				
				2	»Österreich	6.5	(0.6)	17.8	(1.1)	36.2	(1.1)	30.7	(1.2)	8.7	(0.7)	4.0	(0.6)	17.0	(1.5)	37.2	(1.4)	33.5	(1.5)	8.3	(0.8)						
				2	Slowakei	7.5	(1.3)	17.0	(1.2)	33.7	(1.1)	+ 31.1	(1.6)	10.6	(0.8)	–	5.9	(1.0)	15.2	(1.2)	34.8	(1.4)	34.3	(1.1)	9.8	(1.0)	–				
3	3	3	3	3	»Niederlande	2.5	(0.5)	+ 18.2	(1.3)	45.4	(1.4)	30.2	(1.5)	3.7	(0.8)	1.2	(0.4)	+ 12.8	(1.3)	+ 48.5	(1.2)	–	34.6	(1.7)	–	2.9	(0.5)				
					Deutschland	5.8	(0.6)	17.9	(0.9)	35.4	(0.7)	31.3	(1.1)	9.6	(0.7)	–	4.3	(0.7)	+ 17.7	(1.3)	38.6	(1.2)	32.3	(1.3)	7.1	(0.6)					
2		3	2 3		»Italien	5.5	(0.7)	16.1	(0.9)	+ 34.5	(1.0)	+ 30.9	(1.2)	–	13.0	(1.0)	–	5.3	(1.0)	18.9	(1.2)	+ 38.9	(1.4)	+ 29.1	(1.4)	–	7.8	(0.7)	–		
1 2	1 3 6 1	3 1 2			»Neuseeland	13.4	(1.0)	21.7	(0.8)	32.7	(0.9)	24.5	(0.9)	7.8	(0.5)	–	13.8	(0.9)	23.6	(0.9)	34.3	(1.2)	23.1	(0.9)	5.2	(0.5)					
2	2	2	2	2	Georgien	41.1	(2.1)	–	32.9	(1.5)	20.6	(1.7)	+ 4.9	(0.7)	+ 0.5	(0.2)	+ 24.8	(1.6)	31.4	(1.2)	31.3	(1.3)	11.2	(1.1)	1.4	(0.4)					
					»Iran	34.7	(1.9)	29.0	(1.1)	24.6	(1.2)	10.1	(0.9)	1.6	(0.3)	27.9	(1.5)	27.9	(1.1)	28.0	(1.1)	13.4	(0.9)	2.8	(0.4)	–					
					Benchmark-Teilnehmer																										
2 3 6	2	2	2	2	»Dubai, VAE	27.8	(1.4)	–	24.2	(1.3)	–	27.4	(1.2)	16.3	(0.8)	+ 4.2	(0.5)	+ 28.3	(1.1)	–	23.3	(1.0)	–	25.3	(0.9)	+ 17.3	(0.6)	+ 5.7	(0.7)	+	
2	2				»Ontario, Kanada	5.4	(1.0)	15.6	(1.1)	+ 33.6	(1.4)	33.5	(2.0)	11.9	(1.2)	–	5.9	(0.6)	17.5	(1.2)	36.3	(1.2)	31.1	(1.2)	9.3	(0.9)					
2		2 3	3	3	»Québec, Kanada	4.3	(0.6)	21.4	(1.5)	42.3	(1.9)	27.2	(1.8)	4.8	(0.6)	2.8	(0.4)	20.7	(1.4)	47.8	(1.3)	–	25.8	(1.3)	2.9	(0.5)	+				

+ = Prozentwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)– = Prozentwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

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

6 = Abweichender Testzeitpunkt

8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

Unterschiede in den Anteilen der Schülerinnen und Schüler auf den Kompetenzstufen zwischen 2007, 2011, 2015 und 2019: In den Tabellen 4.10.1 und 4.10.2 sind die prozentualen Anteile der Schülerinnen und Schüler auf den verschiedenen Kompetenzstufen bei den bisherigen TIMSS-Zyklen dargestellt, um so Unterschiede und möglicherweise Trends abschätzen zu können. In der Tabelle sind mit Ausnahme von Österreich nur Staaten aufgeführt, die an allen vier Studienzyklen teilgenommen haben. Die Staaten sind entsprechend ihrem Mittelwert der naturwissenschaftlichen Kompetenz in einer Rangreihe angeordnet.

Für Deutschland zeigt sich dabei folgendes Bild: Der Anteil der Schülerinnen und Schüler auf den beiden oberen Kompetenzstufen IV und V hat sich im Vergleich zu TIMSS 2015, 2011 und 2007 um 1 bis 2 Prozentpunkte verringert. Dieser deskriptive Unterschied ist lediglich im Vergleich zu TIMSS 2015 statistisch bedeutsam. Auch im mittleren Leistungsbereich auf der Kompetenzstufe III zeigt sich eine Abnahme um 3 Prozentpunkte im Vergleich zu TIMSS 2015 und 2011 und eine Zunahme von 1 Prozent im Vergleich zu TIMSS 2007. Gleichzeitig nimmt der Anteil der Schülerinnen und Schüler auf den beiden unteren Kompetenzstufen bei TIMSS 2019 zu. Diese Veränderungen sind eher klein (1–3 %) und nur teilweise statistisch bedeutsam. Gleichwohl deuten sie

Tabelle 4.10.2: Prozentuale Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007, 2011, 2015 und 2019

Fußnoten					TIMSS 2015 Kompetenzstufen										TIMSS 2019 Kompetenzstufen														
2007	2011	2015	2019 T	Teilnehmer ^A	I		II		III		IV		V		I		II		III		IV		V						
					%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)					
	2	2	2	»Singapur	2.5	(0.5)	7.2	(0.7)	18.8	(1.1)	34.8	(1.5)	36.7	(2.0)	1.8	(0.4)	5.6	(0.6)	18.3	(1.2)	36.7	(1.3)	37.6	(1.9)					
	2		2	»Russische Föderation	1.5	(0.3)	7.8	(0.9)	29.0	(1.4)	41.8	(1.6)	20.0	(1.5)	0.8	(0.3)	7.1	(0.8)	29.5	(1.4)	44.7	(1.3)	18.0	(1.3)					
			3	»Japan	1.0	(0.2)	6.5	(0.5)	+ 29.2	(1.1)	44.1	(1.2)	19.2	(0.9)	1.8	(0.4)	8.5	(0.7)	30.3	(1.0)	42.0	(1.1)	17.4	(0.8)					
				Taiwan	1.9	(0.3)	10.2	(0.7)	32.3	(0.9)	41.2	(1.2)	14.4	(0.7)	1.4	(0.3)	9.3	(0.9)	31.8	(1.1)	42.4	(1.1)	15.0	(0.9)					
2	3	2	3	USA	4.6	(0.5)	+ 13.9	(0.6)	30.7	(0.8)	34.9	(0.9)	15.8	(0.8)	6.5	(0.6)	15.0	(0.7)	30.6	(0.8)	33.3	(1.0)	14.7	(0.8)					
2	2		2	»Litauen	4.3	(0.5)	17.6	(1.0)	38.7	(1.1)	32.6	(1.3)	6.7	(0.8)	+	3.5	(0.4)	15.1	(1.2)	36.0	(1.2)	34.6	(1.3)	10.9	(0.9)				
			2	Schweden	4.0	(0.8)	13.7	(1.1)	35.0	(1.4)	36.2	(1.6)	11.1	(1.1)	4.0	(0.6)	15.7	(1.2)	35.2	(1.1)	33.9	(1.4)	11.2	(1.0)					
1	3	1	3	1	2	»England	3.0	(0.5)	15.8	(1.0)	38.5	(1.5)	33.0	(1.3)	9.6	(0.8)	3.7	(0.6)	14.9	(0.9)	37.2	(1.3)	33.9	(1.6)	10.3	(1.1)			
			2	Tschechische Republik	3.5	(0.6)	15.6	(0.8)	37.9	(1.1)	34.2	(1.2)	8.7	(0.7)	3.4	(0.5)	15.9	(0.9)	37.9	(1.9)	34.5	(1.9)	8.3	(0.9)					
				Australien	6.3	(0.8)	18.3	(0.9)	– 36.6	(1.0)	31.0	(1.2)	7.9	(0.7)	+	6.2	(0.7)	15.5	(1.0)	34.4	(1.3)	32.8	(1.3)	11.1	(0.9)				
2	3	2		3	»Hongkong	1.9	(0.4)	+ 10.3	(0.9)	+ 32.4	(1.4)	+ 39.3	(1.3)	– 16.2	(1.2)	–	3.9	(0.6)	16.8	(1.4)	38.3	(1.3)	32.5	(1.6)	8.5	(0.9)			
				Ungarn	5.6	(0.9)	13.9	(1.3)	+ 30.2	(1.1)	+ 36.1	(1.2)	– 14.2	(1.1)	–	5.7	(0.7)	17.9	(1.0)	34.5	(1.1)	32.2	(1.0)	9.8	(0.6)				
3	2	3	2	3	»Dänemark	4.2	(0.5)	17.7	(1.0)	39.4	(1.3)	32.1	(1.6)	6.6	(0.6)	4.5	(0.5)	19.2	(1.1)	40.8	(1.1)	29.9	(1.0)	5.7	(0.7)				
			2	»Österreich												5.9	(0.8)	19.4	(1.0)	37.1	(1.1)	30.3	(1.1)	7.2	(0.7)				
			2	Slowakei	9.0	(0.8)	16.8	(0.9)	34.6	(1.1)	+ 30.8	(1.2)	8.7	(0.6)	8.0	(1.2)	15.5	(1.1)	37.9	(1.1)	31.7	(1.2)	6.9	(0.8)					
3	3		3	»Niederlande	3.3	(0.6)	20.5	(1.2)	45.9	(1.1)	27.5	(1.4)	2.8	(0.4)	4.2	(0.6)	20.1	(1.4)	43.0	(1.3)	28.8	(1.8)	3.9	(0.9)					
				Deutschland	4.0	(0.6)	+ 17.6	(1.0)	+ 38.9	(1.0)	– 32.0	(1.6)	7.6	(0.6)	7.2	(0.7)	20.5	(1.0)	35.9	(1.0)	29.6	(1.3)	6.9	(0.9)					
2		3	2	3	»Italien	5.2	(0.7)	20.1	(1.3)	+ 42.8	(1.3)	27.9	(1.2)	4.0	(0.5)	5.0	(0.8)	23.9	(1.1)	43.6	(1.1)	24.2	(1.7)	3.3	(0.7)				
1	2	1	3	6	1	3	1	2	»Neuseeland	11.9	(0.9)	21.2	(1.0)	+ 34.7	(1.0)	25.9	(0.9)	6.3	(0.6)	11.6	(0.8)	24.0	(0.9)	34.5	(1.4)	24.3	(1.1)	5.6	(0.5)
2	2		2	2	Georgien	26.0	(1.7)	32.9	(1.1)	29.4	(1.3)	10.4	(1.0)	1.4	(0.6)	25.1	(1.8)	32.2	(1.1)	30.9	(1.7)	10.3	(1.1)	1.4	(0.4)				
				»Iran	38.6	(1.7)	– 28.2	(1.2)	24.5	(1.4)	8.0	(0.7)	+	0.8	(0.3)	31.8	(1.8)	28.5	(1.1)	26.8	(1.1)	11.5	(0.9)	1.3	(0.3)				
				Benchmark-Teilnehmer																									
2	3	6	2		2	»Dubai, VAE	13.8	(0.5)	– 16.6	(0.8)	– 27.6	(0.8)	28.0	(0.8)	+ 14.0	(0.6)	+	6.2	(0.5)	13.5	(0.6)	28.4	(0.7)	35.3	(1.0)	16.5	(0.8)		
2	2				2	»Ontario, Kanada	4.4	(0.6)	17.0	(1.0)	37.3	(1.3)	32.7	(1.1)	8.5	(0.9)	5.3	(0.8)	19.9	(1.6)	36.7	(1.5)	30.1	(1.3)	7.9	(1.0)			
2			2	3	3	»Québec, Kanada	3.0	(0.6)	19.1	(1.7)	42.6	(1.3)	29.5	(1.9)	5.8	(0.9)	3.4	(0.5)	19.8	(1.5)	42.6	(1.6)	29.0	(1.3)	5.3	(0.7)			

+ = Prozentwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)– = Prozentwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

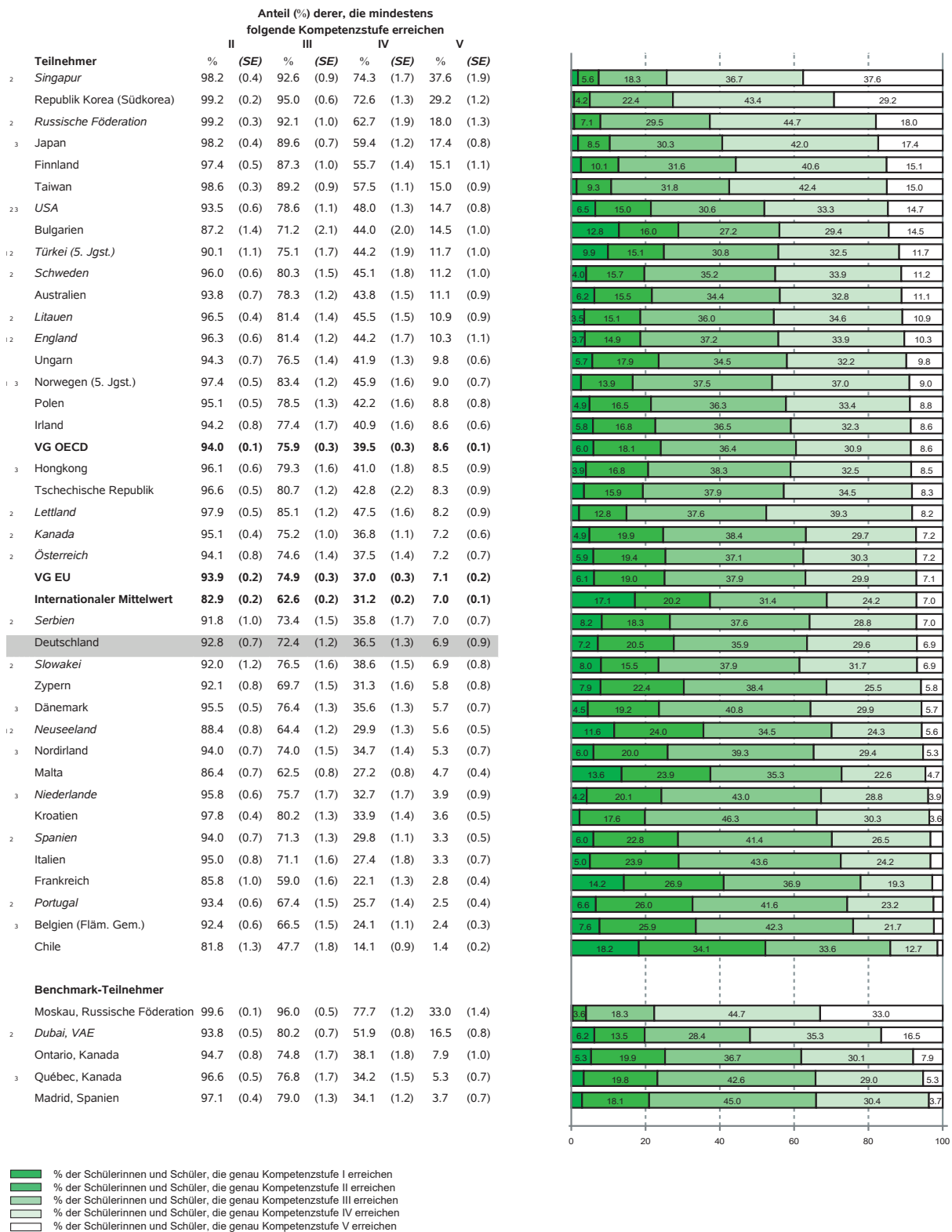
3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.

6 = Abweichender Testzeitpunkt

8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

auf ein ungünstiges Bild hin. So scheint es etwas schlechter zu gelingen, die naturwissenschaftliche Kompetenz leistungsschwacher Schülerinnen und Schüler zu fördern. Auch in den meisten anderen Staaten zeigen sich im Hinblick auf die Verteilung auf die Kompetenzstufen nur geringfügige Verschiebungen von wenigen Prozenten. Staaten, in denen sich substantielle Änderungen in der Verteilung zeigen, sind vor allem solche, in denen sich auch das Niveau substantiell verändert hat. Hierzu gehören beispielsweise Bahrain und Zypern, wo circa 10 Prozent mehr Schülerinnen und Schüler als in den Studienzyklen davor die Kompetenzstufe IV (hoch) erreichen (Mullis et al., 2020).

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

4.5.3 Inhaltsbereiche

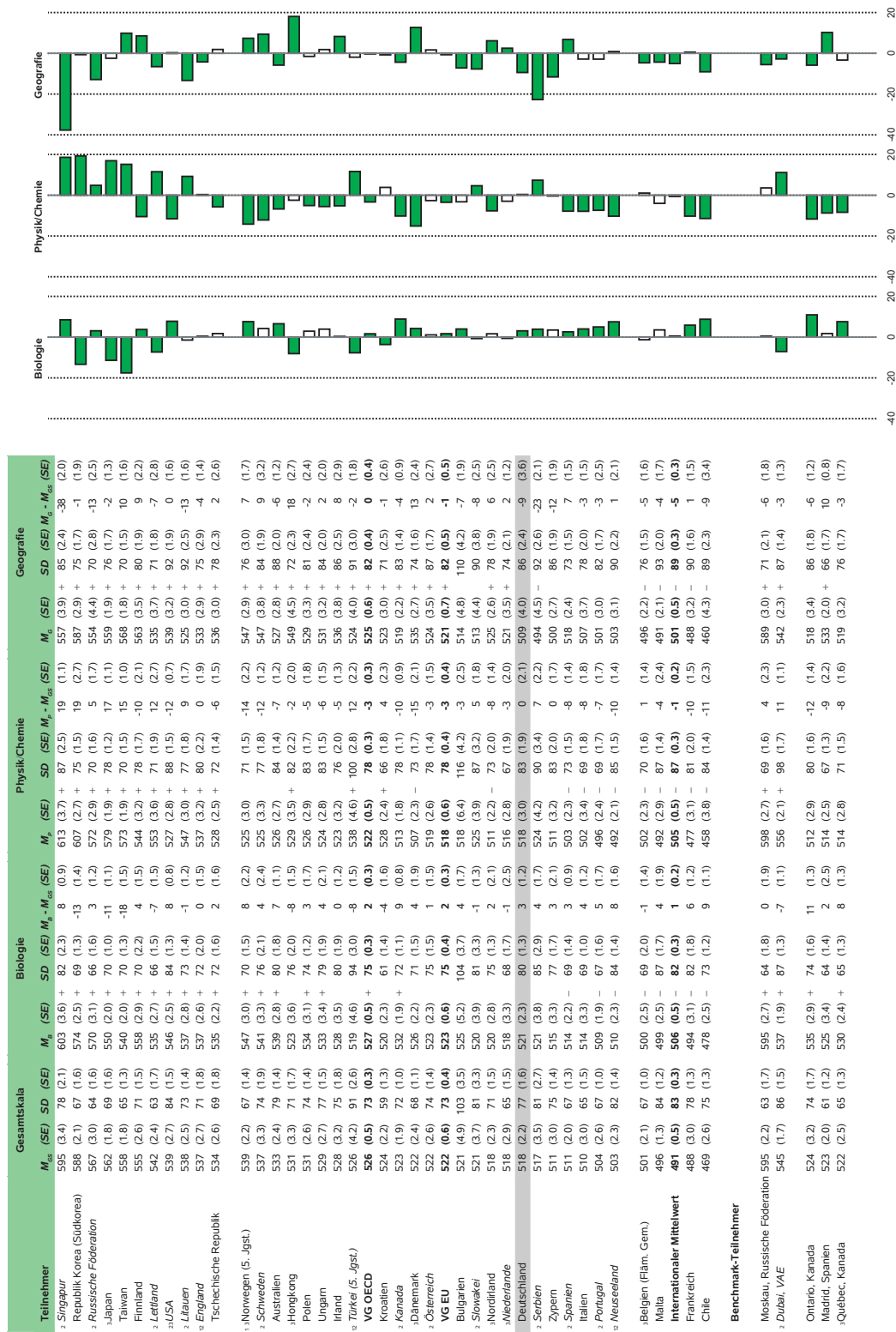
Die naturwissenschaftliche Kompetenz wird in der TIMSS-Rahmenkonzeption in die drei Inhaltsbereiche *Biologie*, *Physik/Chemie* und *Geografie* differenziert. Dies erlaubt einen genaueren Blick auf die Stärken und Schwächen der Schülerinnen und Schüler in den drei Bereichen, die in der Sekundarstufe durch die (teils zusammengefassten) Fächer Biologie, Geografie und Physik und Chemie abgebildet werden. Aus der nationalen Perspektive ist diese Differenzierung auch wichtig, um abzuschätzen, inwiefern die drei Inhaltsbereiche gleichmäßig im Sachunterricht behandelt werden. In diesem Abschnitt werden zunächst die Ergebnisse im internationalen Vergleich und dann die Unterschiede zwischen den Studienzyklen berichtet.

Kompetenzen in den drei Inhaltsbereichen im internationalen Vergleich: In Abbildung 4.8 werden die Kompetenzwerte aufgeschlüsselt nach den Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geografie* sowie die dazugehörigen Standardabweichungen dargestellt. Aufgeführt sind hier nur OECD- und EU-Staaten sowie Staaten, die eine signifikant höhere oder gleich hohe Kompetenz auf der Gesamtskala wie Deutschland aufweisen. Die Anordnung entspricht der Rangfolge auf der Gesamtskala Naturwissenschaften.

Vergleicht man jeweils die Ergebnisse in den drei Inhaltsbereichen mit der Gesamtskala Naturwissenschaften in den teilnehmenden Staaten, zeigt sich kein einheitliches Muster. In der überwiegenden Zahl der Staaten zeigen sich einzelne positive und negative Unterschiede, die von 1 bis 38 Punkten reichen. Generelle Tendenzen, dass zum Beispiel im Inhaltsbereich *Biologie* höhere Werte erzielt werden als im Bereich *Physik/Chemie*, sind nicht erkennbar. Lediglich im Bereich *Geografie* ist erkennbar, dass hier eine etwas größere Anzahl von Staaten (18) eine relative Schwäche aufweist. Die Unterschiede lassen sich eher auf Schwerpunkte in den Bildungsvorgaben der einzelnen Staaten zurückführen. Für Deutschland zeigt sich, dass im Vergleich zur Gesamtskala Naturwissenschaften die mittlere Leistung im Inhaltsbereich *Biologie* um 3 Punkte erhöht ist (521 Punkte). Dieser Unterschied ist (auch wenn er signifikant ist) als gering einzuschätzen. Im Inhaltsbereich *Physik/Chemie* zeigt sich dagegen kein Unterschied zur Gesamtskala (518 Punkte), während die Leistungen im Inhaltsbereich *Geografie* um 9 Punkte niedriger (509 Punkte) sind. Die etwas größere Standardabweichung ($SD = 86$) in diesem Inhaltsbereich deutet zudem auf etwas größere Unterschiede in den Leistungen der Schülerinnen und Schüler hin. Dies liegt vermutlich daran, dass, wie eingangs erläutert, naturgeografische Themen im Sachunterricht weniger in den Lehrplänen und damit auch im Unterricht verankert sind. Die mittleren Kompetenzwerte der EU- und OECD-Staaten in den drei Inhaltsbereichen deuten auf keine ausgeprägten relativen Stärken oder Schwächen hin. Es gibt eine leichte relative Stärke im Bereich der *Biologie* beziehungsweise Schwäche im Bereich der *Physik/Chemie* und im Bereich der *Geografie*. Die Unterschiede zwischen der Gesamtskala Naturwissenschaften und den einzelnen Inhaltsbereichen betragen in beiden Vergleichsgruppen aber maximal 4 Punkte.

Unterschiede in den Kompetenzen in den naturwissenschaftlichen Inhaltsbereichen zwischen TIMSS 2007, 2011, 2015 und 2019: In der Tabelle 4.11 und der Abbildung 4.9 sind die mittleren Kompetenzwerte in den naturwissenschaftlichen Inhaltsbereichen aus TIMSS 2007 bis TIMSS 2019 beziehungsweise Unterschiede zwischen den Kompetenzwerten über die Studienzyklen hin-

Abbildung 4.8: Testleistungen der Schülerinnen und Schüler in Naturwissenschaften im internationalen Vergleich in den Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geografie*



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

Keine statistisch signifikanten Unterschiede zur Gesamtskala ($p > .05$)
 Statistisch signifikante Unterschiede zur Gesamtskala ($p \leq .05$)
 + = Mittelwert statistisch signifikant höher als der Mittelwert von Deutschland ($p \leq .05$)
 - = Mittelwert statistisch signifikant niedriger als der Mittelwert von Deutschland ($p \leq .05$)

IEA: Trends in International Mathematics and Science Study

Tabelle 4.11: Vergleich der Testleistungen in Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 in den Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geografie* I

Fußnoten				Biologie				Physik/Chemie				Geografie				
2007	2011	2015	2019 T	Teilnehmer ^A	2007 ^B	2011 ^B	2015	2019	2007 ^B	2011 ^B	2015	2019	2007 ^B	2011 ^B	2015	2019
					<i>M_{or}</i> (SE)	<i>M₁₁</i> (SE)	<i>M₁₅</i> (SE)	<i>M₁₉</i> (SE)	<i>M_{or}</i> (SE)	<i>M₁₁</i> (SE)	<i>M₁₅</i> (SE)	<i>M₁₉</i> (SE)	<i>M_{or}</i> (SE)	<i>M₁₁</i> (SE)	<i>M₁₅</i> (SE)	<i>M₁₉</i> (SE)
2	2	2	2	Georgien	421 (4.0)	461 (3.6)	459 (4.1)	457 (4.0)	403 (4.9)	440 (4.2)	438 (4.7)	452 (4.6)	416 (5.4)	458 (4.3)	441 (4.3)	435 (4.2)
2	2	2	2	^a Litauen	518 (2.2)	520 (2.9)	529 (3.1)	537 (2.8)	511 (2.0)	514 (3.1)	538 (2.8)	547 (3.0)	508 (2.9)	501 (3.0)	517 (3.9)	525 (3.0)
2	2	2	2	^a Russische Föderation	545 (4.7)	556 (3.6)	569 (3.1)	570 (3.1)	552 (5.4)	548 (4.0)	567 (3.6)	572 (2.9)	541 (5.5)	552 (4.1)	562 (4.7)	554 (4.4)
2	2	2	2	Tschechische Republik	522 (3.5)	550 (3.0)	538 (2.0)	535 (2.2)	509 (3.4)	519 (3.1)	531 (2.4)	528 (2.5)	514 (3.5)	537 (3.4)	531 (3.0)	536 (3.0)
				^{3 a} Japan	536 (2.3)	540 (1.9)	556 (2.2)	550 (2.0)	571 (2.9)	589 (1.9)	587 (2.6)	579 (1.9)	532 (3.5)	551 (1.8)	563 (2.5)	559 (1.9)
				² Schweden	532 (2.8)	534 (2.7)	540 (3.3)	541 (3.3)	509 (3.2)	528 (2.5)	534 (3.6)	525 (3.3)	539 (3.7)	538 (3.2)	552 (4.1)	547 (3.8)
				² Singapur	595 (4.7)	597 (4.3)	607 (4.4)	603 (3.6)	597 (4.3)	598 (3.5)	603 (3.7)	613 (3.7)	565 (4.0)	541 (3.0)	546 (3.7)	557 (3.9)
				^{3 a} Dänemark	527 (3.2)	530 (2.8)	534 (2.4)	526 (2.2)	502 (2.8)	526 (2.5)	516 (2.7)	507 (2.3)	519 (3.3)	527 (3.0)	531 (3.0)	535 (2.7)
				Australien	529 (3.6)	516 (3.1)	531 (3.0)	539 (2.8)	521 (3.8)	514 (3.2)	516 (2.7)	526 (2.7)	536 (4.5)	520 (3.5)	520 (3.3)	527 (2.8)
				^a Iran	437 (5.2)	449 (4.1)	417 (4.5)	430 (4.5)	440 (4.9)	453 (4.0)	423 (5.0)	453 (4.7)	416 (5.0)	457 (3.5)	408 (4.8)	438 (4.2)
				Taiwan	547 (2.8)	538 (2.4)	545 (2.0)	540 (2.0)	564 (2.5)	569 (2.0)	568 (2.0)	573 (1.9)	563 (2.9)	553 (2.5)	555 (2.5)	568 (1.8)
2 3	2 3	2 3	2 3	USA	544 (2.9)	547 (2.1)	555 (2.3)	546 (2.5)	535 (3.1)	544 (2.0)	537 (2.6)	527 (2.8)	537 (3.2)	539 (2.1)	539 (2.4)	539 (3.2)
1 2	1 3 6 1	3	1 2	^a Neuseeland	506 (2.8)	497 (2.5)	511 (2.7)	510 (2.3)	494 (3.4)	493 (2.7)	497 (2.5)	492 (2.1)	513 (3.4)	499 (3.2)	506 (3.4)	503 (3.1)
				² Österreich	528 (2.4)	526 (2.8)		523 (2.3)	517 (3.0)	535 (2.8)		519 (2.6)	535 (2.6)	539 (3.5)		524 (3.5)
1 3	1 3	1	1 2	^a England	536 (3.1)	530 (2.8)	536 (2.5)	537 (2.6)	546 (3.3)	535 (3.5)	540 (2.7)	537 (3.2)	542 (3.4)	522 (3.8)	527 (3.3)	533 (2.9)
3	3	3	3	^{3 a} Niederlande	539 (2.7)	537 (1.8)	525 (2.7)	518 (3.3)	503 (3.1)	526 (2.0)	504 (2.6)	516 (2.8)	524 (3.3)	525 (2.7)	520 (3.0)	521 (3.5)
				² Slowakei	535 (4.6)	534 (3.5)	517 (2.9)	520 (3.9)	512 (5.2)	527 (4.0)	526 (3.4)	525 (3.9)	532 (6.5)	535 (3.8)	514 (3.0)	513 (4.4)
				Ungarn	553 (3.3)	552 (3.5)	550 (3.4)	533 (3.4)	529 (3.7)	520 (3.8)	534 (3.5)	524 (2.8)	517 (4.3)	524 (4.4)	535 (4.0)	531 (3.2)
2 3	2	3	3	^{3 a} Deutschland	531 (2.2)	525 (2.6)	528 (2.0)	521 (2.3)	527 (3.2)	535 (3.1)	532 (2.5)	518 (3.0)	524 (2.8)	520 (3.7)	519 (4.0)	509 (4.0)
2	3	2 3	2 3	^a Hongkong	540 (4.0)	524 (3.7)	550 (3.7)	523 (3.6)	562 (4.0)	539 (4.4)	555 (3.5)	529 (3.5)	568 (4.2)	548 (3.3)	574 (3.1)	549 (4.5)
				^a Italien	555 (3.6)	535 (2.7)	519 (2.7)	514 (3.3)	520 (3.7)	509 (3.0)	513 (2.9)	502 (3.4)	527 (4.1)	523 (3.6)	510 (3.5)	507 (3.7)
Benchmark-Teilnehmer																
2 3 6	2	2	2	^a Dubai, VAE	456 (2.8)	455 (2.9)	518 (2.6)	537 (1.9)	456 (3.5)	460 (3.2)	521 (2.2)	556 (2.1)	461 (3.8)	469 (3.0)	510 (2.9)	542 (2.3)
2	2 3	2 3	2 3	^{3 a} Québec, Kanada	524 (3.0)	524 (2.5)	533 (4.3)	530 (2.4)	509 (3.1)	507 (3.1)	519 (4.9)	514 (2.8)	522 (3.0)	516 (3.5)	515 (4.4)	519 (3.2)
2	2	2	2	^a Ontario, Kanada	539 (4.0)	535 (3.4)	544 (2.6)	535 (2.9)	535 (3.3)	528 (3.2)	522 (2.5)	512 (2.9)	533 (4.2)	514 (3.9)	515 (3.7)	518 (3.4)

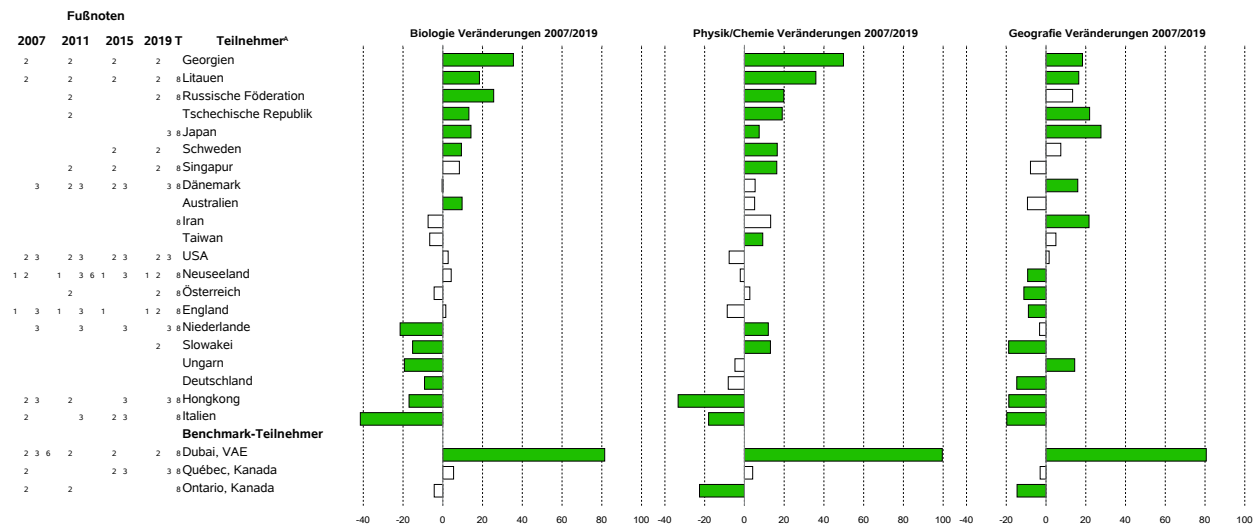
1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfunggrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnehmerquoten auf Schul- und/oder Schullebene erreichen nicht die internationalen Vorgaben.
 6 = Abweichender Testzeitpunkt
 A = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019.
 B = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.
 B = Abweichungen in den berichteten Standardfehlern zur internationalen Berichterstattung sind in einem anderen Berechnungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Abbildung 4.9: Vergleich der Testleistungen in Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 in den Inhaltsbereichen *Biologie, Physik/Chemie und Geografie II*

Fußnoten					Teilnehmer ^A	Biologie Veränderungen ^B			Physik/Chemie Veränderungen ^B			Geografie Veränderungen ^B		
2007	2011	2015	2019	T		$M_{2007}-M_{2019}$ (SE)	$M_{2011}-M_{2019}$ (SE)	$M_{2015}-M_{2019}$ (SE)	$M_{2007}-M_{2019}$ (SE)	$M_{2011}-M_{2019}$ (SE)	$M_{2015}-M_{2019}$ (SE)	$M_{2007}-M_{2019}$ (SE)	$M_{2011}-M_{2019}$ (SE)	$M_{2015}-M_{2019}$ (SE)
2	2	2	2	2	Georgien	36 (5.7) +	-4 (5.4)	-2 (5.8)	50 (6.7) +	12 (6.2) +	15 (6.6) +	18 (6.8) +	-23 (6.0) -	-6 (6.0)
2	2	2	2	2	Litauen	18 (3.5) +	16 (4.0) +	7 (4.2)	36 (3.6) +	33 (4.3) +	10 (4.1) +	16 (4.2) +	24 (4.2) +	8 (4.9)
				2	Russische Föderation	26 (5.7) +	14 (4.8) +	2 (4.4)	20 (6.1) +	24 (5.0) +	5 (4.7)	13 (7.0) +	2 (6.0)	-8 (6.4)
				2	Tschechische Republik	13 (4.1) +	-14 (3.7) -	-3 (3.0)	19 (4.3) +	9 (4.0) +	-3 (3.5)	22 (4.6) +	-2 (4.5)	4 (4.2)
				3	Japan	14 (3.0) +	10 (2.8) +	-6 (3.0) -	7 (3.4) +	-11 (2.7) -	-8 (3.2) -	28 (4.0) +	8 (2.6) +	-4 (3.2)
				2	Schweden	9 (4.3) +	8 (4.3)	2 (4.7)	17 (4.6) +	-3 (4.1)	-9 (4.9)	7 (5.3)	8 (5.0)	-5 (5.6)
				2	Singapur	8 (6.0)	6 (5.7)	-4 (5.7)	16 (5.7) +	15 (5.1) +	10 (5.3)	-8 (5.6)	16 (4.9) +	10 (5.4)
3	2, 3	2, 3	2, 3	3	Dänemark	0 (3.9)	-4 (3.5)	-8 (3.3) -	5 (3.6)	-19 (3.4) -	-9 (3.6) -	16 (4.3) +	8 (4.1) +	4 (4.0)
					Australien	10 (4.5) +	23 (4.2) +	8 (4.1)	5 (4.7)	12 (4.2) +	10 (3.8) +	-9 (5.2)	7 (4.5)	7 (4.3)
					Iran	-7 (6.9)	-19 (6.1) -	13 (6.4) +	13 (6.8)	0 (6.2)	29 (6.9) +	22 (6.5) +	-19 (5.4) -	30 (6.3) +
					Taiwan	-7 (3.5)	2 (3.2)	-4 (2.9)	9 (3.2) +	5 (2.8)	5 (2.7)	5 (3.4)	15 (3.1) +	13 (3.1) +
2, 3	2, 3	2, 3	2, 3	2, 3	USA	3 (3.8)	-1 (3.3)	-9 (3.4) -	-8 (4.2)	-17 (3.4) -	-10 (3.8) -	2 (4.6)	0 (3.9)	-1 (4.0)
1, 2	1, 3, 6, 1	3	1, 2	1, 2	Neuseeland	4 (3.6)	13 (3.4) +	-1 (3.6)	-2 (4.0)	-1 (3.4)	-5 (3.3)	-9 (4.6) -	5 (4.5)	-2 (4.6)
				2	Österreich	-4 (3.3)	-3 (3.6)		3 (4.0)	-16 (3.8) -		-11 (4.3) -	-15 (4.9) -	
1, 3	1, 3, 1	1, 2	1, 2	1, 2	England	2 (4.0)	7 (3.9)	1 (3.6)	-9 (4.6)	2 (4.8)	-3 (4.2)	-9 (4.5)	-11 (4.8) +	5 (4.4)
				3	Niederlande	-21 (4.2) -	-19 (3.8) -	-7 (4.2)	12 (4.2) +	-10 (3.4) -	12 (3.8) +	-3 (4.8)	-4 (4.4)	1 (4.6)
				2	Slowakei	-15 (6.1) -	-14 (5.2) -	3 (4.9)	13 (6.5) +	-2 (5.6)	0 (5.2)	-19 (7.9) -	-22 (5.8) -	-1 (5.3)
					Ungarn	-19 (4.8) -	-18 (4.9) -	-17 (4.8) -	-5 (4.6)	3 (4.7)	-10 (4.5) -	14 (5.3) +	8 (5.4)	-4 (5.1)
					Deutschland	-9 (3.2) -	-4 (3.5)	-7 (3.1) -	-8 (4.4)	-17 (4.3) -	-14 (4.0) -	-15 (4.9) -	-11 (5.4) -	-10 (5.6)
2, 3	2	3	3	3	Hongkong	-17 (5.4) -	-1 (5.1)	-27 (5.2) -	-33 (5.3) -	-10 (5.6)	-26 (5.0) -	-19 (6.2) -	1 (5.5)	-25 (5.4) -
2	3	2, 3	2, 3	2, 3	Italien	-41 (4.9) -	-21 (4.2) -	-5 (4.3)	-18 (5.1) -	-7 (4.6)	-11 (4.5) -	-20 (5.6) -	-16 (5.2) -	-3 (5.1)
					Benchmark-Teilnehmer									
2, 3, 6	2	2	2	2	Dubai, VAE	81 (3.4) +	82 (3.5) +	20 (3.2) +	100 (4.1) +	96 (3.9) +	35 (3.1) +	81 (4.4) +	73 (3.8) +	31 (3.7) +
2	2	2, 3	2, 3	2, 3	Québec, Kanada	5 (3.8)	5 (3.5)	-3 (4.9)	4 (4.2)	7 (4.2)	-6 (5.6)	-3 (4.4)	3 (4.7)	4 (5.4)
2	2				Ontario, Kanada	-4 (4.9)	0 (4.5)	-9 (3.9) -	-23 (4.4) -	-15 (4.3) -	-10 (3.9) -	-15 (5.4) -	4 (5.2)	3 (5.0)



- Keine statistisch signifikanten Veränderungen zwischen 2007 und 2019 ($p > .05$)
 ■ Statistisch signifikante Veränderungen zwischen 2007 und 2019 ($p \leq .05$)
 + = Mittelwert in 2019 statistisch signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)
 - = Mittelwert in 2019 statistisch signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)
 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 6 = Abweichender Testzeitpunkt
 8 = Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019
 A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.
 B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

weg dargestellt. In beiden Abbildungen sind nur Staaten aufgeführt, die an allen vier Studienzyklen teilgenommen haben. Als weiterer Staat ist Österreich aufgenommen, weil hier ähnliche Veränderungen im Bildungssystem wie in Deutschland in den letzten 15 Jahren zu verzeichnen sind (siehe Kapitel 2 in diesem Band). So wurden beispielsweise in Österreich 2008 Bildungsstandards für Deutsch und Mathematik in Klasse 4 eingeführt. Positive Werte in den Spalten ‚Veränderungen‘ in Abbildung 4.9 bedeuten, dass im erstgenannten Studienzyklus mehr Punkte erzielt wurden. Im unteren Teil der Abbildung 4.9 folgt eine grafische Darstellung mittels Differenzbalken. Während weiße Balken statistisch nicht signifikante Veränderungen repräsentieren, weisen die grün markierten Balken auf statistisch bedeutsame Unterschiede zwischen den Erhebungszyklen hin.

Die Abnahme der mittleren Kompetenzwerte der Schülerinnen und Schüler aus Deutschland auf der Gesamtskala Naturwissenschaften spiegelt sich in allen drei Inhaltsbereichen wider. So sind die Kompetenzwerte für *Biologie*, *Physik/Chemie* und *Geografie* aus TIMSS 2019 um 7, 14 beziehungsweise 10 Punkte niedriger als in TIMSS 2015. Statistisch bedeutsam sind allerdings nur die Veränderungen in den ersten beiden Inhaltsbereichen. Am auffälligsten ist dabei der Abfall im Bereich *Physik/Chemie*, dem Bereich, in dem die Schülerinnen und Schüler aus Deutschland zumindest bei TIMSS 2015 und 2011 eine relative Stärke zeigten. Der Bereich *Geografie* war auch in den vorherigen Studienzyklen immer der (relativ betrachtet) schwächste Inhaltsbereich, diese Reihung ändert sich in TIMSS 2019 nicht. Auch der längerfristige Vergleich mit TIMSS 2007 zeigt die Abnahme der mittleren Leistungen um 8 bis 17 Punkte in allen drei Inhaltsbereichen, die allerdings nur in den Bereichen *Biologie* und *Geografie* signifikant ist.

Ähnlich wie bei den Niveauunterschieden in den einzelnen Inhaltsbereichen (siehe Abbildung 4.8) zeigt sich auch bei der Betrachtung der Veränderungen über die letzten vier Studienzyklen kein einheitliches Muster. In den meisten Staaten sind die relativen Schwächen und Stärken eher klein und ähnlich in den verschiedenen Studienzyklen. Die Teilnehmerstaaten, deren mittlere naturwissenschaftliche Kompetenz sich seit TIMSS 2015 sehr auffällig gesteigert hat, wie Malta, Zypern und der Iran, zeigen diese positive Entwicklung auch in den drei einzelnen Inhaltsbereichen (Mullis et al., 2020). Große positive oder negative Veränderungen, die nur einen Inhaltsbereich betreffen, sind dagegen nicht erkennbar.

4.5.4 Kognitive Anforderungsbereiche

Neben den Inhaltsbereichen wird in der TIMSS-Rahmenkonzeption auch zwischen den drei kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* differenziert. Dies erlaubt einen genaueren Einblick in die Art des Wissens, das die Schülerinnen und Schüler im naturwissenschaftlichen Unterricht entwickeln. Dieser Abschnitt ist analog zu dem vorherigen Abschnitt 4.5.3 zu den drei Inhaltsbereichen gegliedert. Es werden zuerst die Ergebnisse zu den kognitiven Anforderungsbereichen im internationalen Vergleich und dann im Vergleich zu den vorherigen Studienzyklen berichtet.

Kompetenzen in den kognitiven Anforderungsbereichen im internationalen Vergleich: In Abbildung 4.10 sind die Kompetenzwerte aufgeschlüsselt nach den Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* sowie die Standardabweichungen dargestellt. Aufgeführt sind auch hier nur OECD- und EU-Staaten sowie Staaten, die eine signifikant höhere oder gleich hohe Kompetenz auf der Gesamtskala wie Deutschland aufweisen. Die Anordnung entspricht der Rangfolge auf der Gesamtskala Naturwissenschaften. Für Deutschland zeigen sich keine relativen Stärken oder Schwächen in einem der Anforderungsbereiche. Die Werte im Anforderungsbereich *Reproduzieren* sind um einen Punkt höher im Vergleich zur Gesamtskala Naturwissenschaften, die in den anderen beiden Anforderungsbereichen *Anwenden* und *Problemlösen* etwas niedriger (2 Punkte) beziehungsweise gleich (0 Punkte). Die Unterschiede sind statistisch nicht bedeutsam. Ein ähnliches Bild zeigt sich auch für die

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



Vergleichsgruppe der EU- und OECD-Teilnehmerstaaten. Auch hier zeigen sich keine relativen Stärken und Schwächen in den drei Anforderungsbereichen.

Die Staaten, die in einem Leistungsbereich wie Deutschland liegen, zeigen überwiegend auch eher geringe bis moderate Unterschiede zwischen den drei Anforderungsbereichen, zum Beispiel 20 Punkte in Serbien und 9 Punkte in den Niederlanden. Auch in den meisten anderen Teilnehmerstaaten finden sich in den drei Anforderungsbereichen allenfalls kleinere positive oder negative relative Verschiebungen von 1 bis 10 Punkten im Vergleich zu der jeweiligen Gesamtskala. Eine gewisse Ausnahme stellt Japan dar, hier zeigt sich ein Unterschied von 45 Punkten zwischen dem stärksten Anforderungsbereich (*Problemlösen*) und dem schwächsten Anforderungsbereich (*Reproduzieren*). Diese relativen Stärken und Schwächen hängen vermutlich mit unterschiedlichen Unterrichtstraditionen in den verschiedenen Teilnehmerstaaten zusammen. Zusammenfassend zeigt sich, dass in den meisten Staaten eine gleichmäßige Förderung der Kompetenzen im Unterricht gelingt, die notwendig sind, um die Aufgaben in den verschiedenen Anforderungsbereichen zu bearbeiten.

Unterschiede in den Kompetenzen in den kognitiven Anforderungsbereichen zwischen TIMSS 2007, 2011, 2015 und 2019: In der Tabelle 4.12 und der Abbildung 4.11 sind die mittleren Kompetenzwerte in den drei kognitiven Anforderungsbereichen aus TIMSS 2007 bis 2019 sowie Unterschiede zwischen den Kompetenzwerten über die Studienzyklen hinweg dargestellt. Wie in Tabelle 4.11 und Abbildung 4.9 sind abgesehen von Österreich nur Staaten aufgeführt, die an allen vier Studienzyklen teilgenommen haben. Positive Differenzwerte in den Spalten ‚Veränderungen‘ im oberen Teil der Abbildung 4.11 bedeuten, dass höhere Kompetenzwerte im jeweils erstgenannten Studienzyklus erzielt wurden. Im unteren Teil von Abbildung 4.11 folgt eine grafische Darstellung mittels Differenzbalken. Während weiße Balken statistisch nicht signifikante Veränderungen repräsentieren, markieren die grünen Balken statistisch bedeutsame Unterschiede zwischen den Erhebungszyklen.

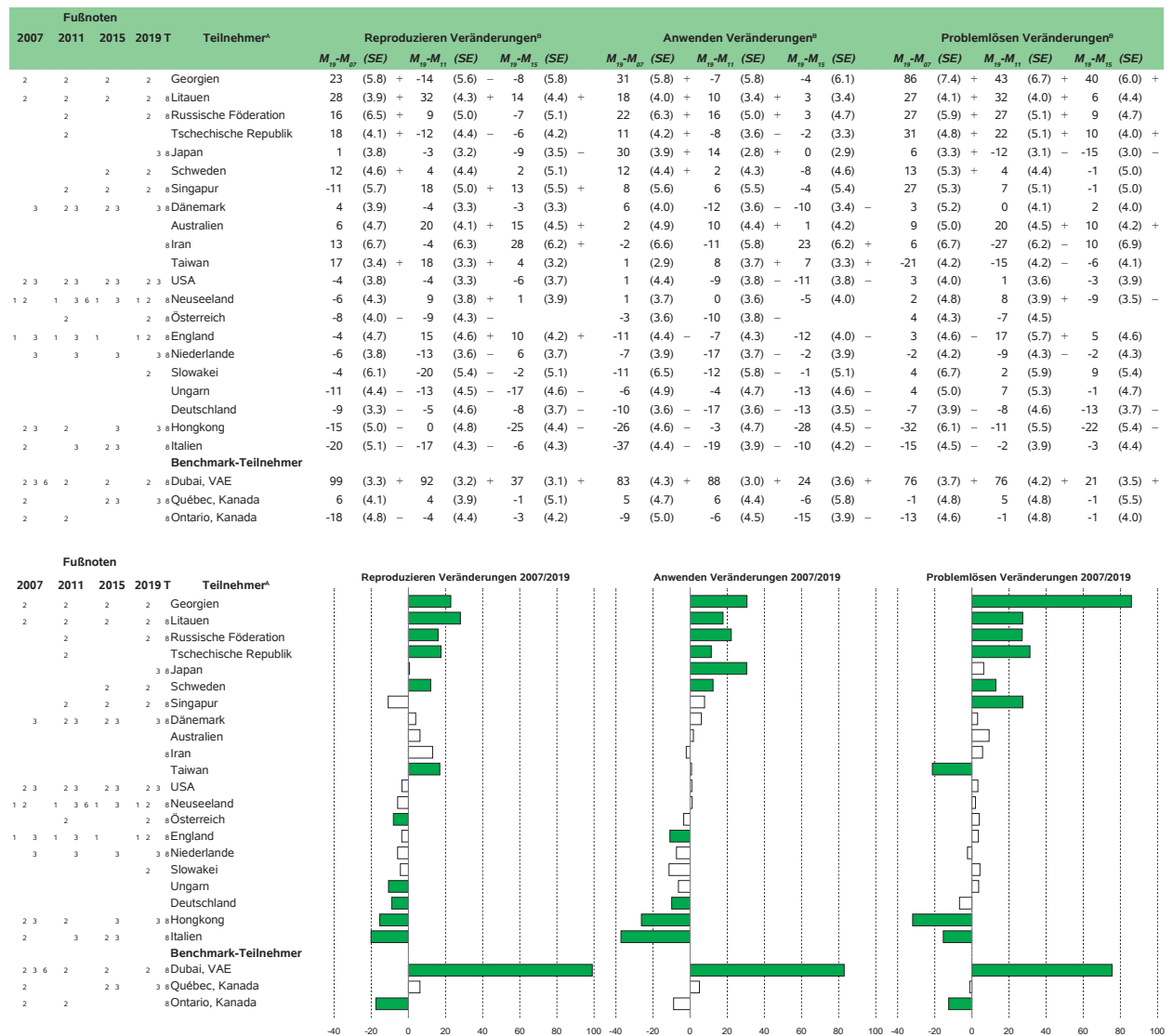
Im Vergleich zu TIMSS 2015 lässt sich Abbildung 4.11 entnehmen, dass Deutschland sich in allen drei Anforderungsbereichen signifikant verschlechtert hat. Wie sich in Abbildung 4.10 bereits andeutet, lässt sich also die generelle Abnahme des mittleren Kompetenzwertes Naturwissenschaften nicht auf eine spezifische Schwäche in einem der Anforderungsbereiche zurückführen. Vergleicht man die Kompetenzwerte des aktuellen Studienzyklus mit denen aus TIMSS 2011, dann zeigt sich eine signifikante Abnahme lediglich im Anforderungsbereich *Anwenden* beziehungsweise im Vergleich zu TIMSS 2007 in den Anforderungsbereichen *Reproduzieren* und *Anwenden*. Auch diese Unterschiede deuten nicht auf systematische Schwächen bei der Wiedergabe von Wissen oder den komplexeren Anforderungen hin. In den anderen Staaten zeigen sich auch keine systematischen Veränderungen zwischen den Studienzyklen.

Tabelle 4.12: Vergleich der Testleistungen in Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 in den kognitiven Anforderungsbereichen *Reproduzieren, Anwenden und Problemlösen*¹

Fußnoten				Reproduzieren								Anwenden								Problemlösen							
2007	2011	2015	2019 T	Teilnehmer ^a		2007 ^B	2011 ^B	2015	2019	2007 ^B	2011 ^B	2015	2019	2007 ^B	2011 ^B	2015	2019										
						M_{op} (SE)	M_{11} (SE)	M_{15} (SE)	M_{19} (SE)	M_{op} (SE)	M_{11} (SE)	M_{15} (SE)	M_{19} (SE)	M_{op} (SE)	M_{11} (SE)	M_{15} (SE)	M_{19} (SE)										
2	2	2	2	2	Georgien	429 (4.3)	466 (3.9)	460 (4.2)	452 (3.9)	415 (4.5)	452 (4.4)	449 (4.8)	445 (3.7)	379 (6.0)	422 (5.0)	426 (4.0)	465 (4.4)										
2	2	2	2	2	Litauen	511 (2.3)	508 (2.9)	526 (3.2)	539 (3.1)	513 (3.3)	521 (2.5)	529 (2.5)	531 (2.3)	521 (2.9)	515 (2.8)	541 (3.3)	548 (2.9)										
2	2	2	2	2	«Russische Föderation	546 (5.6)	553 (3.8)	569 (3.9)	562 (3.3)	550 (5.3)	556 (3.6)	568 (3.3)	572 (3.4)	542 (5.2)	542 (4.2)	561 (3.8)	569 (2.8)										
2	2	2	2	2	Tschechische Republik	521 (2.9)	551 (3.3)	545 (3.0)	538 (2.9)	515 (3.4)	534 (2.6)	528 (2.1)	526 (2.5)	507 (3.6)	516 (4.0)	529 (2.4)	539 (3.2)										
				3	«Japan	534 (2.7)	538 (1.8)	544 (2.3)	535 (2.6)	546 (3.2)	562 (1.6)	576 (1.8)	576 (2.2)	573 (2.3)	591 (2.0)	594 (1.8)	579 (2.4)										
				2	Schweden	528 (3.0)	536 (2.8)	538 (3.8)	540 (3.4)	520 (3.0)	531 (3.0)	540 (3.4)	532 (3.1)	528 (4.3)	537 (3.0)	542 (3.8)	541 (3.2)										
				2	«Singapur	599 (4.4)	570 (3.4)	574 (4.1)	588 (3.7)	587 (4.2)	590 (4.0)	599 (4.0)	595 (3.7)	576 (4.0)	597 (3.8)	605 (3.6)	604 (3.5)										
3	2	3	2	3	«Dänemark	517 (3.3)	524 (2.6)	524 (2.6)	521 (2.0)	513 (3.2)	532 (2.5)	529 (2.4)	519 (2.5)	524 (4.5)	527 (3.1)	526 (2.9)	527 (2.7)										
					Australien	532 (3.6)	517 (2.8)	523 (3.3)	538 (3.0)	522 (3.8)	513 (3.0)	522 (2.7)	524 (3.2)	528 (4.1)	518 (3.4)	527 (3.0)	538 (3.0)										
					«Iran	431 (5.0)	448 (4.3)	416 (4.1)	444 (4.6)	443 (4.9)	452 (3.8)	417 (4.5)	440 (4.3)	427 (4.6)	459 (3.9)	422 (4.9)	432 (4.9)										
					Taiwan	544 (2.8)	542 (2.7)	557 (2.5)	560 (1.9)	560 (2.2)	552 (3.1)	553 (2.6)	561 (2.0)	574 (3.2)	568 (3.2)	558 (3.1)	552 (2.7)										
2	3	2	3	2	3	USA	546 (2.7)	546 (1.9)	548 (2.5)	542 (2.7)	534 (3.1)	544 (2.1)	546 (2.2)	535 (3.1)	535 (3.0)	537 (2.3)	542 (2.7)	538 (2.7)									
1	2	1	3	6	1	«Neuseeland	511 (3.4)	496 (2.7)	504 (2.8)	505 (2.7)	496 (2.7)	497 (2.6)	502 (3.1)	497 (2.6)	503 (4.0)	497 (2.9)	514 (2.4)	505 (2.6)									
					2	«Österreich	531 (2.4)	532 (3.0)	523 (3.1)	527 (2.7)	533 (2.9)	523 (2.4)	523 (3.1)	514 (2.8)	525 (3.1)	518 (3.3)	518 (3.3)										
1	3	1	3	1	1	«England	547 (3.4)	529 (3.2)	533 (2.6)	544 (3.3)	537 (3.2)	532 (3.1)	538 (2.7)	526 (3.0)	540 (2.8)	526 (4.4)	539 (2.7)	544 (3.7)									
3	3	3	3	3	«Niederlande	521 (2.6)	528 (2.3)	508 (2.4)	515 (2.8)	525 (2.4)	534 (2.0)	519 (2.4)	517 (3.1)	526 (2.7)	532 (2.9)	526 (2.9)	523 (3.2)										
				2	Slowakei	531 (4.8)	547 (3.8)	530 (3.3)	527 (3.9)	527 (4.9)	528 (4.0)	517 (2.8)	515 (4.3)	512 (5.3)	514 (4.2)	507 (3.4)	516 (4.2)										
					Ungarn	544 (3.5)	547 (3.7)	550 (3.8)	533 (2.7)	532 (3.8)	530 (3.6)	539 (3.4)	526 (3.1)	528 (4.2)	525 (4.5)	533 (3.9)	532 (2.6)										
						Deutschland	529 (2.4)	524 (4.0)	527 (2.8)	520 (2.3)	526 (2.5)	533 (2.6)	529 (2.4)	516 (2.5)	525 (2.6)	526 (3.6)	532 (2.3)	518 (2.9)									
2	3	2	3	3	«Hongkong	553 (3.9)	537 (3.6)	562 (3.0)	537 (3.2)	552 (3.4)	529 (3.5)	554 (3.3)	526 (3.1)	563 (4.9)	541 (4.2)	552 (4.1)	531 (3.6)										
2	3	2	3	3	«Italien	535 (4.2)	532 (3.0)	521 (3.1)	515 (3.0)	541 (3.4)	523 (2.7)	513 (3.1)	504 (2.7)	523 (3.6)	510 (2.9)	511 (3.5)	508 (2.7)										
Benchmark-Teilnehmer																											
2	3	6	2	2	2	«Dubai, VAE	461 (2.6)	467 (2.5)	523 (2.3)	560 (2.1)	458 (3.7)	453 (2.0)	517 (2.8)	456 (3.0)	455 (3.7)	510 (2.9)	531 (2.1)										
2	2	2	2	2	3	«Québec, Kanada	517 (3.0)	519 (2.7)	524 (4.3)	523 (2.8)	515 (3.0)	514 (2.5)	525 (4.5)	520 (3.6)	526 (3.7)	520 (3.7)	526 (4.6)	525 (3.0)									
2	2	2	2	2	2	«Ontario, Kanada	542 (3.7)	529 (3.1)	527 (2.8)	525 (3.1)	529 (3.9)	526 (3.3)	534 (2.5)	520 (3.1)	540 (3.4)	529 (3.7)	529 (2.8)	528 (3.0)									

- 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
3 = Die Teilnehmerquoten auf Schul- und/oder Schülerschicht erreichen nicht die internationalen Vorgaben.
6 = Abweichender Testzeitpunkt.
8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019.
A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.
B = Abweichungen in den berichteten Standardabweichungen zur internationalen Berichterstattung sind in einem differenzierten Berechnungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

Abbildung 4.11: Vergleich der Testleistungen in Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 in den kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen*

□ Keine statistisch signifikanten Veränderungen zwischen 2007 und 2019 ($p > .05$)

■ Statistisch signifikante Veränderungen zwischen 2007 und 2019 ($p \leq .05$)

+ = Mittelwert in 2019 statistisch signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

- = Mittelwert in 2019 statistisch signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

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

6 = Abweichender Testzeitpunkt

8 = Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

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

4.5.5 Einstellung und Selbstkonzept

Neben den kognitiven Kompetenzen hat der Sachunterricht (wie andere Fächer) auch das Ziel, motivationale Merkmale von Schülerinnen und Schülern wie Interesse, Einstellungen oder fachbezogene Selbstkonzepte zu fördern. Entsprechende multikriteriale Ziele sind auch in den meisten Lehrplänen und dem Perspektivrahmen verankert. Diese motivationalen Merkmale werden in der Grundschule angebahnt und im Verlauf der weiteren Schullaufbahn weiterentwickelt und ausdifferenziert. Sie gelten als zentrale Bedingungsfaktoren für erfolgreiche Bildungsverläufe (Kriegbaum, Becker & Spinath, 2018). Schülerinnen und Schüler, die sich für kompetent in einem Fach halten und die Inhalte als interessant und bedeutsam einschätzen, sind motivierter, sich mit einem Inhalt auseinanderzusetzen und entwickeln wiederum höhere Kompetenzen (Wigfield & Eccles, 2000). Gleichzeitig beeinflusst die Kompetenz, welche fachbezogenen Fähigkeiten sich Schülerinnen und Schüler zutrauen und wie interessiert und motiviert sie sind. Es wird also eine wechselseitige Beziehung zwischen motivationalen Merkmalen und kognitiven Kompetenzen angenommen.

Im Rahmen der TIMSS-Konzeption werden die Einstellung gegenüber den Naturwissenschaften und das naturwissenschaftsbezogene Selbstkonzept als zwei zentrale motivationale Merkmale aufgegriffen. Einstellungen sind affektive Bewertungen der Naturwissenschaften (Osborne, Simon & Collins, 2003), während unter dem naturwissenschaftsbezogenen Selbstkonzept die Selbsteinschätzung der eigenen Leistungsstärke in diesem Bereich verstanden wird (Möller & Trautwein, 2009). Diese Selbsteinschätzungen beruhen auf Kompetenzerfahrungen, die jemand in einem Schulfach oder einem Kompetenzbereich macht, dabei spielen soziale Vergleichsprozesse mit Mitschülerinnen und Mitschülern eine wichtige Rolle.

In diesem Abschnitt werden die Ergebnisse aus TIMSS 2019 zu beiden Merkmalen der Schülerinnen und Schüler aus Deutschland dargestellt und mit den Ergebnissen aus den vorherigen Studienzyklen verglichen. Auf den internationalen Vergleich wird an dieser Stelle verzichtet, da bei Einstellungs- und Selbstkonzeptskalen kulturspezifische Referenzgruppeneffekte zum Tragen kommen (Seaton, Marsh & Craven, 2009; Vande Gaer, Grisay, Schulz & Gebhardt, 2012), die dazu führen, dass die verwendeten Messmodelle für internationale Vergleiche nicht universell gültig sind. Mittelwertunterschiede zwischen Staaten lassen sich so nicht valide interpretieren.

Die Herausforderung bei der Erhebung von naturwissenschaftsbezogener Einstellung und Selbstkonzept von Grundschülerinnen und Grundschülern in Deutschland ist, dass die Kinder zwar den Sachunterricht kennen, aber möglicherweise nicht zwischen dem Sachunterricht als Fach mit verschiedenen Inhaltsbereichen und den naturwissenschaftlichen Anteilen differenzieren. Sachunterrichtslehrkräfte benennen in der Regel die einzelnen Perspektiven nicht gezielt oder grenzen diese voneinander ab, da das Selbstverständnis des Sachunterrichts ein perspektivenintegrierender Ansatz ist. Fragt man lediglich nach den Naturwissenschaften, ist wiederum unklar, ob alle Schülerinnen und Schüler eine klare Vorstellung von den dazugehörigen Inhalten und Vorgehensweisen haben. Die Situation in anderen Staaten, in denen Schülerinnen und Schüler zum Beispiel das Fach *Science* kennen, stellt sich anders dar. Aus diesem Grund wurden die Schülerinnen und Schüler in Deutschland vor den Fragen zu Einstellungen und Selbstkonzept auf die *naturwissenschaftlichen* Anteile des Sachunterrichts hingewiesen, in dem ihnen konkrete naturwissenschaftliche

Themen, zum Beispiel Tiere und Pflanzen, Wetter, elektrischer Strom und Licht und Schatten, genannt wurden. Dieses Vorgehen ist auch aus anderen Studien bekannt (Kleickmann et al., 2020). Zudem ist anzunehmen, dass die Schülerinnen und Schüler durch den naturwissenschaftlichen Kompetenztest, den sie vor dieser Befragung bearbeiten, ein gewisses *Priming* auf die Naturwissenschaften erfahren. Gleichwohl sind die Ergebnisse mit einer gewissen Unsicherheit behaftet, welche Vorstellung die Schülerinnen und Schüler für ihre Einschätzungen zugrunde gelegt haben. Aus diesem Grund sprechen wir hier von Einstellung gegenüber dem Sachunterricht.

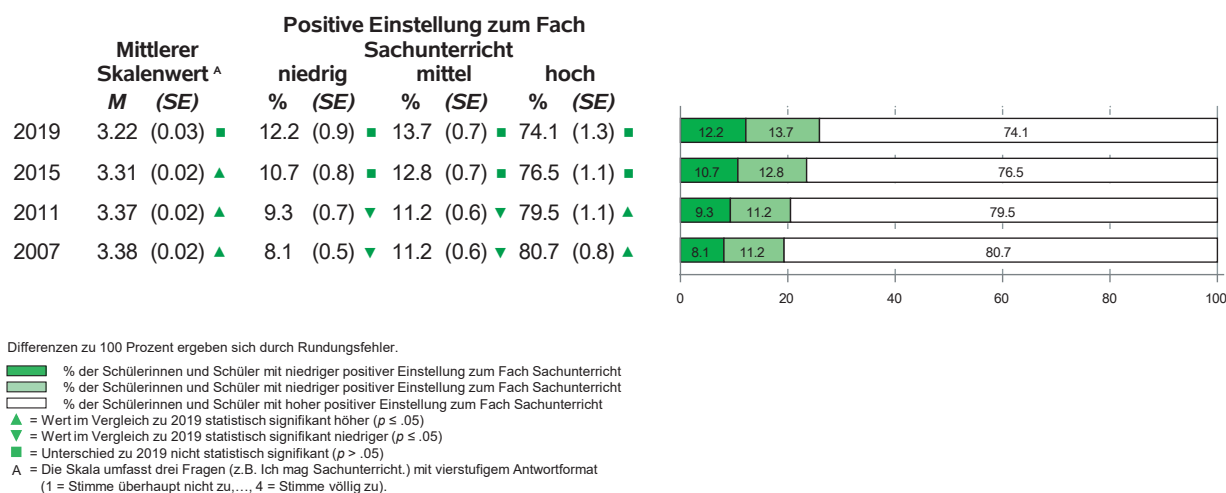
Die Skala zur Erhebung der Einstellung zum Sachunterricht umfasst drei Aussagen, bei denen die Schülerinnen und Schüler angeben sollten, ob sie ihnen völlig zustimmen, eher zustimmen, eher nicht zustimmen oder überhaupt nicht zustimmen. Folgende Aussagen wurden dabei eingeschätzt:

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

Die Antworten der Schülerinnen und Schüler wurden bei Aussage 1 und 3 mit 1 (*stimme überhaupt nicht zu*) bis 4 (*stimme völlig zu*) Punkten kodiert, bei Aussage 2 in umgekehrter Reihenfolge. Der Mittelwert über die drei Aussagen drückt aus, wie stark positiv die Einstellung des Kindes zum Sachunterricht ausgeprägt ist. Die interne Konsistenz der Skala ist hoch (*Cronbachs Alpha* = .84). Im nächsten Schritt wurden die Schülerinnen und Schüler anhand ihrer Mittelwerte in drei Gruppen mit *niedrig*, *mittel* und *hoch* ausgeprägter positiver Einstellung zum Sachunterricht eingeteilt. Schülerinnen und Schüler mit einem Mittelwert (*M*) von $1 \leq M < 2$ bildeten die Gruppe *niedrig*, mit einem Mittelwert von $2 \leq M < 3$ die Gruppe *mittel* und diejenigen mit einem Mittelwert von $3 \leq M \leq 4$ die Gruppe *hoch*.

In Abbildung 4.12 sind links die Mittelwerte und Standardfehler der Einstellung der Schülerinnen und Schüler aus Deutschland im aktuellen Studienzyklus sowie in den drei vorherigen Studienzyklen dargestellt. Rechts sind in den

Abbildung 4.12: Mittlere positive Einstellung von Schülerinnen und Schülern zum Sachunterricht sowie prozentuale Verteilungen nach niedriger, mittlerer und hoher positiver Einstellung – TIMSS 2007, 2011, 2015 und 2019 im Vergleich



horizontalen Stapelbalken die Anteile der Schülerinnen und Schüler mit *niedrig*, *mittel* und *hoch* positiven Einstellungen gegenüber dem Sachunterricht dargestellt.

Der Mittelwert der selbstberichteten Einstellung zum Sachunterricht liegt bei 3.22 Punkten auf der vierstufigen Skala und ist damit sehr hoch. Fast drei Viertel der Schülerinnen und Schüler haben eine sehr positive Einstellung zu diesem Fach. Im Vergleich zu TIMSS 2015 ist der Mittelwert um 0.09 Punkte niedriger, im Vergleich zu TIMSS 2007 um 0.16 Punkte. Beide Veränderungen sind statistisch signifikant. Es handelt sich aber allenfalls um kleine Effekte ($d = .11$ bzw. $.20$). Entsprechende Verschiebungen zeigen sich auch in den drei Gruppen zuungunsten der sehr positiven Gruppe.

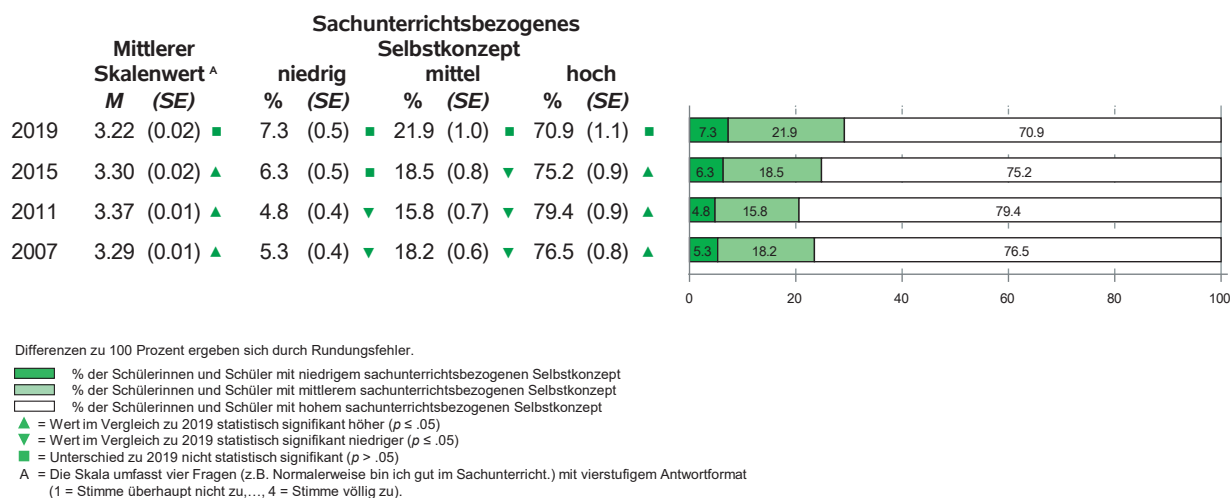
Als zweites motivationales Merkmal wurde das sachunterrichtsbezogene Selbstkonzept erhoben. Auch hier sollten die Schülerinnen und Schüler Aussagen mit den oben beschriebenen Abstufungen von *stimme überhaupt nicht zu* bis *stimme völlig zu* einschätzen. Die Skala umfasste dabei vier Aussagen:

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

Die Antworten der Schülerinnen und Schüler wurden bei Aussage 1 und 4 mit 1 (*stimme überhaupt nicht zu*) bis 4 (*stimme völlig zu*) Punkten kodiert, bei den Aussagen 2 und 3 in umgekehrter Reihenfolge. Die interne Konsistenz dieser Skala ist zwar etwas niedriger im Vergleich zu der internen Konsistenz der Skala Einstellung, liegt aber bei einem *Cronbachs Alpha* von $\alpha = .78$ und ist somit immer noch als gut einzuschätzen.

In Abbildung 4.13 sind analog zur Abbildung 4.12 die Mittelwerte und Standardfehler sowie die Verteilung auf die drei Gruppen von Schülerinnen und Schülern mit einem niedrigen, mittleren und hohen sachunterrichtsbezogenen Selbstkonzept in den vier Studienzyklen dargestellt. Ähnlich wie die positiven Einstellungen ist auch in TIMSS 2019 das sachunterrichtsbezogene Selbstkonzept mit einem Mittelwert von 3.22 Punkten sehr hoch ausgeprägt. Das heißt, die Mehrheit der Schülerinnen und Schüler (ca. 71 %) schätzt ihre eigene Leistungsfähigkeit im Sachunterricht hoch ein. Dieses positive Ergebnis spiegelt sich auch in den Anteilen in den drei Gruppen wider. Gleichzeitig zeigt sich wie auch schon in TIMSS 2015 im Vergleich zu den vorherigen Studienzyklen eine Abnahme der Schülerinnen und Schüler mit besonders hohem sachunterrichtsbezogenen Selbstkonzept zugunsten der Gruppen *mittel* und *niedrig*. Die Unterschiede zwischen TIMSS 2019 und TIMSS 2015 betragen 0.08 Skalenpunkte beziehungsweise 0.07 Skalenpunkte zu TIMSS 2007. Auch hierbei handelt es sich zwar um signifikante, aber lediglich sehr kleine Änderungen mit geringen Effekten ($d = .12$ bzw. $.11$).

In diesem Abschnitt wurde eingangs die wechselseitige Beziehung zwischen Leistung und motivationalen Merkmalen beschrieben. So sind verschiedene motivationale Merkmale wichtig für erfolgreiche Lernprozesse und umgekehrt wirkt sich eine günstige Leistungsentwicklung positiv auf motivationale Merkmale aus (Marsh, Trautwein, Lüdtke, Köller & Baumert, 2005). Entsprechend sind motivationale Merkmale wie Einstellungen, Interesse, Lernfreude sowie fachbezogene Selbstkonzepte und Selbstwirksamkeitserwartungen wichtige Aspekte von Unterricht. Vor diesem Hintergrund soll hier analysiert werden, inwieweit die

Abbildung 4.13: Mittleres sachunterrichtsbezogenes Selbstkonzept von Schülerinnen und Schülern sowie prozentuale Verteilungen nach hohem, mittlerem und niedrigem Selbstkonzept – TIMSS 2007, 2011, 2015 und 2019 im Vergleich

IEA: Trends in International Mathematics and Science Study

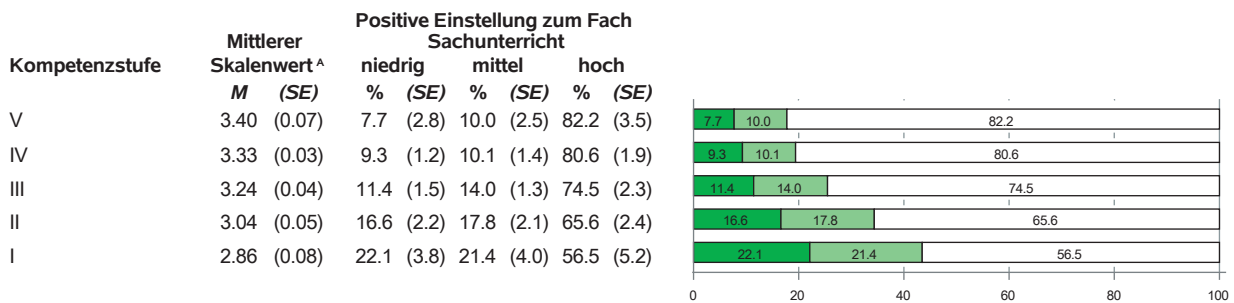
© TIMSS 2019

Einstellung und das sachunterrichtsbezogene Selbstkonzept von Schülerinnen und Schülern davon abhängig sind, welche Kompetenzstufe die Schülerinnen und Schüler erreichen. Hierzu sind in den Abbildungen 4.14 und 4.15 die Mittelwerte und die beschriebene Einteilung in die drei Gruppen *niedrig*, *mittel* und *hoch* der Einstellung beziehungsweise des fachbezogenen Selbstkonzepts für die Schülerinnen und Schüler der Kompetenzstufen I (rudimentär) bis V (fortgeschritten) dargestellt.

Aus Abbildung 4.14 lässt sich entnehmen, dass die positive Einstellung von leistungsstärkeren Schülerinnen und Schülern stärker ausgeprägt ist als von den leistungsschwächeren Schülerinnen und Schülern auf den unteren Kompetenzstufen. So nimmt der Anteil der Schülerinnen und Schüler der Gruppe *hoch* von den höchsten beiden Kompetenzstufen zu den untersten beiden Kompetenzstufen um circa 15 beziehungsweise 25 Prozent ab. Gleichwohl zeigt sich, dass auch die leistungsschwächeren Schülerinnen und Schüler immer noch relativ positive Einstellungen zum Fach Sachunterricht aufweisen.

Noch deutlicher ausgeprägt ist der Befund, wenn man das sachunterrichtsbezogene Selbstkonzept der Schülerinnen und Schüler der verschiedenen Kompetenzstufen vergleicht (Abbildung 4.15). So nimmt hier der Mittelwert zwischen der Kompetenzstufe V (3.60) und der Kompetenzstufe I (2.67) um circa eine Stufe auf der vierstufigen Skala ab. Gleichwohl ist der Mittelwert auch auf der Kompetenzstufe I immer noch über dem theoretischen Mittelwert der Skala.

Zusammenfassend zeigt sich, dass die hier erhobenen motivationalen Merkmale der Schülerinnen und Schüler hoch ausgeprägt sind und die Mehrheit, also circa drei Viertel der Schülerinnen und Schüler, günstige Voraussetzungen für das Lernen und die Weiterentwicklung von Interesse, Motivation und das Zutrauen in die eigenen Fähigkeiten im Kontext der Naturwissenschaften mitbringen. Gleichzeitig muss die (zumindest leicht) wachsende Gruppe der Schülerinnen und Schüler im Blick behalten werden, die dem Sachunterricht und damit auch den Naturwissenschaften und auch dem eigenen Lernen dort weniger positiv gegenüber eingestellt ist. Dies betrifft insbesondere leistungsschwächere Schülerinnen und Schüler. Die Befunde verweisen darauf, dass es wichtig ist, diese Gruppe

Abbildung 4.14: Prozentuale Verteilung von Schülerinnen und Schülern auf den fünf Kompetenzstufen nach Einstellungen zum Fach Sachunterricht

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

% der Schülerinnen und Schüler mit niedriger positiver Einstellung zum Sachunterricht

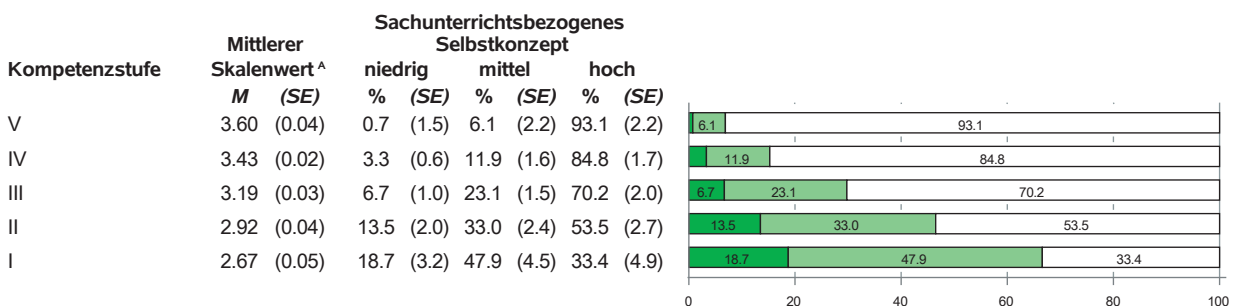
% der Schülerinnen und Schüler mit mittlerer positiver Einstellung zum Sachunterricht

% der Schülerinnen und Schüler mit hoher positiver Einstellung zum Sachunterricht

A = Die Skala umfasst drei Fragen (z.B. Ich mag Sachunterricht.) mit vierstufigem Antwortformat (1 = Stimme überhaupt nicht zu, ..., 4 = Stimme völlig zu).

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Abbildung 4.15: Prozentuale Verteilung von Schülerinnen und Schülern auf den fünf Kompetenzstufen nach sachunterrichtsbezogenem Selbstkonzept

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

% der Schülerinnen und Schüler mit niedrigem sachunterrichtsbezogenen Selbstkonzept

% der Schülerinnen und Schüler mit mittlerem sachunterrichtsbezogenen Selbstkonzept

% der Schülerinnen und Schüler mit hohem sachunterrichtsbezogenen Selbstkonzept

A = Die Skala umfasst vier Fragen (z.B. Normalerweise bin ich gut im Sachunterricht.) mit vierstufigem Antwortformat (1 = Stimme überhaupt nicht zu, ..., 4 = Stimme völlig zu).

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

nicht nur bei der Wissensentwicklung, sondern auch bei der Entwicklung motivationaler Merkmale gezielt zu fördern. Im Hinblick auf das sachunterrichtsspezifische Selbstkonzept muss allerdings beachtet werden, dass es dabei nicht um die Entwicklung von unrealistischen Fähigkeitseinschätzungen geht.

4.6 Zusammenfassung

In diesem Kapitel wurde dargestellt, über welche naturwissenschaftlichen Kompetenzstände Schülerinnen und Schüler am Ende der vierten Klasse im Jahr 2019 verfügen und welche Unterschiede sich im Vergleich zu den vorangegangenen Studienzyklen in den Jahren 2007, 2011 und 2015 zeigen. An dieser Stelle sind Teile der Ergebnisse zusammengefasst, die insbesondere aus der nationalen Perspektive relevant sind.

Niveau und Streuung der naturwissenschaftlichen Kompetenz: Die mittlere naturwissenschaftliche Kompetenz der Schülerinnen und Schüler aus Deutschland liegt bei 518 Punkten auf der TIMSS-Skala. Deutschland liegt damit über dem internationalen Mittelwert von 491 Punkten und ist vergleichbar mit den vielen anderen EU-Staaten (Mittelwert EU-Staaten: 522 Punkte). Staaten, die ein ähnliches Kompetenzniveau aufweisen wie Deutschland, sind zum Beispiel die Niederlande, Serbien, Nordirland und Bulgarien. Im Vergleich zum Mittelwert der Vergleichsgruppe OECD schneidet Deutschland signifikant, aber lediglich um 8 Punkte schlechter ab. Gleichzeitig ist der Abstand zu Staaten, die am besten abschneiden, mit zum Beispiel 77 und 70 Punkten zu Singapur beziehungsweise der Republik Korea beträchtlich. Der Mittelwert für Deutschland liegt außerdem 10 Punkte unter dem Mittelwert der vorherigen Studienzyklen, der seit 2007 stabil bei 528 Punkten lag. Ähnliche Abnahmen zeigen sich auch in manchen anderen EU-Staaten wie Österreich, Polen und Ungarn, während zum Beispiel in Schweden, Dänemark, den Niederlanden und England allenfalls kleinere Unterschiede in der mittleren Kompetenz zu beobachten sind. Eine weitere Veränderung in Deutschland, die sich in diesem Studienzyklus zeigt, ist die leichte Zunahme der Streuung der Leistungen, so zeigen die mittleren 90 Prozent der Schülerinnen und Schüler eine Streuung von 251 Punkten, was circa 20 Punkte mehr ist im Vergleich zu den beiden vorherigen Studienzyklen. Die Streuung ergibt sich vor allem durch einen niedrigeren Kompetenzwert am unteren Ende des Leistungsspektrums. So zeigen hier Schülerinnen und Schüler einen deutlich niedrigeren Kompetenzwert als bislang (5-Prozent-Perzentil: 338 Punkte). Zusammengefasst hat also nicht nur das Niveau, sondern auch die Homogenität der Leistungen in Deutschland abgenommen.

Kompetenzstufen: Für TIMSS 2019 zeigt die Verteilung auf die Kompetenzstufen, dass fast drei Viertel der Schülerinnen und Schüler in Deutschland die mittlere Kompetenzstufe III erreichen oder übertreffen. 37 Prozent der Schülerinnen und Schüler erreichen die beiden höchsten Kompetenzstufen, diese Schülerinnen und Schüler haben bereits anwendbares inhaltspezifisches Wissen von Inhalten aus verschiedenen Bereichen und zum Teil auch ein erstes Verständnis von naturwissenschaftlichen Denk- und Arbeitsweisen. Diese Gruppe ist (sehr) gut vorbereitet auf den naturwissenschaftlichen und (natur-)geografischen Unterricht der weiterführenden Schulen. Auf den unteren Kompetenzstufen I und II befinden sich 7 beziehungsweise 20 Prozent der Schülerinnen und Schüler. Diese verfügen kaum über die notwendigen Grundlagen für den Unterricht der weiterführenden Schulen. Besonders kritisch ist dabei die Gruppe der Schülerinnen und Schüler auf der Kompetenzstufe I, die lediglich über rudimentäres Alltagswissen verfügt. Diese Werte sind am oberen wie am unteren Ende des Leistungsspektrums vergleichbar mit den Mittelwerten der EU- und OECD-Staaten.

Die Trendanalysen zeigen, dass es kleinere Verschiebungen bei den Verteilungen auf die Kompetenzstufen zwischen den Jahren 2011 und 2019 gegeben hat. Während die Besetzung der oberen Kompetenzstufen relativ stabil geblieben ist, ist der Anteil der Schülerinnen und Schüler, die lediglich die unteren Kompetenzstufen erreichen, in TIMSS 2019 leicht gestiegen. Diese teils signifikanten Veränderungen betragen nur wenige Prozentpunkte und können als klein eingeschätzt werden. Sie bestätigen aber das zuvor beschriebene ungünstige Befundmuster der Zunahme von Schülerinnen und Schülern am unteren Ende des Leistungsspektrums.

Inhalts- und Anforderungsbereiche: Die Rahmenkonzeption von TIMSS erlaubt die Differenzierung der Ergebnisse nach den drei Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geografie* und nach den drei kognitiven Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen*. Zwischen den Inhaltsbereichen zeigen sich signifikante Unterschiede in den relativen Leistungen zugunsten der *Biologie* und zuungunsten der *Geografie*. Der Unterschied im Inhaltsbereich *Biologie* ist sehr klein, während der im Bereich *Geografie* etwas größer ist. Möglicherweise setzen Lehrkräfte in der geografischen Perspektive des Sachunterrichts ihre Schwerpunkte eher auf regionalgeografische Themen, während die Themen, die im TIMSS-Inhaltsbereich *Geografie* erfragt werden, weniger im Unterricht behandelt werden. Das schwächere Abschneiden in diesem Bereich ist auch aus den vorherigen Studienzyklen bekannt. Themen wie das Sonnensystem, Erdrotation und -revolution werden in den meisten Curricula in Deutschland im Geografieunterricht der Jahrgänge 5 und 6 (nochmal) aufgegriffen und vertieft. Tendenziell zeigt sich die relative Schwäche auch in vielen anderen Staaten.

Der Vergleich mit den vorherigen Studienzyklen, an denen Deutschland teilgenommen hat, macht deutlich, dass das insgesamt schwächere Abschneiden der Schülerinnen und Schüler aus Deutschland in TIMSS 2019 nicht an einem bestimmten Inhaltsbereich liegt, sondern sich die Abnahme der Kompetenz über alle drei Inhaltsbereiche verteilt. Ähnliches zeigt sich auch im Hinblick auf die kognitiven Anforderungsbereiche. Hier zeigen sich keine relativen Stärken oder Schwächen zwischen dem *Reproduzieren*, *Anwenden* und *Problemlösen*. Der Vergleich mit TIMSS 2015 zeigt eine signifikante Abnahme der mittleren Kompetenzen in allen drei Anforderungsbereichen. Das bedeutet, dass zum Beispiel keine spezifischen Schwächen in den komplexeren Anforderungen wie *Anwenden* und *Problemlösen* erkennbar sind. Im naturwissenschaftsbezogenen Sachunterricht sollten weiterhin die drei Inhalts- und Anforderungsbereiche gleichmäßig gefördert werden.

Sachunterrichtsbezogene Einstellung und Selbstkonzept: Neben kognitiven Aspekten sind für gelingende naturwissenschaftliche Bildungsprozesse auch motivationale Merkmale entscheidend, da sie den Erfolg von Lernprozessen beeinflussen und auch für spätere Schwerpunktsetzungen in der Schule sowie Ausbildung und Studium wesentlich sind. In TIMSS werden daher auch die Einstellung zum Sachunterricht und das sachunterrichtsbezogene Selbstkonzept erhoben. Den Schülerinnen und Schülern wurde dabei vor der Befragung der Hinweis auf die naturwissenschaftlichen Anteile des Sachunterrichts durch konkrete Beispiele von Themen gegeben. Gleichwohl sind die Ergebnisse mit Vorsicht zu interpretieren, da sicher nicht alle Schülerinnen und Schüler zwischen den naturwissenschaftsbezogenen und anderen Anteilen des Sachunterrichts differenzieren können. Es besteht also eine gewisse Ungewissheit, inwiefern sich die Angaben der Kinder nur auf die Naturwissenschaften oder eher auf den Sachunterricht generell beziehen. Jeweils knapp drei Viertel der Schülerinnen und Schüler weisen eine sehr positive Einstellung zum Sachunterricht beziehungsweise ein positives sachunterrichtsbezogenes Selbstkonzept auf. Im Vergleich zu den vorherigen Studienzyklen sind die Werte leicht niedriger. Dieser Abfall ist zwar signifikant, aber klein. Allerdings ist auffällig, dass für die positive Einstellung seit 2007 und für das sachunterrichtsbezogene Selbstkonzept seit 2011 eine kontinuierliche Abnahme zu verzeichnen ist. Die Differenzierung der Mittelwerte der Einstellung und des sachunterrichtsbezogenen Selbstkonzepts aus TIMSS 2019 nach Schülerinnen

und Schülern, die unterschiedliche Kompetenzstufen erreichen, verdeutlicht, dass insbesondere bei Schülerinnen und Schülern am unteren Leistungsspektrum beide Merkmale etwas niedriger ausgeprägt sind.

Der Rückgang der im Durchschnitt erreichten Leistungen in den Naturwissenschaften und den naturwissenschaftsbezogenen motivationalen Merkmalen sowie die Zunahme der Heterogenität in den Leistungen in TIMSS 2019 im Vergleich zu den vorherigen Studienzyklen sind zwar signifikant, aber dennoch als eher klein einzuschätzen. Gleichzeitig muss man sagen, dass sie ein ungünstiges Befundmuster darstellen. Erstrebenswert für Bildungssysteme sind ein möglichst hohes Niveau, Homogenität der Kompetenzwerte und relativ hohe Kompetenzwerte auch am unteren Leistungsspektrum. Die Ergebnisse entsprechen in der Tendenz den leicht negativen Befunden aus der Mathematik in der Primarstufe aus TIMSS 2015 (Selter, Walter, Walther & Wendt, 2016) sowie dem IQB-Bildungstrend 2016 (Stanat, Schipolowski, Rjosk, Weirich & Haag, 2017). Auch hier zeigten sich Rückgänge in den mittleren Leistungen, auch wenn die Höhe dieses negativen Trends unterschiedlich ist. Die aktuellen TIMSS-Befunde aus der Mathematik zeigen keine weitere Abnahme, aber auch keine Verbesserung der mittleren mathematischen Leistungen (siehe Kapitel 3 in diesem Band). Ursächlich für diese Befunde können Veränderungen im Schulsystem sein (Autorengruppe Bildungsberichterstattung, 2020), wie die Ausweitung der gemeinsamen Beschulung von Kindern mit und ohne sonderpädagogischen Förderbedarf oder die Zunahme an Schülerinnen und Schülern mit Migrationshintergrund, die häufiger als Schülerinnen und Schüler ohne Migrationshintergrund geringere sprachliche Kompetenzen aufweisen (Rjosk, Haag, Heppt & Stanat, 2017; siehe Kapitel 12 in diesem Band). Diese Veränderungen können zu Veränderungen im Niveau und der Kompetenzstreuung führen. Um zu verhindern, dass die aktuellen Ergebnisse in den Naturwissenschaften sich tatsächlich zu einem negativen Trend entwickeln, erscheint es besonders zentral, Schülerinnen und Schüler mit ungünstigen Voraussetzungen gezielt zu fördern – ohne dabei die leistungsstarken Schülerinnen und Schüler zu vernachlässigen. Naturwissenschaftsspezifische Maßnahmen, in denen geeignete Bildungsangebote mit damit einhergehenden Fortbildungen für die Lehrkräfte umgesetzt werden, scheinen hierfür geeignete Ansatzpunkte. Aufgrund der besonderen Bedeutung von sprachlichen Kompetenzen für das Lernen sind aber zudem auch Maßnahmen der Sprachförderung und der Sprachförderung speziell im Sachunterricht geeignet, um einem möglichen Negativtrend entgegenzuwirken.

Literatur

- Autorengruppe Bildungsberichterstattung. (Hrsg.). (2020). *Bildung in Deutschland 2020. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt*. Bielefeld: wbv. Verfügbar unter: https://www.bildungsbericht.de/static_pdfs/bildungsbericht-2020.pdf
- Bayerisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst. (Hrsg.). (2014). *LehrplanPLUS Grundschule. Lehrplan für die bayerische Grundschule*. Verfügbar unter: <https://www.lehrplanplus.bayern.de/schulart/grundschule>
- Blaseio, B. (2009). Natur in den Bildungsplänen des Elementarbereichs. In R. Lauterbach, H. Giest & B. Marquardt-Mau (Hrsg.), *Lernen und kindliche Entwicklung. Elementarbildung und Sachunterricht* (Probleme und Perspektiven des Sachunterrichts, Bd. 19, S. 85–92). Bad Heilbrunn: Klinkhardt.

- Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.). (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Wendt, H., Köller, O. & Selter, C. (Hrsg.). (2012). *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bybee, R. & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, 33(1), 7–26.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9. <https://doi.org/10.1016/j.learninstruc.2013.07.001>
- Gesellschaft für die Didaktik des Sachunterrichts. (2013). *Perspektivrahmen Sachunterricht* (vollständig überarbeitete und erweiterte Aufl.). Bad Heilbrunn: Klinkhardt.
- Harlen, W. (2015). *Working with big ideas*. Trieste: Science Education Programme (SEP) of IAP.
- Kleickmann, T., Brehl, T., Saß, S., Prenzel, M. & Köller, O. (2012). Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In W. Bos, H. Wendt, O. Köller & C. Selter (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 123–169). Münster: Waxmann.
- Kleickmann, T., Steffensky, M. & Praetorius, A.-K. (2020). Quality of teaching in elementary science education: More than three basic dimensions? In A.-K. Praetorius & E. Klieme (Hrsg.), *Empirische Forschung zu Unterrichtsqualität. Theoretische Grundfragen und quantitative Modellierungen* (Zeitschrift für Pädagogik, 66. Beiheft, S. 37–55). Weinheim: Beltz Juventa.
- Kriegbaum, K., Becker, N. & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review*, 25, 120–148. <https://doi.org/10.1016/j.edurev.2018.10.001>
- KMK (Kultusministerkonferenz). (Hrsg.). (2004). *Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. Verfügbar unter: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Chemie.pdf
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, 76(2), 397–416. <https://doi.org/10.1111/j.1467-8624.2005.00853.x>
- Ministerium für Kultus, Jugend und Sport Baden-Württemberg. (Hrsg.). (2016). *Bildungsplan der Grundschule Sachunterricht*. Verfügbar unter: <http://www.bildungsplaene-bw.de/site/bildungsplan/get/documents/lsbw/export-www.bildungsplaene-bw.de>
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (Hrsg.). (2008). *Richtlinien für die Grundschule in Nordrhein-Westfalen*. Düsseldorf.
- Möller, J. & Trautwein, U. (2009). Selbstkonzept. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 179–203). Heidelberg: Springer. https://doi.org/10.1007/978-3-540-88573-3_8
- Mullis, I. V. S. & Martin, M. O. (Hrsg.). (2017). *TIMSS 2019 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Verfügbar unter: <http://timssandpirls.bc.edu/timss2019/international-results/>
- Mullis, I. V. S., Martin, M. O., Goh, S. & Cotter, K. (Hrsg.). (2016). *TIMSS 2015 encyclopedia: Education policy and curriculum in mathematics and science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Verfügbar unter: <http://timssandpirls.bc.edu/timss2015/encyclopedia/>
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Committee on a conceptual framework for new K-12 science education standards. Board on science education, division of behavioral and social sciences and education*. Washington, DC: The National Academies Press.
- Niedersächsisches Kultusministerium. (Hrsg.). (2017). *Kerncurriculum für die Grundschule Schuljahrgänge 1–4 Sachunterricht*. Verfügbar unter: <https://cuvo.nibis.de/cuvo>

- php?key_lev0_0=Dokumentenart&svalue_lev0_0=Kerncurriculum&key_lev0_1=Fach&svalue_lev0_1=Russisch&fulltextsearch_lev0_ov=&key_lev0_0_ov=Dokumentenart&svalue_lev0_0_ov=&key_lev0_1001_ov=Schulbereich&svalue_lev0_1001_ov=&key_lev0_1002_ov=Schulform&svalue_lev0_1002_ov=&key_lev0_1_ov=Fach&svalue_lev0_1_ov=Sachunterricht&p=search
- OECD. (2016). *PISA 2015 Assessment and analytical framework: Science, reading, mathematical and financial literacy* (PISA). Paris: OECD publishing.
- OECD Global Science Forum. (Hrsg.). (2006). *Evolution of student interest in science and technology studies. Policy report*. Verfügbar unter: <http://www.oecd.org/data-oecd/16/30/36645825.pdf>
- Osborne, J., Simon, S. & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), 1049–1079.
- Rjosk, C., Haag, N., Heppt, B. & Stanat, P. (2017). Zuwanderungsbezogene Disparitäten. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich & N. Haag (Hrsg.), *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* (S. 237–276). Münster: Waxmann.
- Sächsisches Staatsministerium für Kultus. (Hrsg.). (2004/2009). *Lehrplan Grundschule Sachunterricht*. Verfügbar unter: <http://www.widerstreit-sachunterricht.de/lehrplaene/Sachsen/2004-2009%20LP.pdf>
- Seaton, M., Marsh, H. W. & Craven, R. G. (2009). Big-fish-little-pond effect. Generalizability and moderation. Two sides of the same coin. *American Educational Research Journal*, 47(2), 1–44. <https://doi.org/10.3102/0002831209350493>
- Selter, C., Walter, D., Walther, G. & Wendt, H. (2016). Mathematische Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 75–132). Münster: Waxmann.
- Smith, C. L., Wiser, M., Anderson, C. W. & Krajcik, J. (2006). FOCUS ARTICLE: Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic-molecular theory. *Measurement: Interdisciplinary Research and Perspective*, 4(1/2), 1–98. <https://doi.org/10.1080/15366367.2006.9678570>
- Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S. & Haag, N. (Hrsg.). (2017). *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Münster: Waxmann.
- Steffensky, M., Kleickmann, T., Kasper, D. & Köller, O. (2016). Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015 Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 133–184). Münster: Waxmann.
- Stubbe, T., Schwippert, K. & Wendt, H. (2016). Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (1. Auflage, neue Ausgabe, S. 297–314). Münster: Waxmann.
- Vande Gaer, E., Grisay, A., Schulz, W. & Gebhardt, E. (2012). The reference group effect. *Journal of Cross-Cultural Psychology*, 43(8), 1205–1228. <https://doi.org/10.1177/0022022111428083>
- Wendt, H., Bos, W., Kasper, D., Walzebug, A., Goy, M. & Jusufi, D. (2016). Ziele, Anlage und Durchführung der Trends in International Mathematics and Science Study (TIMSS 2015). In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 31–77). Münster: Waxmann
- Wendt, H., Bos, W., Selter, C., Köller, O., Schwippert, K. & Kasper, D. (Hrsg.). (2016). *TIMSS 2015 Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Wendt, H., Schwippert, K. & Stubbe, T. (2016). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In H.

- Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015 Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 315–329). Münster: Waxmann.
- Wendt, H., Steinmayr, R. & Kasper, D. (2016). Geschlechtsunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015 Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 255–296). Münster: Waxmann.
- Wigfield, A. & Eccles, J. S. (2000). Expectancy – Value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81.
- Wittwer, J., Saß, S. & Prenzel, M. (2008). Naturwissenschaftliche Kompetenz im internationalen Vergleich. Testkonzeption und Ergebnisse. In W. Bos, M. Bensen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 87–124). Münster: Waxmann.

Kapitel 5

Leistungsveränderungen in TIMSS zwischen 2015 und 2019: Die Rolle des Testmediums und des methodischen Vorgehens bei der Trendschätzung

Alexander Robitzsch, Oliver Lüdtke, Knut Schwippert, Frank Goldhammer, Ulf Kroehne und Olaf Köller

5.1 Einleitung

Die *Trends in International Mathematics and Science Study* (TIMSS) 2019 hat wie auch das *Programme for International Student Assessment* (PISA) den Erhebungsmodus gewechselt (Reiss, Weis, Klieme & Köller, 2019): An die Stelle von Papier-Testheften, die mit Stiften durch die Schülerinnen und Schüler bearbeitet werden (*Paper-based-Assessment* – PBA), ist die Testung mithilfe von Computern getreten (*Computer-based-Assessment* – CBA). Der Wechsel des Testmediums ist für Trendschätzungen in internationalen Schulleistungsstudien voraussetzungsreich, zeigt doch eine Vielzahl von Studien (Kroehne, Buerger, Hahnel & Goldhammer, 2019; Robitzsch, Lüdtke, Goldhammer, Kroehne & Köller, 2020), dass sich Testitems in ihren Eigenschaften (Schwierigkeit und Diskrimination) verändern können, wenn das Testmedium gewechselt wird. So konnten Robitzsch et al. (2017) für die PISA-2015-Erhebung in Deutschland zeigen, dass Items, die am Computer gelöst wurden, im Mittel schwieriger waren als bei der Bearbeitung im analogen Testheft. Goldhammer et al. (2019) haben diese Effekte auch für PISA 2018 belegen können. Als Folge zeigte sich in Sekundäranalysen der PISA-Datensätze, dass bei einer nationalen Schätzung und Berücksichtigung dieses Moduseffekts (die Veränderung der Itemeigenschaften bei Übertragung von Papier auf den Computer) Schülerinnen und Schüler in Deutschland höhere Kompetenzstände aufwiesen, als von der OECD offiziell publiziert.

In TIMSS 2019 hat die internationale Studienleitung ein Testdesign umgesetzt, das es für die internationalen Datensätze, aber auch für den nationalen deutschen Datensatz erlaubt, Moduseffekte zu schätzen und damit auch zu untersuchen, in welchem Ausmaß für die Trendschätzungen in Mathematik und den Naturwissenschaften der Effekt des Testmediums berücksichtigt werden sollte. In diesem Kapitel wird der Frage nach potenziellen Moduseffekten mit vertiefen-

den Analysen nachgegangen. Jenseits der Aufdeckung dieser Effekte wird auch untersucht, ob die Art der gewählten Trendschätzung (originaler vs. marginaler Trend; Carstensen, Prenzel & Baumert, 2009; Gebhardt & Adams, 2007) Effekte auf die Interpretation der Veränderungen der Leistungen in Deutschland hat. Originale Trendschätzungen basieren auf den Itemparametern, die in der internationalen TIMSS-Gesamtstichprobe ermittelt wurden, marginale Trendschätzungen beschränken sich auf Itemparameter, die in der nationalen (deutschen) Stichprobe geschätzt wurden.

5.2 Trendschätzungen in internationalen Schulleistungstudien

Vorhaben wie TIMSS stellen den Teilnehmerstaaten wichtige Informationen über die Leistungsfähigkeit ihrer Bildungssysteme bereit. Die in festen Zeitabständen wiederholten Erhebungen liefern Daten über die Entwicklung des Bildungssystems beziehungsweise bieten Hinweise, ob nachfolgende Kohorten von Schülerinnen und Schülern (der 4. bzw. 8. Jahrgangsstufe) im Vergleich zu früheren Kohorten höhere Kompetenzstände aufweisen. Die Interpretation von Trendinformationen als Beleg für die Leistungsfähigkeit von Bildungssystemen ist allerdings voraussetzungsreich (Mazzeo & von Davier, 2008). Letztendlich können solche Trendinformationen nur valide gewonnen werden, wenn sichergestellt wird, dass die Erhebungsinstrumente, die in unterschiedlichen Zyklen und/oder Kohorten eingesetzt werden, zu vergleichbaren Messungen desselben Merkmals (Konstrukt) aufseiten der Schülerinnen und Schüler führen. Im Kontext der PISA-Studie haben Mazzeo und von Davier (2008) folgende Empfehlungen abgegeben, um eine valide Schätzung von Trends zu erreichen:

- *Möglichst große Zahlen von Trenditems*: Auswahl solcher Items, die von Zyklus zu Zyklus immer wieder eingesetzt werden können.
- *Beachtung von Testleteffekten*: In internationalen Studien wird oftmals ein relativ langer Aufgabenstamm mit relativ vielen Items präsentiert; dies birgt die Gefahr von Itemabhängigkeiten, dementsprechend sollte ein Testlet eher wenige Items enthalten.
- *Beachtung – besser Kontrolle – von Kontexteffekten*: Items können ihre Eigenschaften in Abhängigkeit der Items, die vorher administriert wurden, ändern. Insofern ist es sinnvoll, hier von Zyklus zu Zyklus die Kontexte möglichst stabil zu halten.
- *Beachtung und Kontrolle von Positionseffekten*: Itemparameter schwanken in Abhängigkeit von ihrer Position in einem Testheft. Von Zyklus zu Zyklus sollten also die Positionen von Testlets in Heften konstant gehalten werden.
- *Hinreichend viele Short-Answer und Multiple-Choice-Items*: Dies sichert im Gegensatz zu *Open-(Constructed)-Response-Items* eine hohe Objektivität bei der Kodierung in richtig/falsch.

Im einfachsten Fall wechselt man nicht das Instrument und verwendet es in nachfolgenden Erhebungszyklen immer wieder („When measuring change, do not change the measure“; Beaton, 1990, S. 10). Diese Strategie wurde in der Vergangenheit in TIMSS verfolgt, indem eine Teilmenge von Testheften mit relativ großen Zahlen entsprechender Trenditems unverändert wieder eingesetzt wurde (Martin, Mullis & Hooper, 2016). Die übrigen Items wurden publiziert. Testlets enthalten in TIMSS üblicherweise wenige Items; *Short-Answer*- und

Multiple-Choice-Items sind in relativ großer Zahl vorhanden. Im internationalen Vorgehen in TIMSS werden das 2PL- (offene Antwortformate) und das 3PL-Modell (*Multiple-Choice-Items*) bei der Skalierung eingesetzt (Martin et al., 2016). Letztendlich wurde durch dieses Vorgehen zwischen TIMSS 2007 und TIMSS 2015 sichergestellt, dass robuste Trendschätzungen der mathematischen und naturwissenschaftlichen Leistungen von Kindern in vierten Klassen in Deutschland publiziert werden konnten.

5.2.1 Marginale und originale Trendschätzung

In internationalen Schulleistungsstudien werden die Itemparameter üblicherweise anhand der internationalen Stichprobe, die Daten aus allen Teilnehmerstaaten enthält, geschätzt. Für Trendanalysen wählt man eine Referenzerhebung (für TIMSS 2019 ist die Referenzerhebung TIMSS 2015), führt eine konkurrente Skalierung (gemeinsame Skalierung internationaler PBA-Datensätze von 2015 und 2019) durch und betrachtet die Veränderung der Verteilung der Kompetenzen (Fishbein, Martin, Mullis & Foy, 2018). So ergibt sich dann auf der Basis internationaler Itemparameter die Leistungsveränderung in jedem Teilnehmerstaat. Dabei wird in der konkurrenten Skalierung die Annahme getroffen, dass die Itemparameter invariant über die verschiedenen Staaten sind, das heißt, es gibt keinen Interaktionseffekt ‚Item \times Country‘. Gleichwohl hat sich in der Vergangenheit gezeigt, dass für ausgewählte Items zumindest kleine Interaktionseffekte gefunden werden konnten (Klieme & Baumert, 2001). Diese Interaktionen haben Konsequenzen für die Trendschätzung in zwei aufeinanderfolgenden Zyklen, wenn sich Trenditems (solche, die im ersten und zweiten Zyklus eingesetzt werden) von Nicht-Trenditems (solche, die nur in einem Zyklus eingesetzt werden) im mittleren Interaktionseffekt ‚Item \times Country‘ unterscheiden. Gilt beispielsweise, dass die Trenditems in einem Staat spezifisch schwieriger sind – im Gegensatz zu den Nicht-Trenditems im ersten Zyklus – so wird beim Ignorieren dieses Effekts die internationale (originale) Trendschätzung typischerweise dazu führen, dass für diesen Staat die Trendschätzung negativ verzerrt ist. Das heißt, eine positive Veränderung wird in diesem Staat unter-, eine negative Veränderung überschätzt.

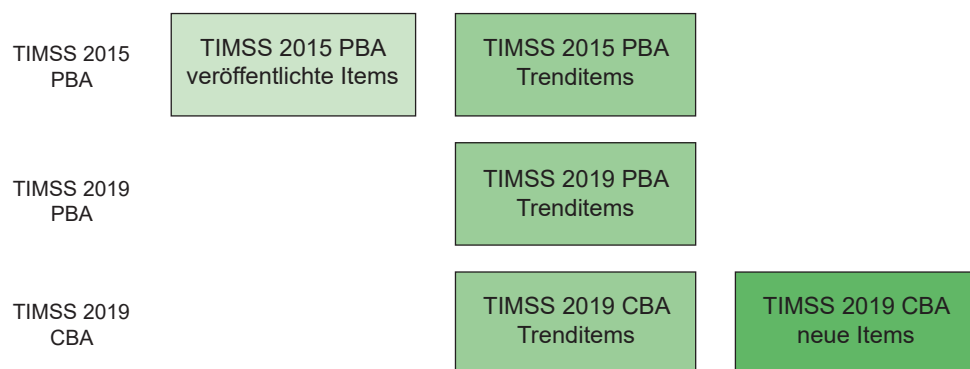
Um solche Phänomene besser untersuchen zu können, werden häufig sogenannte marginale Trendschätzungen ergänzend vorgenommen, bei denen nur die Daten eines einzelnen Staates berücksichtigt werden. Die nationalen Itemparameter werden getrennt für die Zyklen geschätzt und anschließend eine *Linking*-Methode eingesetzt, sodass die Verteilungen der Kompetenzen unterschiedlicher Zyklen auf derselben Metrik abgetragen werden können. Als Folge basiert die marginale Trendschätzung in einem Staat – im Gegensatz zur originalen Trendschätzung – nur auf der Information der Trenditems. Die übrigen Items (Nicht-Trenditems), die nur einmalig eingesetzt wurden, beeinflussen diese Schätzung nicht. In der Literatur zu originalen und marginalen Trends ist dementsprechend wiederholt dokumentiert, dass diese deutlich unterschiedlich ausfallen können (Gebhardt & Adams, 2007; Robitzsch & Lüdtke, 2019; Robitzsch, Lüdtke et al., 2020). So konnte schon im Rahmen von PISA 2006 gezeigt werden, dass die Interpretation der Veränderung der naturwissenschaftlichen Kompetenzen in Deutschland positiver ausfällt, wenn anstelle des originalen der marginale Trend geschätzt wurde (Carstensen et al., 2009). Robitzsch und Lüdtke (2019)

zeigen zudem, dass marginale Trendschätzungen häufig präziser als originale Trendschätzungen sind (vgl. auch Sachse, Roppelt & Haag, 2016).

5.2.2 Trendschätzung in TIMSS 2019 – Designvorgaben

Mit TIMSS 2019 wurden die mathematischen und naturwissenschaftlichen Leistungen in TIMSS erstmalig per CBA erfasst. Letztendlich trägt dieser Paradigmenwechsel dem Umstand Rechnung, dass auch im schulischen Kontext ein Transformationsprozess von analogen zu digitalen Lehr-Lernmedien stattfindet und digitale Test- und Prüfungsformate zukünftig in Schulen bzw. Klassen gegenüber analogen dominieren dürften. Um den damit verbundenen Problemen bei der Trendschätzung (Robitzsch, Lüdtke et al., 2020) zu begegnen, hat sich die internationale TIMSS-Studienleitung entschieden, einmalig eine sogenannte Brückenstudie durchzuführen, bei der in ausgewählten Teilnehmerstaaten eine große nationale Stichprobe die CBA-Instrumente und eine kleinere, auch national repräsentative Stichprobe, die PBA-Instrumente erhalten hat. Wie in früheren Zyklen enthielten sowohl die PBA- als auch die CBA-Instrumente eine große Anzahl von Items, die bereits in TIMSS 2015 eingesetzt worden waren. Die Abbildung 5.1 zeigt das realisierte Design und seine Verzahnung mit TIMSS 2015.

Abbildung 5.1: Testdesign für deutsche Stichproben in TIMSS 2015 und TIMSS 2019



IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Das 2019 realisierte internationale PBA-Design enthält gemeinsame Items beziehungsweise Itemcluster mit den 2015 realisierten PBA-Tests. Durch dieses Design konnten die internationalen PBA-Datensätze von 2015 und 2019 gemeinsam skaliert werden, um so zu gemeinsamen Itemparametern und einer Trendschätzung zu gelangen. Die Skalierung unter Berücksichtigung der Trendsitems im CBA-Modus bei gleichzeitiger Fixierung der PBA-Trendsitemparameter aus der vorherigen konkurrenten Skalierung erlaubt zum einen die Schätzung von Moduseffekten, zum anderen den Ausschluss der CBA-Items, deren Itemparameter von den PBA-Parametern bei der Trendschätzung abweichen. Danach würden für die Bestimmung eines Trends zwischen 2015 und 2019 auf Basis der CBA-Items nur diejenigen CBA-Items in die Trendschätzung eingehen, die identische Itemparameter (d.h. invariant über den Testmodus unter Verwendung einer additiven Konstanten, der für mittlere Moduseffekte kon-

trolliert; Kroehne, Buerger et al., 2019) wie die PBA-Items haben. Dieses Vorgehen nutzt für die Trendschätzung in jedem Staat die im internationalen Datensatz gewonnenen Itemparameter, das heißt, sie basiert auf den Daten aller Teilnehmerstaaten, die sowohl PBA- als auch CBA-Testungen vorgenommen haben. Deshalb wird angenommen, dass der Interaktionseffekt ‚Item \times Modus‘ in allen Teilnehmerstaaten identisch ist, das heißt, der Wechsel von PBA auf CBA soll in allen Staaten zu identischen Veränderungen der Itemcharakteristiken führen (Fishbein et al., 2018). Die zuvor bereits erwähnten Sekundäranalysen der PISA-Daten (Jerrim, Micklewright, Heine, Sälzer & McKeown, 2018; Robitzsch et al., 2017; Robitzsch, Lüdtke et al., 2020) machen allerdings deutlich, dass dies keineswegs der Fall sein muss, Moduseffekte also nach Teilnehmerstaat variieren können. Als Folge können marginale und originale Trendschätzungen unterschiedlich ausfallen, wenn nämlich in einem spezifischen Staat andere Items vom Moduseffekt betroffen sind, als dies im internationalen Datensatz der Fall ist.

5.2.3 Moduseffekte in Schulleistungsstudien

Wechselt man den Erhebungsmodus, indem man Items, die bisher im PBA-Modus administriert wurden, auf dem Computer darbietet, dann können sich die psychometrischen Eigenschaften der Messung ändern. Solche Moduseffekte können Eigenschaften einzelner Items betreffen und damit zusammenhängend auch Eigenschaften einer ganzen Skala (z.B. die Metrik, Messgenauigkeit, Interpretation des Testwertes). Aus Metaanalysen ist bekannt, dass Richtung und Stärke von Moduseffekten von unterschiedlichen Faktoren abhängen können (Bennett et al., 2008; Jerrim, 2016; Kingston, 2008; Mangen, Walgermo & Brønnick, 2013; Wang, Jiao, Young, Brooks & Olson, 2008). Moduseffekte können dabei als das Ergebnis einer (unbekannten) Mischung von Einzeleffekten geänderter Merkmale der Messung verstanden werden (Kroehne & Martens, 2011), beispielsweise wenn sich in einer Leseverständnisaufgabe infolge der Digitalisierung das Format des zu lesenden Textes ändert oder auch das Antwortformat. Moduseffekte können sowohl direkt als auch vermittelt durch den Bearbeitungsprozess auftreten. Beispielsweise kann sich in Folge des Moduswechsels die Bearbeitungsgeschwindigkeit ändern (Kroehne, Hahnel & Goldhammer, 2019), was sich in einem unterschiedlichen Ausmaß nicht erreichter Aufgaben auswirken kann (Kroehne, Deribo & Goldhammer, 2020). Das bedeutet, dass in jeder Studie eine eigene empirische Überprüfung von Moduseffekten erforderlich ist. Dies trifft auf TIMSS 2019 umso mehr zu, als dass die in Deutschland getestete Population der Viertklässlerinnen und Viertklässler typischerweise keine Computer oder eher selten Computer im Unterricht nutzt. Dementsprechend dürfte es für viele Kinder ungewohnt sein, Aufgaben im Mathematik- oder Sachunterricht am Computer zu bearbeiten. Zu vermuten ist daher, dass die Items beim CBA gegenüber dem PBA schwieriger werden könnten. Für Fünfzehnjährige ließen sich genau solche Effekte in Deutschland belegen, und zwar sowohl für PISA 2015 als auch für PISA 2018 (Goldhammer et al., 2019; Robitzsch et al., 2017). Darüber hinaus wurde beispielhaft gezeigt, dass die Moduseffekte das gemessene Konstrukt nicht wesentlich verändern (d.h., Konstruktäquivalenz vorliegt, Goldhammer et al., 2019; Kroehne, Buerger et al., 2019).

5.2.4 Moduseffekte in TIMSS 2019

Das in Abbildung 5.1 gezeigte Testdesign von TIMSS 2019 lässt die Schätzung von Moduseffekten auf Basis des internationalen Datensatzes zu, indem die Itemparameter der Trenditems zwischen PBA und CBA verglichen werden. Im Fall der Invarianz über die Testmodi hinweg wären Abweichungen in den Itemparametern zwischen PBA und CBA lediglich zufällig und ließen sich inferenzstatistisch nicht absichern. Die internationale Studienleitung hat hierzu entsprechende Ergebnisse vorgelegt, die sich in der vierten Jahrgangsstufe auf die Daten von etwa 46 000 Schülerinnen und Schülern beziehen (Foy, Yin & Fishbein, 2020). Für die Schwierigkeitsparameter der Items ergaben sich hohe Korrelationen zwischen PBA und CBA (Mathematik: $r = .94$, Naturwissenschaften: $r = .95$). Dabei wurde berichtet, dass insgesamt 72 Prozent der Mathematikitems und 86 Prozent der Naturwissenschaftsitems invariante PBA- und CBA-Schwierigkeitsparameter aufwiesen. Auf der *Logit*-Metrik der Itemparameter ergab sich, dass die nicht invarianten Items in der vierten Jahrgangsstufe auf dem Computer im Mittel schwieriger (in positiver Richtung verschoben) waren. Im Mittel waren so die CBA-Trenditems um 0.096 *Logits* (Mathematik) beziehungsweise 0.061 *Logits* (Naturwissenschaften) schwieriger. Auf der internationalen Metrik ($M = 500$, $SD = 100$) entspricht dies ungefähr 6 beziehungsweise 10 Punkten. Der Befund steht im Einklang mit einer für TIMSS 2019 durchgeführten Äquivalenzstudie (Fishbein et al., 2018), in der auch der mittlere Moduseffekt in Naturwissenschaften geringer ausfiel als in Mathematik (Hamhuis, Glas & Meelissen, 2020). Dementsprechend musste in der Trendschätzung 2015/2019 der CBA-Stichprobe eine Adjustierung vorgenommen werden, um eine negative Verzerrung der Veränderung zu kompensieren. Letztendlich führte dieses Vorgehen dazu, dass sich für den internationalen TIMSS-2019-Datensatz aller Schülerinnen und Schüler die Testleistungen der CBA- und der PBA-Stichprobe nicht aufgrund unterschiedlicher Testmodi unterscheiden. Beobachtbare Differenzen zwischen beiden Stichproben sollten sich lediglich im Zufallsbereich bewegen.

Für die Berichtslegung stellt die internationale Studienleitung staatspezifische Kompetenzverteilungen (*plausible values*; Lüdtke & Robitzsch, 2017) für Mathematik und Naturwissenschaften sowohl für die PBA- als auch die CBA-Stichprobe zur Verfügung. Unterschiede in den Verteilungen für die PBA- und CBA-Stichprobe können auf staatspezifische Moduseffekte hinweisen, die von den auf internationaler Ebene gewonnenen Moduseffekten abweichen (Foy et al., 2020).

5.2.5 Fragestellungen

Mit den nachfolgenden vertiefenden Analysen der nationalen TIMSS-Daten werden im Wesentlichen zwei Ziele verfolgt. Im ersten Schritt interessiert das Ausmaß des Moduseffekts in der nationalen Stichprobe der Viertklässlerinnen und Viertklässler. Dazu werden die Unterschiede in den Itemparametern der PBA- und CBA-Trenditems geschätzt. Analog zu früheren Befunden aus Stichproben mit älteren Schülerinnen und Schülern (Robitzsch et al., 2017; Robitzsch, Lüdtke et al., 2020) wird erwartet, dass die CBA-Items im Mittel schwieriger sind und bei keiner entsprechenden Adjustierung zu niedrigeren Leistungswerten in TIMSS 2019 führen.

Über den Moduseffekt hinaus wird dem international publizierten originalen Trend von TIMSS 2015 nach TIMSS 2019 der marginale Trend gegenübergestellt, der sich allein auf den nationalen Datensatz und die Trenditems bezieht. Die Analysen zeigen, in welchem Ausmaß und in welcher Richtung sich der marginale vom originalen Trend unterscheidet. Zusätzlich soll ein bedingter marginaler Trend berichtet werden, der die Veränderung der deutschen Testleistungen von 2015 nach 2019 abbildet, wenn die Verteilung relevanter demografischer Kovariaten der Schülerinnen und Schüler von 2015 nach 2019 konstant geblieben wäre.

5.3 Methodisches Vorgehen

5.3.1 Moduseffekte

Für Deutschland liegen in TIMSS 2019 Daten für eine PBA- und eine CBA-Stichprobe in der vierten Jahrgangsstufe vor. Die Schülerinnen und Schüler stammen aus insgesamt 245 Schulen. In 36 dieser Schulen wurden sowohl CBA- als auch PBA-Testungen (allerdings mit verschiedenen Schülerinnen und Schülern) durchgeführt. Insgesamt 42 Schulen nahmen ausschließlich an der PBA-Testung teil, 167 Schulen ausschließlich an der CBA-Testung. In die Analyse der Moduseffekte gehen $N = 1495$ Schülerinnen und Schüler der PBA-Testung und $N = 3429$ Schülerinnen und Schüler der CBA-Testung ein, für die Itemantworten für die PBA-Trenditems beziehungsweise CBA-Trenditems in Mathematik und Naturwissenschaften vorliegen. Die Bestimmung der Moduseffekte beruht auf Skalierungen, in denen nur die Trenditems verwendet werden. Insgesamt werden 92 Trenditems in Mathematik und 94 Trenditems in Naturwissenschaften verwendet. Für die beiden Domänen liegen im Mittel Antworten von 490 Schülerinnen und Schülern pro Item ($Min = 482$, $Max = 501$) im CBA-Modus und 374 Schülerinnen und Schülern ($Min = 367$, $Max = 380$) im PBA-Modus vor.

Zur Bestimmung der Standardfehler wird im internationalen Vorgehen in TIMSS 2019 das *Jackknife*-Verfahren (Kolenikov, 2010) auf Basis von 75 *Jackknife*-Zonen angewendet. Die *Jackknife*-Zonen wurden von der internationalen Studienleitung für die PBA- und CBA-Stichprobe unabhängig voneinander definiert. Da die Analyse der Moduseffekte auf der PBA- und der CBA-Stichprobe beruht, liegt für knapp die Hälfte der Schulen der PBA-Stichprobe auch eine CBA-Testung vor. Diese zusätzliche Abhängigkeit erlaubt eine bessere Kontrolle unbeobachteter Effekte auf der Schulebene für den Vergleich beider Modi. Standardfehler fallen daher typischerweise kleiner aus als bei unabhängiger Behandlung der Schülerinnen und Schüler aus der PBA- und der CBA-Stichprobe. Daher wurden für die Analyse der Moduseffekte 42 *Jackknife*-Zonen neu definiert, sodass dieselben Schulen der PBA- und der CBA-Stichprobe in derselben *Jackknife*-Zone liegen. Die Information über das Stratum für die Stichprobenziehung der Schulen ging dabei in die Definition der *Jackknife*-Zonen ein.

Die deutsche PBA- und CBA-Stichprobe ist unter Verwendung von Stichprobengewichten repräsentativ. Zur Verringerung zufallsbedingter Abweichungen der PBA-Stichprobe von der CBA-Stichprobe wurden die von der internationalen Studienleitung bereitgestellten Stichprobengewichte adjustiert, sodass die marginalen Verteilungen relevanter demografischer Kovariaten der Schülerinnen und Schüler mit denen der PBA-Stichprobe übereinstimmen. Dabei wurde die

Raking-Methode (auch: *iterative proportional fitting*; Kolenikov, 2014) verwendet. Aus dem Schülerfragebogen gingen dabei folgende Variablen ein: *Geschlecht*, *Häufigkeit*, mit der zu Hause Deutsch gesprochen wird, *Migrationshintergrund*, *Anzahl Bücher zu Hause*, die Besitzstandsvariablen *Computer/Tablet*, *eigener Schreibtisch*, *eigenes Kinderzimmer*, *eigenes Handy*. Aus dem Elternfragebogen gingen die Variablen *häusliche Lernressourcen*, *höchster Berufsstatus der Eltern*, *höchster Bildungsabschluss der Eltern* sowie die *Anzahl der Bücher zu Hause* ein. Zusätzlich wurde die Stratumvariable verwendet. In der *Raking*-Methode wurden alle Variablen als kategorial behandelt. Fehlende Werte wurden als eine eigene Kategorie definiert. Zur Vermeidung extremer Stichprobengewichtungen wurde ein Trimmen der Gewichte vorgenommen (Kolenikov, 2014). Als Ergebnis der Adjustierung der Stichprobengewichte in der PBA-Stichprobe stimmen die marginalen Verteilungen der gewichteten PBA- und CBA-Stichprobe überein.

Für die Bestimmung von Moduseffekten wurde auf Modelle der *Item-Response-Theory* (IRT; Rost, 2004) zurückgegriffen. Dabei wird im Gegensatz zum internationalen Vorgehen in TIMSS ein 2PL-Modell (dichotome Items; Birnbaum, 1968) beziehungsweise generalisiertes *Partial-Credit*-Modell (polytome Items; Muraki, 1992) verwendet, da die Stichprobengrößen pro Item keine zuverlässigen Schätzungen der Itemparameter im 3PL-Modell erlauben. Die Itemantworten der Schülerinnen und Schüler wurden zunächst separat für die PBA- und die CBA-Stichprobe unter Verwendung von Stichprobengewichten skaliert. Alle fehlenden Itemantworten (*omitted* und *not reached*) wurden in den Skalierungen als falsch kodiert (Pohl & Becker, 2020, für einen Überblick alternativer Modelle). Die in den Skalierungen gewonnenen Itemparameter wurden mit der *Linking*-Methode nach Haebara (Kolen & Brennan, 2014) auf dieselbe Metrik gebracht. Dabei wurde die Verteilung der PBA-Stichprobe als Referenz gewählt, das heißt $M_{\text{PBA}} = 0$ und $SD_{\text{PBA}} = 1$. Der Mittelwert M_{CBA} und die Standardabweichung SD_{CBA} werden als Ergebnis der Haebara-Methode gewonnen. Im Fall von $M_{\text{CBA}} \neq 0$ oder $SD_{\text{CBA}} \neq 1$ wird vom Vorliegen eines Moduseffekts gesprochen. Zusätzlich zur (gewöhnlichen) Haebara-Methode wurde zur Validierung der Befunde die sogenannte robuste Haebara-Methode eingesetzt (mit der Potenz $p = 0.5$; He & Cui, 2020; Robitzsch, 2020). Die robuste Haebara-Methode entspricht dabei empirisch dem Vorgehen einer Skalierung unter partieller Invarianz, wobei Items mit stark abweichenden Itemparametern zwischen PBA- und CBA-Testmodus modusspezifische Parameter erhalten (Robitzsch & Lüdtke, 2020; Davier, Khorrarnadel, He, Shin & Chen, 2019).

Für die Analyse der Moduseffekte werden folgende Größen betrachtet. Zunächst wird ein standardisierter Mittelwertsunterschied $d = (M_{\text{CBA}} - M_{\text{PBA}}) / SD_{\text{PBA}}$ zwischen PBA- und CBA-Modus berechnet (Robitzsch et al., 2017). Diese Effektgröße d bildet primär mittlere Unterschiede zwischen den Testmodi in den Itemschwierigkeiten ab. Mittlere Unterschiede in den Itemdiskriminationen werden durch das Verhältnis der Standardabweichungen $e = SD_{\text{CBA}} / SD_{\text{PBA}}$ erfasst. Die itemspezifische Heterogenität der Moduseffekte wird durch Korrelationen der Itemschwierigkeiten und Itemdiskriminationen zwischen den Testmodi bestimmt (Goldhammer et al., 2019).

Für die statistische Inferenz der gewonnenen Größen wird die durch Personen und Items verursachte Variabilität quantifiziert. Der Standardfehler (*standard error*, *SE*) für die Stichprobenziehung von Schülerinnen und Schülern wurde mit der *Jackknife*-Methode bestimmt. Dabei wurde die Skalierung mit anschließendem *Linking* entsprechend der Anzahl der *Jackknife*-Zonen wiederholt durchgeführt. Die Unsicherheit bezüglich der Itemauswahl wird als *Linking*-Fehler (*lin-*

linking error, LE) bezeichnet (Monseur, Sibberns & Hastedt, 2008; Wu, 2010). Der *Linking-Fehler* wird unter *Jackknife* von Items bestimmt (Monseur & Berezner, 2007), indem jeweils ein Item für die Berechnung herausgelassen wird. Bei I verwendeten Items kann die durch Items bedingte Unsicherheit auf ein einzelnes Item bezogen werden, indem die Größe $LE_1 = \sqrt{I} \cdot LE$ berechnet wird. Der auf ein Item bezogene *Linking-Fehler* LE_1 hat den Vorteil, unabhängig von der Anzahl der Items zu sein. Für den d -Wert bedeutet dies beispielsweise, dass LE_1 für d die Standardabweichung der itemspezifischen Moduseffekte schätzt. Der totale Fehler (*total error*, TE) einer interessierenden Größe wurde gemäß $TE = \sqrt{(SE^2 + LE^2)}$ berechnet und für die statistische Inferenz verwendet. In früheren TIMSS-Publikationen wurde der *Linking-Fehler* für Trendschätzungen im PBA-Modus häufig nicht betrachtet, da die Größe der *Linking-Fehler* im Vergleich zum Standardfehler als vernachlässigbar angesehen werden kann (Martin & Mullis, 2019). Allerdings betrachtet die für TIMSS vorgeschlagene Methode zur Berechnung der *Linking-Fehler* (Martin, Mullis, Foy, Brossman & Stanco, 2012) im Gegensatz zu den in PISA eingesetzten Methoden (Monseur & Berezner, 2007; OECD, 2014, 2017) nicht die durch die Itemauswahl bedingte Unsicherheit.

Alle statistischen Analysen dieses Kapitels wurden mit der Statistiksoftware R (R Core Team, 2020) durchgeführt. Für die IRT-Skalierung wurde das R-Paket TAM (Robitzsch, Kiefer & Wu, 2020) verwendet.

5.3.2 Trendschätzung

Für die Schätzung des Trends von TIMSS 2015 zu 2019 wurden die PBA-Stichprobe aus TIMSS 2015 ($N = 3\,948$) sowie die PBA-Stichprobe ($N = 1\,505$) und CBA-Stichprobe ($N = 3\,437$) aus TIMSS 2019 verwendet. Für die originalen und marginalen Trends wurden jeweils die von der internationalen Studienleitung gelieferten Stichprobengewichte genutzt. Eine Ausnahme stellt der bedingte marginale Trend im PBA-Modus dar, für den die deutsche PBA-Stichprobe in TIMSS 2019 mit dem Verfahren des *Raking* (siehe Abschnitt 5.3.1) neu gewichtet wurde, sodass die marginalen Verteilungen relevanter demografischer Kovariaten der Schülerinnen und Schüler mit den Verteilungen aus der PBA-Stichprobe in TIMSS 2015 übereinstimmen. Als Kovariaten wurden nur Variablen aus dem Schülerfragebogen verwendet.¹ Folgende in TIMSS 2015 und 2019 enthaltenen Variablen wurden für das *Raking* genutzt: *Geschlecht*, *Häufigkeit*, mit der zu Hause Deutsch gesprochen wird, *Migrationshintergrund*, *Anzahl Bücher zu Hause*, Besitzstandsvariablen *Schreibtisch*, *eigenes Kinderzimmer*, *verfügbare Internetverbindung*, *eigenes Handy*.

Für Mathematik wurden in den Skalierungen 77 veröffentlichte Items verwendet, die nur in TIMSS 2015 (PBA) auftraten. 92 Trenditems liegen sowohl im PBA- als auch CBA-Modus in TIMSS 2019 vor. Insgesamt 79 neue Items liegen nur im CBA-Modus in TIMSS 2019 vor. Für die Naturwissenschaften bestand der Itempool aus 74 veröffentlichten Items aus TIMSS 2015 (PBA),

¹ Da der Elternfragebogen in den beiden TIMSS-Erhebungen 2015 und 2019 nur in etwa 65 Prozent der Fälle ausgefüllt vorlag (siehe Kapitel 2 in diesem Band; Wendt et al., 2016), können in den Analysen nur die Daten der Schülerinnen und Schüler berücksichtigt werden. Mithilfe der multiplen Imputation wären genauere Analysen mit den Angaben der Erziehungsberechtigten möglich gewesen. Diese sind jedoch deutlich aufwendiger und können in diesem Rahmen nicht geleistet werden.

94 Trenditems im PBA- und CBA-Modus in TIMSS 2019 sowie 74 neuen Items im CBA-Modus in TIMSS 2019.

Für originale Trendschätzungen wurden die von der internationalen Studienleitung zur Verfügung gestellten *plausible values* aus TIMSS 2015 (PBA) und TIMSS 2019 (PBA und CBA) unter Verwendung der Stichprobengewichte genutzt (orig. PBA→PBA und orig. adj. PBA→CBA). Für die marginalen Trendschätzungen wurden wiederum das 2PL- beziehungsweise generalisierte *Partial-Credit*-Modell als Skalierungsmodell unter Verwendung der Stichprobengewichte verwendet. Zunächst wurden drei separate Skalierungen für die nationalen Stichproben 2015 (PBA) und 2019 (PBA und CBA) durchgeführt. Die Itemparameter wurden als Input für die (robuste) Haebara-Methode genutzt, sodass marginale PBA→PBA-Trends 2015→2019 im PBA-Modus (marg. PBA→PBA) sowie ein unadjustierter PBA→CBA-Trend von PBA 2015 nach CBA 2019 (marg. unadj. PBA→CBA) ermittelt wurden. In den marginalen Trends wurden die Mittelwerte und Standardabweichungen so linear transformiert, dass M und SD den Werten aus TIMSS 2015 (PBA) entsprechen.

Für die Berechnung der Standardfehler wurden die originalen 75 *Jackknife*-Zonen verwendet. Dabei wurde die Skalierung mit anschließendem *Linking* wiederholt für die 75 *Jackknife*-Stichproben durchgeführt. Sowohl für den PBA-Trend als auch den PBA-CBA-Trend wurden *Linking*-Fehler bestimmt, indem jeweils ein Item (sowohl Trenditems als auch Nicht-Trenditems) aus den Analysen herausgelassen wurde. In den empirischen Analysen zeigte sich jedoch, dass der *Linking*-Fehler zum größten Teil auf die Trenditems (mehr als 95 % der Variabilität) zurückging, sodass für den auf ein Item bezogenen *Linking*-Fehler in der Formel $LE_1 = \sqrt{I} \cdot LE$ für I die Anzahl der Trenditems in der jeweiligen Domäne gewählt wurde.

5.4 Ergebnisse

5.4.1 Moduseffekte

Im Folgenden werden Moduseffekte für die nationale Stichprobe in TIMSS 2019 berichtet. In Tabelle 5.1 sind die Effektgrößen für Moduseffekte d der standardisierten Mittelwertdifferenz und e für das Verhältnis der Standardabweichungen angegeben. Items auf dem Computer waren im Mittel in beiden Domänen signifikant schwieriger als auf Papier (Mathematik: $d = -0.152$, $p = .044$; Naturwissenschaften: $d = -0.159$, $p = .034$). Die bedeutsamen *Linking*-Fehler für den d -Wert (Mathematik: $LE_1 = 0.396$; Naturwissenschaften: $LE_1 = 0.336$) deuten darauf hin, dass nicht von einer homogenen Verschiebung der Itemschwierigkeiten auszugehen ist, sondern dass Moduseffekte in starkem Ausmaß itemspezifisch ausfallen (siehe auch die Befunde in PISA; Goldhammer et al., 2019).

Der Moduseffekt d unterscheidet sich jedoch nicht signifikant zwischen Mathematik und Naturwissenschaften ($\Delta d = 0.007$, $SE = 0.033$, $LE = 0.054$, $TE = 0.063$, $t = 0.11$, $p = .916$). Mit der robusten Haebara-Methode wurden sowohl in Mathematik ($d = -0.159$, $SE = 0.074$, $LE = 0.043$, $TE = 0.085$) als auch in den Naturwissenschaften ($d = -0.151$, $SE = 0.089$, $LE = 0.043$, $TE = 0.099$) sehr ähnliche Ergebnisse erhalten, sodass nicht davon auszugehen ist, dass die Moduseffekte bei einer konkurrenten Skalierung unter partieller Invarianz deutlich verschieden ausfallen würden.

Tabelle 5.1: Moduseffekte im Hinblick auf Itemschwierigkeiten (*d*) und Itemdiskriminationen (*e*) für TIMSS 2019

Domäne	Par	Est	SE	LE	LE _i	TE	t	p
Mathematik	<i>d</i>	-0.152	0.063	0.041	0.396	0.076	-2.02	.044
Naturwissenschaften		-0.159	0.067	0.035	0.336	0.075	-2.12	.034
Mathematik	<i>e</i>	0.894	0.036	0.037	0.357	0.051	-2.06	.040
Naturwissenschaften		1.001	0.046	0.036	0.347	0.058	0.03	.980

Anmerkungen: Par = Parameter; *d* = standardisierter Moduseffekt bzgl. des Mittelwertes; *e* = Moduseffekt bzgl. der SD; Est = Schätzung; SE = Standardfehler; LE = Linking-Fehler; LE_i = auf ein Item bezogener Linking-Fehler; TE = totaler Fehler

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Die mittleren Itemdiskriminationen zwischen beiden Testmodi sind in Tabelle 5.1 durch die Effektgröße *e* abgebildet. In Mathematik diskriminieren die Items im CBA-Modus signifikant geringer als im PBA-Modus ($e = 0.894$, $p = .040$). Dieser Befund wurde allerdings nicht in Naturwissenschaften vorgefunden ($e = 1.001$, $p = .980$). Der Unterschied im Moduseffekt *e* bezüglich der Standardabweichung zwischen Mathematik und Naturwissenschaften wurde jedoch knapp nicht signifikant ($\Delta e = -0.107$, $SE = 0.036$, $LE = 0.052$, $TE = 0.063$, $t = -1.71$, $p = .088$).²

Die Korrelationen der Itemschwierigkeiten zwischen dem PBA- und dem CBA-Modus fielen mit .960 in Mathematik und .952 in den Naturwissenschaften hoch aus. Ähnliche Ergebnisse wurden sowohl für die internationale Stichprobe (Foy et al., 2020) als auch für PISA 2018 berichtet (Goldhammer et al., 2019). Analog zu Befunden in PISA 2018 (Goldhammer et al., 2019) fielen die Korrelationen der Itemdiskriminationen jedoch nur in moderater Höhe aus (Mathematik: $r = .614$; Naturwissenschaften: $r = .596$).

5.4.2 Trendschätzung

Nachfolgend werden die in den Berichtsbänden dargestellten originalen Trendschätzungen von TIMSS 2015 nach 2019 mit marginalen Trendschätzungen verglichen, die nur auf den Stichproben der Schülerinnen und Schüler in Deutschland beruhen. Tabelle 5.2 zeigt Mittelwerte und Standardabweichungen in TIMSS 2015 und 2019.

Für die deutsche PBA-Stichprobe fielen die auf der originalen, der marginalen und der bedingt marginalen Trendschätzung beruhenden Mittelwerte und Standardabweichungen sehr ähnlich aus. Der von der internationalen Studienleitung berichtete Mittelwert in der deutschen CBA-Stichprobe ($M = 521.0$) war praktisch identisch. Insgesamt kann daher ein kohärentes Muster gleichbleibender Mathematikleistungen der Schülerinnen und Schüler in Deutschland von TIMSS 2015 zu 2019 festgestellt werden. Für die CBA-Stichprobe wurde außerdem ein marginaler unadjustierter Mittelwert von $M = 509.8$ bestimmt, der sich deutlich von der marginalen Schätzung des Mittelwertes von $M = 521.5$ unterscheidet. Die beiden Schätzungen erlauben auch eine indirekte Schätzung des Moduseffekts gemäß $d = (M_{\text{CBA}} - M_{\text{PBA}}) / SD_{\text{PBA}} = (509.8 - 521.5) / 72.4 = -0.161$, der sehr ähnlich

2 Würde man für die statistische Inferenz den Linking-Fehler ignorieren (d.h. man verwendet den Standardfehler *SE* anstelle des totalen Fehlers *TE* für die Bestimmung von *t* und *p*), so ergäbe sich ein signifikantes Ergebnis ($t = -2.99$, $p = .003$).

Tabelle 5.2: Mittelwerte und Standardabweichungen in TIMSS 2015 (PBA) und TIMSS 2019 (PBA und CBA)

		Mathematik				Naturwissenschaften			
		<i>M</i>		<i>SD</i>		<i>M</i>		<i>SD</i>	
		<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>	<i>Est</i>	<i>SE</i>
orig.	2015 PBA	521.6	2.0	65.4	1.2	528.5	2.4	69.7	1.3
orig.	2019 PBA	519.3	4.2	72.0	2.7	522.2	4.7	77.9	3.2
marg.	2019 PBA	521.5	4.8	72.4	2.9	521.9	5.4	80.6	3.4
bed. marg.	2019 PBA	521.2	4.7	71.1	2.9	522.4	5.2	79.8	3.6
orig. adj.	2019 CBA	521.0	2.3	69.6	1.4	518.3	2.2	77.0	1.6
marg. unadj.	2019 CBA	509.8	3.0	64.1	2.5	511.3	3.2	77.7	2.5

Anmerkungen: orig. = originale Schätzung (international berichtet); marg. = marginale Schätzung; bed. marg. = bedingte marginale Schätzung; orig. adj. = originale adjustierte Schätzung (international berichtet); marg. unadj. = marginale unadjustierte Schätzung; *Est* = Schätzung; *SE* = Standardfehler
 In Berichtsbänden publizierte Ergebnisse auf Basis der von der internationalen Studienleitung zur Verfügung gestellten plausible values sind fett markiert.
 Abweichungen zu den in 2015 berichteten Werten (Selter, Walter, Walther & Wendt, 2016; Steffensky, Kleickmann, Kasper & Köller, 2016) sind im Rundungsverfahren begründet.

zum aus der nationalen Stichprobe 2019 erhaltenen Moduseffekt $d = -0.152$ ausfällt (siehe Abschnitt 5.4.1). Für Naturwissenschaften fielen wiederum die originale, marginale und bedingt marginale Schätzung des Mittelwertes sehr ähnlich aus. Der mit der CBA-Stichprobe berechnete originale Mittelwert war mit $M = 518.3$ etwas geringer. Allerdings war der Unterschied zum originalen PBA-Mittelwert von $M = 522.2$ nicht statistisch signifikant. Wie für Mathematik ließ sich auch für die Naturwissenschaften der Moduseffekt d indirekt auch auf Basis des marginalen PBA- und CBA-Mittelwertes schätzen. Dabei ergibt sich $d = -0.132$, was leicht vom direkt ermittelten Moduseffekt von $d = -0.159$ (siehe Abschnitt 5.4.1) abweicht.

In Tabelle 5.3 sind die verschiedenen Trendschätzungen dargestellt. Für Mathematik fallen der PBA-Trend und der adjustierte PBA→CBA-Trend praktisch identisch aus. Der marginale unadjustierte signifikante PBA→CBA-Trend von -11.8 Punkten unterscheidet sich aber deutlich vom marginalen PBA-Trend. In Naturwissenschaften ist der Unterschied im marginalen PBA-Trend von -6.5 Punkten nicht signifikant, während der Leistungsabfall von -10.1 Punkten signifikant ausfällt ($\Delta M = -10.1$, $SE = 3.3$, $t = -3.11$, $p = .002$). Beide Trendschätzungen unterscheiden sich jedoch nicht signifikant voneinander. Der marginale unadjustierte PBA→CBA-Trend ist mit -17.2 Punkten noch etwas größer. Die Ergebnisse auf Basis der robusten Haebara-Methode fielen wie in Abschnitt 5.4.1 sehr ähnlich aus, sodass diese hier nicht näher aufgeführt werden. Dies illustriert, dass es nicht der Fall ist, dass nur einzelne Items deutlich schwieriger im CBA-Modus als im PBA-Modus ausfallen und die übrigen Items keine oder nur geringe Moduseffekte aufweisen, sondern dass von einer näherungsweise symmetrischen Verteilung itemspezifischer Moduseffekte ausgegangen werden kann.

Erwartungsgemäß fallen die itemspezifischen *Linking*-Fehler LE_1 für den reinen PBA-Trend (Mathematik: $LE_1 = 19.2$; Naturwissenschaften: $LE_1 = 24.9$) etwas niedriger als für den PBA→CBA-Trend (Mathematik: $LE_1 = 23.3$; Naturwissenschaften: $LE_1 = 32.9$) aus. Während in LE_1 für den PBA-Trend nur die längsschnittliche Variation in Itemparametern abgebildet wird, sind im PBA→CBA-Trend zusätzlich itemspezifische Moduseffekte enthalten.

Tabelle 5.3: Trendschätzungen für Mathematik und Naturwissenschaften für TIMSS 2015 nach TIMSS 2019

		Mathematik						
		<i>Est</i>	<i>SE</i>	<i>LE</i>	<i>LE_i</i>	<i>TE</i>	<i>t</i>	<i>p</i>
orig.	PBA → PBA	-2.4	4.7	–	–	4.7	-0.51	.614
marg.	PBA → PBA	-0.1	5.2	2.0	19.2	5.6	-0.02	.984
bed. marg.	PBA → PBA	-0.4	5.1	2.1	20.5	5.5	-0.07	.941
orig. adj.	PBA → CBA	-0.7	3.1	–	–	3.1	-0.21	.831
marg. unadj.	PBA → CBA	-11.8	3.6	2.4	23.3	4.4	-2.71	.007

		Naturwissenschaften						
		<i>Est</i>	<i>SE</i>	<i>LE</i>	<i>LE_i</i>	<i>TE</i>	<i>t</i>	<i>p</i>
orig.	PBA → PBA	-6.3	5.3	–	–	5.3	-1.18	.237
marg.	PBA → PBA	-6.5	5.9	2.6	24.9	6.5	-1.01	.312
bed. marg.	PBA → PBA	-6.1	5.7	2.6	25.6	6.3	-0.97	.333
orig. adj.	PBA → CBA	-10.1	3.3	–	–	3.3	-3.11	.002
marg. unadj.	PBA → CBA	-17.2	4.0	3.4	32.9	5.2	-3.27	.001

Anmerkungen: orig. = originale Schätzung (international berichtet); marg. = marginale Schätzung; bed. marg. = bedingte marginale Schätzung; orig. adj. = originale adjustierte Schätzung (international berichtet); marg. unadj. = marginale unadjustierte Schätzung; *Est* = Schätzung; *SE* = Standardfehler; *LE* = Linking-Fehler; *LE_i* = auf ein Item bezogener Linking-Fehler; *TE* = totaler Fehler
 In Berichtsbänden publizierte Ergebnisse auf Basis der von der internationalen Studienleitung zur Verfügung gestellten plausible values sind fett markiert.

5.5 Zusammenfassung

Mit den in diesem Kapitel vorgestellten vertiefenden Analysen der TIMSS-2019-Daten wurde zum einen der Frage nachgegangen, ob – ähnlich wie in anderen *Large-Scale-Assessments* (z.B. PISA 2015, PISA 2018) – der Wechsel von Papier-und-Bleistift-Tests auf computerbasierte Tests Effekte auf Itemeigenschaften hat. Erwartet wurde, dass die Itemschwierigkeiten beim CBA höher ausfallen als beim PBA. Hinsichtlich der Itemdiskriminationsparameter wurden keine gerichteten Hypothesen formuliert. Darüber hinaus wurde auch der Frage nachgegangen, ob ein alternatives Vorgehen bei der Trendschätzung – die marginale anstelle der originalen – Effekte auf die Ergebnisse der in Deutschland getesteten Schülerinnen und Schüler hat. Zwei wesentliche Befunde lassen sich auf Basis der Analysen festhalten:

- Ebenso wie für die Fünfzehnjährigen in PISA 2015 und PISA 2018 ergab sich für die Viertklässlerinnen und Viertklässler in TIMSS 2019, dass der Moduswechsel einen Effekt auf die Itemeigenschaften hatte: Die mit dem Computer administrierten Aufgaben sind im Mittel in beiden getesteten Domänen schwieriger und weisen (im Fach Mathematik) im Mittel geringere Itemdiskriminationen auf als die auf Papier vorgegebenen Aufgaben.
- Die im internationalen Bericht vorgestellten originalen Trendschätzungen für Deutschland, für die eine Adjustierung der CBA-Items bereits vorgenommen wurde, können durch die Analysen der nationalen Daten weitgehend gestützt werden, wenn auch die marginale PBA-Trendschätzung in den Naturwissenschaften für Deutschland etwas günstigere Ergebnisse aufweist.

Im Folgenden sollen beide Befunde diskutiert werden, bevor ein Ausblick auf zukünftige Erhebungen in TIMSS vorgenommen wird.

5.5.1 Moduseffekte

Die Umstellung der Testungen auf den Computer ist ein stetig fortschreitender Modernisierungsprozess internationaler Schulleistungsstudien. Als Folge der Covid-19-Pandemie wird die Bedeutung digitaler Medien im schulischen Kontext weiter steigen, sodass Leistungstestungen am Computer immer mehr dem Alltag der Schülerinnen und Schüler entsprechen werden. Mit dieser Umstellung verbunden – das zeigen die hier vorgenommenen Analysen – können sich Itemeigenschaften verändern, beispielsweise, indem die Items schwieriger werden. Bei Betrachtung der nationalen Daten fällt auch auf, dass der Moduseffekt in Deutschland größer ausfällt als international, dies gilt vor allem für die naturwissenschaftlichen Aufgaben. Die Anlage der TIMS-Studie 2019 ist ein gutes Beispiel dafür, wie man solchen unerwünschten Effekten bei der originalen und marginalen Trendschätzung begegnen kann. Ein Testdesign, das die Analyse von Moduseffekten erlaubt, führt dann auch zu weitgehend validen Schätzungen der Veränderungen, indem man die Leistungen um den Moduseffekt korrigiert.

Auch zukünftige TIMSS-Zyklen sollen computerbasiert durchgeführt werden. Es bleibt offen, ob aufgrund zunehmender Vertrautheit mit digitalen Medien die Leistungen am Computer in künftigen TIMSS-Zyklen nicht mehr geringer ausfallen als die Leistungen bei der papierbasierten Testung. Interindividuelle Unterschiede im Moduseffekt konnten durch basale Computerfähigkeiten nicht erklärt werden (Kroehne, Buerger et al., 2019). Daher könnte es für nachfolgende Erhebungen gewinnbringend sein, auch den unadjustierten CBA-Trend anzugeben, um festzustellen, ob die unadjustierten CBA-Testleistungen allmählich zu den PBA-Testleistungen aufschließen. Der (originale) adjustierte CBA-Trend birgt die Gefahr, dass eine vermeintlich positive Leistungsentwicklung eines Staates nur einen über die Zyklen kleiner werdenden Moduseffekt darstellt (Robitzsch, Lüdtke et al., 2020).

Es soll aber nicht der Eindruck erweckt werden, dass CBA-Testungen für Trendschätzungen generell kritisch zu hinterfragen sind. Vielmehr wird deutlich, dass die valide Interpretation von Testleistungen immer an konkrete Rahmenbedingungen der Testadministration geknüpft ist. Mithilfe randomisierter Moduseffekt-Studien lassen sich Moduseffekte korrigieren, solange die Voraussetzung der Konstruktäquivalenz (Pommerich, 2016) vorliegt. Künftige Forschung zu Moduseffekten sollte vermehrt vermittelnde Prozesse (z.B. Bearbeitungsgeschwindigkeit, Kroehne, Hahnel & Goldhammer, 2019), fehlende Antworten (z.B. hoher Anteil übersprungener Aufgaben, Buerger, Kroehne, Koehler & Goldhammer, 2019), strukturelle Unterschiede im Antwortprozess (z.B. für Textantworten, Zehner, Kroehne, Hahnel & Goldhammer, 2020) und das Antwortverhalten (z.B. die Anzahl geratener Antworten, Kroehne et al., 2020) in den Blick nehmen.

5.5.2 Trendanalysen

Für die Bildungspolitik in Deutschland ist es von großer Bedeutung, Trends im Bildungssystem zu erkennen und deren Implikationen für weiteres Handeln zu bewerten. Die international berichteten Befunde weisen in Deutschland auf stabile Leistungen in Mathematik zwischen 2015 und 2019 hin, in den Naturwissenschaften sinken die Leistungen leicht ab. Die marginalen Trendschätzungen stützen die internationale Befundlage, wenngleich sie das Absinken der Leistungen in den Naturwissenschaften etwas geringer erscheinen lassen. Die Abweichungen dürften vermutlich vor allem dadurch zustande kommen, dass der nationale Moduseffekt in Deutschland in den Naturwissenschaften stärker ausfällt als der internationale und die Adjustierung des originalen Trends in Deutschland etwas zu gering ausgefallen ist. Insgesamt bleibt es aber bei der in den Kapiteln 3 und 4 in diesem Band festgestellten Befundlage, dass sich die Leistungen zwischen 2015 und 2019 nur geringfügig verändert haben. Die Einordnung dieser Befunde wurde in den einschlägigen Kapiteln vorgenommen. Dort wurde auch diskutiert, wie Fördermaßnahmen aussehen können, um langfristig zu höheren Kompetenzständen der Schülerinnen und Schüler in Deutschland zu kommen. Es ist anzustreben, dass die in TIMSS gewählten Studiendesigns auch in Zukunft valide Trendschätzungen in Mathematik und den Naturwissenschaften zulassen.

Literatur

- Beaton, A. E. (1990). Introduction. In A. E. Beaton & R. Zwick (Hrsg.), *The effect of changes in the national assessment: Disentangling the NAEP 1985-86 reading anomaly* (Bericht Nr. 17-TR-21) (S. 1–13). Princeton, NJ: National Assessment of Educational Progress.
- Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B. & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *The Journal of Technology, Learning and Assessment*, 6(9).
- Birnbaum, A. (1968). Some latent trait models. In F. M. Lord & M. R. Novick (Hrsg.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Buerger, S., Kroehne, U., Koehler, C. & Goldhammer, F. (2019). What makes the difference? The impact of item properties on mode effects in reading assessments. *Studies in Educational Evaluation*, 62, 1–9. <https://doi.org/10.1016/j.stueduc.2019.04.005>
- Carstensen, C., Prenzel, M. & Baumert, J. (2009). Trendanalysen in PISA: Wie haben sich die Kompetenzen in Deutschland zwischen PISA 2000 und PISA 2006 entwickelt? *Vertiefende Analysen zu PISA 2006. Zeitschrift für Erziehungswissenschaft, Sonderheft 10* (S. 11–34). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91815-0_2
- Davies, M. von, Khorramdel, L., He, Q., Shin, H. J. & Chen, H. (2019). Developments in psychometric population models for technology-based large-scale assessments: An overview of challenges and opportunities. *Journal of Educational and Behavioral Statistics*, 44(6), 671–705. <https://doi.org/10.3102/1076998619881789>
- Fishbein, B., Martin, M. O., Mullis, I. V. & Foy, P. (2018). The TIMSS 2019 item equivalence study: Examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education*, 6, 11. <https://doi.org/10.1186/s40536-018-0064-z>
- Foy, P., Yin, L., & Fishbein, B. (2020, Juni). *Scaling the TIMSS 2019 achievement data*. Talk held at the 8th Meeting of National Research Coordinators.
- Gebhardt, E. & Adams, R. J. (2007). The influence of equating methodology on reported trends in PISA. *Journal of Applied Measurement*, 8(3), 305–322.
- Goldhammer, F., Harrison, S., Bürger, S., Kröhne, U., Lüdtke, O., Robitzsch, A., Köller, O., Heine, J.-H. & Mang, J. (2019). Vertiefende Analysen zur Umstellung des Modus

- von Papier auf Computer. In K. Reiss, M. Weis, E. Klieme & O. Köller (Hrsg.), *PISA 2018. Grundbildung im internationalen Vergleich* (S. 163–186). Münster: Waxmann.
- Hamhuis, E., Glas, C. & Meelissen, M. (2020). Tablet assessment in primary education: Are there performance differences between TIMSS' paper-and-pencil test and tablet test among Dutch grade-four students? *British Journal of Educational Technology*. Advance online publication. <https://doi.org/10.1111/bjet.12914>
- He, Y. & Cui, Z. (2020). Evaluating robust scale transformation methods with multiple outlying common items under IRT true score equating. *Applied Psychological Measurement*, 44(4), 296–310. <https://doi.org/10.1177/0146621619886050>
- Jerrim, J. (2016). PISA 2012: How do results for the paper and computer tests compare? *Assessment in Education: Principles, Policy & Practice*, 23(4), 495–518. <https://doi.org/10.1080/0969594X.2016.1147420>
- Jerrim, J., Micklewright, J., Heine, J.-H., Sälzer, C. & McKeown, C. (2018). PISA 2015: how big is the 'mode effect' and what has been done about it? *Oxford Review of Education*, 44(4), 476–493. <https://doi.org/10.1080/03054985.2018.1430025>
- Kingston, N. M. (2008). Comparability of computer- and paper-administered multiple-choice tests for K–12 populations: A synthesis. *Applied Measurement in Education*, 22(1), 22–37. <https://doi.org/10.1080/08957340802558326>
- Klieme, E. & Baumert, J. (2001). Identifying national cultures of mathematics education: Analysis of cognitive demands and differential item functioning in TIMSS. *European Journal of Psychology of Education* 16(3), 385–402. <https://doi.org/10.1007/BF03173189>
- Kolen, M. J. & Brennan, R. L. (2014). *Test equating, scaling, and linking*. New York: Springer. <https://doi.org/10.1007/978-1-4939-0317-7>
- Kolenikov, S. (2010). Resampling variance estimation for complex survey data. *The Stata Journal*, 10(2), 165–199. <https://doi.org/10.1177/1536867X1001000201>
- Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (raking). *The Stata Journal*, 14(1), 22–59. <https://doi.org/10.1177/1536867X1401400104>
- Kroehne, U., Buerger, S., Hahnel, C. & Goldhammer, F. (2019). Construct equivalence of PISA reading comprehension measured with paper-based and computer-based assessments. *Educational Measurement: Issues and Practice*, 38(3), 97–111. <https://doi.org/10.1111/emip.12280>
- Kroehne, U., Deribo, T. & Goldhammer, F. (2020). Rapid guessing rates across administration mode and test setting. *Psychological Test and Assessment Modeling*, 62(2), 147–177.
- Kroehne, U., Hahnel, C. & Goldhammer, F. (2019). Invariance of the response processes between gender and modes in an assessment of reading. *Frontiers in Applied Mathematics and Statistics*, 5, 2. <https://doi.org/10.3389/fams.2019.00002>
- Kroehne, U. & Martens, T. (2011). Computer-based competence tests in the national educational panel study: The challenge of mode effects. *Zeitschrift für Erziehungswissenschaft*, 14(2), 169–186. <https://doi.org/10.1007/s11618-011-0185-4>
- Lüdtke, O. & Robitzsch, A. (2017). Eine Einführung in die Plausible-Values-Technik für die psychologische Forschung. *Diagnostica*, 63(3), 193–205. <https://doi.org/10.1026/0012-1924/a000175>
- Mangen, A., Walgermo, B. R. & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, 58, 61–68. <https://doi.org/10.1016/j.ijer.2012.12.002>
- Martin, M. O. & Mullis, I. V. (2019). TIMSS 2015: Illustrating advancements in large-scale international assessments. *Journal of Educational and Behavioral Statistics*, 44(6), 752–781. <https://doi.org/10.3102/1076998619882030>
- Martin, M. O., Mullis, I. V., Foy, P., Brossman, B. & Stanco, G. M. (2012). Estimating linking error in PIRLS. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 5, 35–47.
- Martin, M. O., Mullis, I. V. S. & Hooper, M. (Hrsg.). (2016). *Methods and procedures in TIMSS 2015*. Verfügbar unter: <http://timssandpirls.bc.edu/publications/timss/2015-methods.html>
- Mazzeo, J. & Davier, M. von (2008). Review of the Programme for International Student Assessment (PISA) test design: Recommendations for fostering stability in assessment results. *Education Working Papers EDU/PISA/GB (2008)*, 28.

- Monseur, C. & Berezner, A. (2007). The computation of equating errors in international surveys in education. *Journal of Applied Measurement*, 8(3), 323–335.
- Monseur, C., Sibberns, H. & Hastedt, D. (2008). Linking errors in trend estimation for international surveys in education. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 1, 113–122.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159–176.
<https://doi.org/10.1177/014662169201600206>
- OECD. (2014). *PISA 2012 technical report*. Paris: OECD Publishing.
- OECD. (2017). *PISA 2015 technical report*. Paris: OECD Publishing.
- Pohl, S. & Becker, B. (2020). Performance of missing data approaches under nonignorable missing data conditions. *Methodology*, 16(2), 147–165.
<https://doi.org/10.5964/meth.2805>
- Pommerich, M. (2016). The fairness of comparing test scores across different tests or modes of administration. In N. J. Dorans & L. L. Cook (Hrsg.), *Fairness in educational assessment and measurement* (S. 111–134). New York, NY: Routledge.
<https://doi.org/10.4324/9781315774527>
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria. Verfügbar unter: <https://www.R-project.org/>
- Reiss, K., Weis, M., Klieme, E. & Köller, O. (Hrsg.). (2019). *PISA 2018. Grundbildung im internationalen Vergleich*. Münster: Waxmann.
<https://doi.org/10.31244/9783830991007>
- Robitzsch, A. (2020). Robust Haebara linking for many groups: Performance in the case of uniform DIF. *Psych*, 2(3), 155–173. <https://doi.org/10.3390/psych2030014>
- Robitzsch, A., Kiefer, T. & Wu, M. (2020). *TAM: Test analysis modules*. R package version 3.5-19. Verfügbar unter: <http://CRAN.R-project.org/package=TAM>
- Robitzsch, A. & Lüdtke, O. (2019). Linking errors in international large-scale assessments: Calculation of standard errors for trend estimation. *Assessment in Education: Principles, Policy & Practice*, 26(4), 444–465.
<https://doi.org/10.1080/0969594X.2018.1433633>
- Robitzsch, A. & Lüdtke, O. (2020). A review of different scaling approaches under full invariance, partial invariance, and noninvariance for cross-sectional country comparisons in large-scale assessments. *Psychological Test and Assessment Modeling*, 62(2), 233–279.
- Robitzsch, A., Lüdtke, O., Goldhammer, F., Kroehne, U. & Köller, O. (2020). Reanalysis of the German PISA data: A comparison of different approaches for trend estimation with a particular emphasis on mode effects. *Frontiers in Psychology*, 11, 884.
<https://doi.org/10.3389/fpsyg.2020.00884>
- Robitzsch, A., Lüdtke, O., Köller, O., Kroehne, U., Goldhammer, F. & Heine, J.-H. (2017). Herausforderungen bei der Schätzung von Trends in Schulleistungsstudien: Eine Skalierung der deutschen PISA-Daten. *Diagnostica*, 63(2), 148–165.
<https://doi.org/10.1026/0012-1924/a000177>
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.
- Sachse, K. A., Roppelt, A. & Haag, N. (2016). A comparison of linking methods for estimating national trends in international comparative large-scale assessments in the presence of cross-national DIF. *Journal of Educational Measurement*, 53(2), 152–171.
<https://doi.org/10.1111/jedm.12106>
- Selter, C., Walter, D., Walther, G. & Wendt, H. (2016). Mathematische Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 79–136). Münster: Waxmann.
- Steffensky, M., Kleickmann, T., Kasper, D. & Köller, O. (2016). Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 137–188). Münster: Waxmann.
- Wang, S., Jiao, H., Young, M. J., Brooks, T. & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments: A meta-analysis

- of testing mode effects. *Educational and Psychological Measurement*, 68(1), 219–238.
<https://doi.org/10.1177/0013164407305592>
- Wendt, H., Bos, W., Selter, C., Köller, O., Schwippert, K. & Kasper, D. (Hrsg.). (2016). *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Wu, M. (2010). Measurement, sampling, and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice*, 29(4), 15–27.
<https://doi.org/10.1111/j.1745-3992.2010.00190.x>
- Zehner, F., Kroehne, U., Hahnel, C. & Goldhammer, F. (2020). PISA reading: Mode effects unveiled in short text responses. *Psychological Test and Assessment Modeling*, 62(1), 85–105.

Kapitel 6

Einblicke in die Gestaltung des Mathematik- und naturwissenschaftsbezogenen Sachunterrichts an Grundschulen in Deutschland

Justine Stang¹, Chantal Lepper¹, Mirjam Steffensky und Nele McElvany

6.1 Einleitung

Unter dem Begriff *Unterrichtsqualität* werden verschiedene Unterrichtsmerkmale, die in einem positiven Zusammenhang mit Unterrichts- und Erziehungszielen stehen, zusammengefasst (Einsiedler, 2017). Die Unterrichtsmerkmale können in Sicht- und Tiefenstrukturen eingeteilt werden, wobei insbesondere die nicht direkt beobachtbaren Tiefenstrukturmerkmale des Unterrichts als relevant für die Lern- und Entwicklungsprozesse im Unterricht gelten (Kunter & Trautwein, 2013; Seidel & Shavelson, 2007). Im deutschsprachigen Raum hat sich eine dreigliedrige Konzeptualisierung etabliert, welche zentrale Tiefenstrukturmerkmale in Klassenführung, kognitive Aktivierung und konstruktive Unterstützung² differenziert (Klieme, Lipowsky, Rakoczy & Ratzka, 2006; Klieme & Rakoczy, 2008; Kunter & Voss, 2011; Praetorius, Klieme, Herbert & Pinger, 2018). Diese drei Tiefenstrukturmerkmale werden als die Basisdimensionen von Unterrichtsqualität beschrieben. Auch international werden vergleichbare Konzeptionen zugrunde gelegt. Beispielsweise wird im generischen CLASS-System (Pianta & Hamre, 2009) zwischen der Klassenführung, der instruktionalen Unterstützung und der emotionalen Unterstützung unterschieden. Die instruktionale und die emotionale Unterstützung zeigen Überschneidungen mit der kognitiven Aktivierung respektive der konstruktiven Unterstützung.

¹ geteilte Erstautorenschaft

² In TIMSS 2015 wurde für dieselbe inhaltliche Skala der Begriff *Unterstützendes Klima* verwendet, der einen wichtigen und großen Teilaspekt von konstruktiver Unterstützung ausmacht.

6.2 Basisdimensionen der Unterrichtsqualität

Die drei Basisdimensionen Klassenführung, kognitive Aktivierung und konstruktive Unterstützung sind durch verschiedene Charakteristika gekennzeichnet. Die Klassenführung fokussiert auf eine möglichst effektive Nutzung der Unterrichtszeit für Lernprozesse, indem beispielsweise Unterrichtsstörungen und Disziplinprobleme durch die Einführung von Klassenregeln, routinisierten Abläufen und transparenten Strukturen präventiv vermieden werden (Hochweber, Hosenfeldt & Klieme, 2014; Kuger, 2016). Die Klassenführung wird als wichtige Rahmenbedingung für produktive Lernprozesse erachtet, da sie als Fundament für die Dimensionen kognitive Aktivierung und konstruktive Unterstützung im Unterricht dient. Die kognitive Aktivierung im Unterricht zielt darauf ab, Lernende im Unterricht zu anspruchsvollen Denkprozessen anzuregen. Dabei sind zum Beispiel die Vorwissensaktivierung, die Anregung von Problemlösungsstrategien und vertieften Denkprozessen wesentlich, die beispielsweise durch die Bereitstellung angemessen herausfordernder Lernaufgaben adressiert werden können (Baumert et al., 2010; Lipowsky et al., 2009). Die konstruktive Unterstützung der Lernenden im Unterricht umfasst Merkmale, die zu einer lernförderlichen und schülerorientierten Lernumgebung beitragen. Hierzu gehören beispielsweise sowohl eine positive Fehlerkultur im Unterricht als auch ein wertschätzender Umgang zwischen Lernenden untereinander sowie zwischen Lernenden und der Lehrkraft (Kunter & Voss, 2011; Rakoczy, 2008). Zudem umfasst die konstruktive Unterstützung strukturierende Maßnahmen wie die geeignete Sequenzierung von Inhalten, mit denen Lernende unterstützt werden, kognitiv anspruchsvolle Aufgaben zu bewältigen (Kunter & Voss, 2011). Insgesamt zielt die konstruktive Unterstützung darauf ab, motivationsrelevanten Bedürfnissen im Unterricht wie dem Erleben von Autonomie, Kompetenz und sozialer Eingebundenheit zu entsprechen (Ryan & Deci, 2000).

Empirisch zeigte sich in nationalen wie internationalen (Vergleichs-)Arbeiten, dass die drei Basisdimensionen sowohl mit der Schülerinnen- und Schülerleistung als auch mit der Schülerinnen- und Schülermotivation positiv zusammenhängen (z.B. Dietrich, Dicke, Kracke & Noack, 2015; Fauth, Decristan, Rieser, Klieme & Büttner, 2014a; Kunter et al., 2013; Praetorius et al., 2018). Die Bedeutung der Basisdimensionen zeichnet sich dabei einerseits für verschiedene Fächer und andererseits auch für verschiedene Domänen und Bildungsstufen ab (z.B. Decristan et al., 2016; Pianta & Hamre, 2009). Rieser, Stahns, Walzebug und Wendt (2016) konnten anhand der Daten der *Trends in International Mathematics and Science Study* (TIMSS) 2015 aufzeigen, dass Zusammenhänge zwischen der Wahrnehmung der Unterrichtsqualität und dem Leistungsstand der Lernenden bestehen: Der Anteil an Lernenden, der sich im Mathematikunterricht kognitiv aktiviert fühlte, nahm mit zunehmender Leistung ab. Ein ähnliches Muster wurde für den Sachunterricht beobachtet (Rieser et al., 2016). Insgesamt muss einschränkend festgehalten werden, dass die Befundlage zur Bedeutsamkeit der drei Basisdimensionen für leistungsbezogene und motivationale Lernergebnisse sowohl hinsichtlich der Stärke als auch der Richtung der Effekte inkonsistent ist (z.B. Praetorius et al., 2018). Im Grundschulkontext liegen zudem weniger Befunde vor als für die Sekundarstufe (z.B. Grundschule: IGEL, Fauth, Decristan, Rieser, Klieme & Büttner, 2014b; PERLE, Lotz, 2016; PLUS, Kleickmann, Steffensky & Praetorius, 2020; Sekundarstufe: PISA, Schiepe-Tiska et al., 2013; Pythagoras, Klieme et al., 2006).

Um Einblicke in die Gestaltung des Grundschulunterrichts geben zu können, wurden die drei Basisdimensionen der Unterrichtsqualität in den Unterrichtsfächern Mathematik und Sachunterricht in TIMSS 2019 erhoben und genau wie in TIMSS 2015 erfasst (Rieser et al., 2016). Im Folgenden wird berichtet, wie die Unterrichtsqualität aus der Sicht von Lernenden am Ende der Grundschulzeit in den jeweiligen Fächern aktuell ausgeprägt ist, wo sie sich im internationalen Vergleich verorten lässt und wie sie sich im Vergleich zu TIMSS 2015 verändert hat.

6.3 Aktueller Stand TIMSS 2019 – internationaler und nationaler Vergleich

Um die drei Basisdimensionen der Unterrichtsqualität aus der Sicht von Schülerinnen und Schülern darzustellen, wurden drei Skalen gebildet, wobei die Zuordnung der Items zu den einzelnen Dimensionen in Übereinstimmung mit den Basisdimensionen der Unterrichtsqualität in TIMSS 2015 erfolgte (Wendt, Bos, Goy & Jusufi, 2017). Die Lernenden wurden auf einer vierstufigen Likertskala gefragt, wie stark sie verschiedenen Aussagen zustimmen (1 = *in jeder Stunde/stimme völlig zu* bis 4 = *nie oder fast nie/stimme überhaupt nicht zu*). Die Klassenführung im Mathematik- beziehungsweise Sachunterricht wurde mithilfe von jeweils fünf Items ermittelt. Zur Erfassung der Klassenführung im Mathematik- beziehungsweise Sachunterricht wurde beispielsweise erfragt, ob die Lehrkräfte lange warten müssen, bis alle Schülerinnen und Schüler leise sind. Die kognitive Aktivierung wurde mittels sieben Items erhoben. Lernende wurden zum Beispiel gefragt, ob die Lehrkraft im Mathematik- beziehungsweise Sachunterricht möchte, dass die Schülerinnen und Schüler ihre Antworten erklären. Die konstruktive Unterstützung fokussiert sowohl auf affektiv-motivationale Aspekte, beispielsweise die Fehlerkultur oder die Beziehungsqualität zwischen Lehrkraft und Lernenden, als auch auf unterstützende Aspekte wie zum Beispiel die Lernbegleitung und wurde anhand von fünf Items erfasst. In den Abbildungen 6.1 und 6.2 sind jeweils die prozentualen Verteilungen der Antworthäufigkeiten sowie die mittleren Skalenwerte der drei Basisdimensionen für die verschiedenen Domänen dargestellt. Alle Items wurden so kodiert, dass höhere Werte höhere Ausprägungen darstellen. Die Reliabilitäten der drei Skalen zur Erfassung der Dimensionen der Unterrichtsqualität waren zufriedenstellend bis gut.

Zur Befragung von Schülerinnen und Schülern zum Sachunterricht ist Folgendes anzumerken: Im Rahmen von TIMSS interessieren in erster Linie die naturwissenschaftsbezogenen Anteile des Sachunterrichts. Der Sachunterricht ist aber ein mehrperspektivisches Fach, das sozialwissenschaftliche und naturwissenschaftlich-technische Perspektiven umfasst (siehe Kapitel 4 in diesem Band). Befragungen von Lernenden zu einzelnen Bereichen des Sachunterrichts sind daher mit der Herausforderung verbunden, ob die Kinder zwischen dem Sachunterricht als Fach mit verschiedenen Inhaltsbereichen und den naturwissenschaftlichen Anteilen differenzieren können.

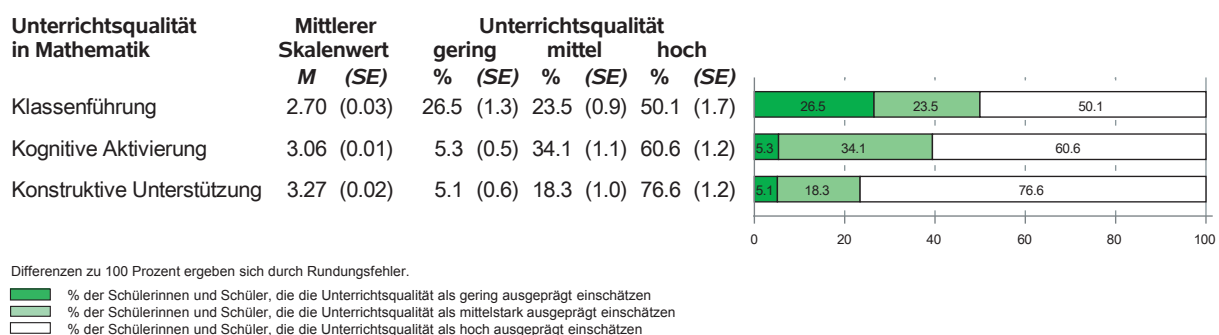
Sachunterrichtslehrkräfte benennen in der Regel die einzelnen Perspektiven nicht gezielt oder grenzen diese nicht voneinander ab, da das Selbstverständnis des Sachunterrichts ein perspektivenintegrierender Ansatz ist. Fragt man lediglich nach dem naturwissenschaftlichen Unterricht, ist unklar, ob alle Schülerinnen und Schüler eine Vorstellung von den dazugehörigen Inhalten und Vorgehensweisen

haben. Gerade im Hinblick auf die kognitive Aktivierung als stärker fachspezifischem Anteil der Unterrichtsqualität sind Unterschiede in der Ausprägung in den verschiedenen Inhaltsbereichen denkbar, da die Lehrkräfte unterschiedlich viel fachbezogenes Professionswissen in den einzelnen Bereichen haben. Aus diesem Grund wurden die Schülerinnen und Schüler in Deutschland vor den Fragen zum Sachunterricht auf die naturwissenschaftlichen Anteile des Sachunterrichts hingewiesen, indem ihnen konkrete naturwissenschaftliche Themen, zum Beispiel Tiere und Pflanzen, Wetter, elektrischer Strom und Licht und Schatten, genannt wurden. Dieses Vorgehen ist auch aus anderen Studien bekannt (Kleickmann et al., 2020). Gleichwohl sind die Ergebnisse mit einer gewissen Unsicherheit behaftet, welche Vorstellung die Schülerinnen und Schüler für ihre Einschätzungen zugrunde gelegt haben.

6.3.1 Zur wahrgenommenen Qualität des Mathematikunterrichts aus Sicht der Schülerinnen und Schüler

Anhand Abbildung 6.1 wird ersichtlich, dass die Hälfte der befragten Lernenden in Deutschland (50.1 %) die Klassenführung im Mathematikunterricht als hoch ausgeprägt wahrnimmt. Je etwa ein Viertel empfindet die Klassenführung als wenig effizient (26.5 %) beziehungsweise als mittelstark (23.5 %). Weiterführenden Analysen lässt sich entnehmen, dass sich die Angaben zur wahrgenommenen Klassenführung der Lernenden je nach Leistungsstand unterscheiden: Schülerinnen und Schüler, die sich durch überdurchschnittliche Leistungen in Mathematik (Kompetenzstufen IV und V) auszeichnen, erleben die Klassenführung häufiger als sehr effizient (59.4 %) im Vergleich zu den Lernenden, deren Leistungen durchschnittlich (Kompetenzstufe III) oder unterdurchschnittlich (Kompetenzstufen I und II) sind. Während immerhin noch 49.4 Prozent der Lernenden mit durchschnittlichen Testleistungen die Klassenführung im Unterricht als hoch effizient wahrnehmen, teilen nur 34.1 Prozent der Schülerinnen und Schüler mit unterdurchschnittlichen Leistungen diesen Eindruck. In Bezug auf die kognitive Aktivierung im Mathematikunterricht berichtet die Mehrheit der Schülerinnen und Schüler (60.6 %) von einem hoch kognitiv aktivierenden Unterricht, während gut ein Drittel der Lernenden (34.1 %) die kognitive Aktivierung als mittelstark ausgeprägt einschätzt. Dabei geben insbesondere eher leistungsschwächere Lernende an, im Unterricht stark kognitiv aktiviert zu werden (67.1 %). Dies deutet darauf hin, dass die kognitive

Abbildung 6.1: Wahrgenommene Unterrichtsqualität im Mathematikunterricht (Anteile der Schülerinnen und Schüler in Prozent)

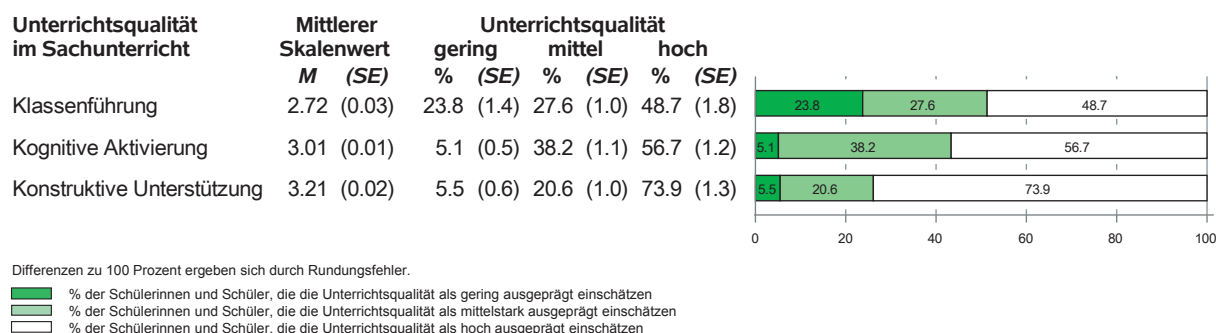


Aktivierung von diesen Schülerinnen und Schülern möglicherweise eher als Herausforderung wahrgenommen wird. Für Lernende mit mindestens durchschnittlichen Leistungen fallen die Angaben der als hoch eingestuften kognitiven Aktivierung im Mathematikunterricht etwas geringer aus (Kompetenzstufe III: 60.4%; Kompetenzstufen IV und V: 57.3%). Hinsichtlich der konstruktiven Unterstützung im Unterricht geben die Lernenden mehrheitlich an, diese als positiv zu erleben (76.6%). Während die kognitive Aktivierung im Unterricht insbesondere von leistungsschwächeren Schülerinnen und Schülern als hoch ausgeprägt wahrgenommen wird, zeigt sich für die Einschätzung der konstruktiven Unterstützung ein anderes Bild: Im Vergleich zu leistungsschwächeren Lernenden schätzen vor allem leistungsstärkere Lernende die konstruktive Unterstützung im Mathematikunterricht als stärker gegeben ein (83.6%). Je ein Großteil der Lernenden mit durchschnittlichen Leistungen (73.5%) und der Lernenden mit unterdurchschnittlichen Leistungen (69.2%) erlebt die konstruktive Unterstützung im Unterricht ebenfalls als äußerst positiv.

6.3.2 Zur wahrgenommenen Qualität des naturwissenschaftsbezogenen Sachunterrichts aus Sicht der Schülerinnen und Schüler

Abbildung 6.2 illustriert, wie Lernende die einzelnen Basisdimensionen der Unterrichtsqualität im naturwissenschaftsbezogenen Sachunterricht bewerten. Für die Klassenführung besteht ein ähnliches Muster wie für den Mathematikunterricht. Sie wird von rund der Hälfte der Viertklässlerinnen und Viertklässler (48.7%) als hoch ausgeprägt eingestuft. Jedoch berichtet auch jeweils gut ein Viertel der Lernenden von einer wenig effizienten (23.8%) beziehungsweise von einer mittelhoch ausgeprägten (27.6%) Klassenführung. Im Einklang mit der Beurteilung der Klassenführung in Mathematik lässt sich weiterführenden Analysen entnehmen, dass auch im Sachunterricht vor allem leistungsstärkere Schülerinnen und Schüler die Klassenführung als hoch effizient und strukturiert wahrnehmen (56.7%). Die Lernenden mit durchschnittlichen oder unterdurchschnittlichen Leistungen geben vergleichsweise etwas seltener an, eine sehr effiziente Klassenführung im Sachunterricht zu erleben (Kompetenzstufe III: 48.2%; Kompetenzstufen I und II: 36.3%). Hinsichtlich der kognitiven Aktivierung berichten 56.7 Prozent der Schülerinnen und Schüler, sich im Unterricht stark kognitiv aktiviert zu fühlen. Mehr als ein Drittel der Lernenden bewertet die kognitive Aktivierung im Unterricht als mittelstark ausgeprägt (38.2%). Zudem

Abbildung 6.2: Wahrgenommene Unterrichtsqualität im naturwissenschaftsbezogenen Sachunterricht (Anteile der Schülerinnen und Schüler in Prozent)



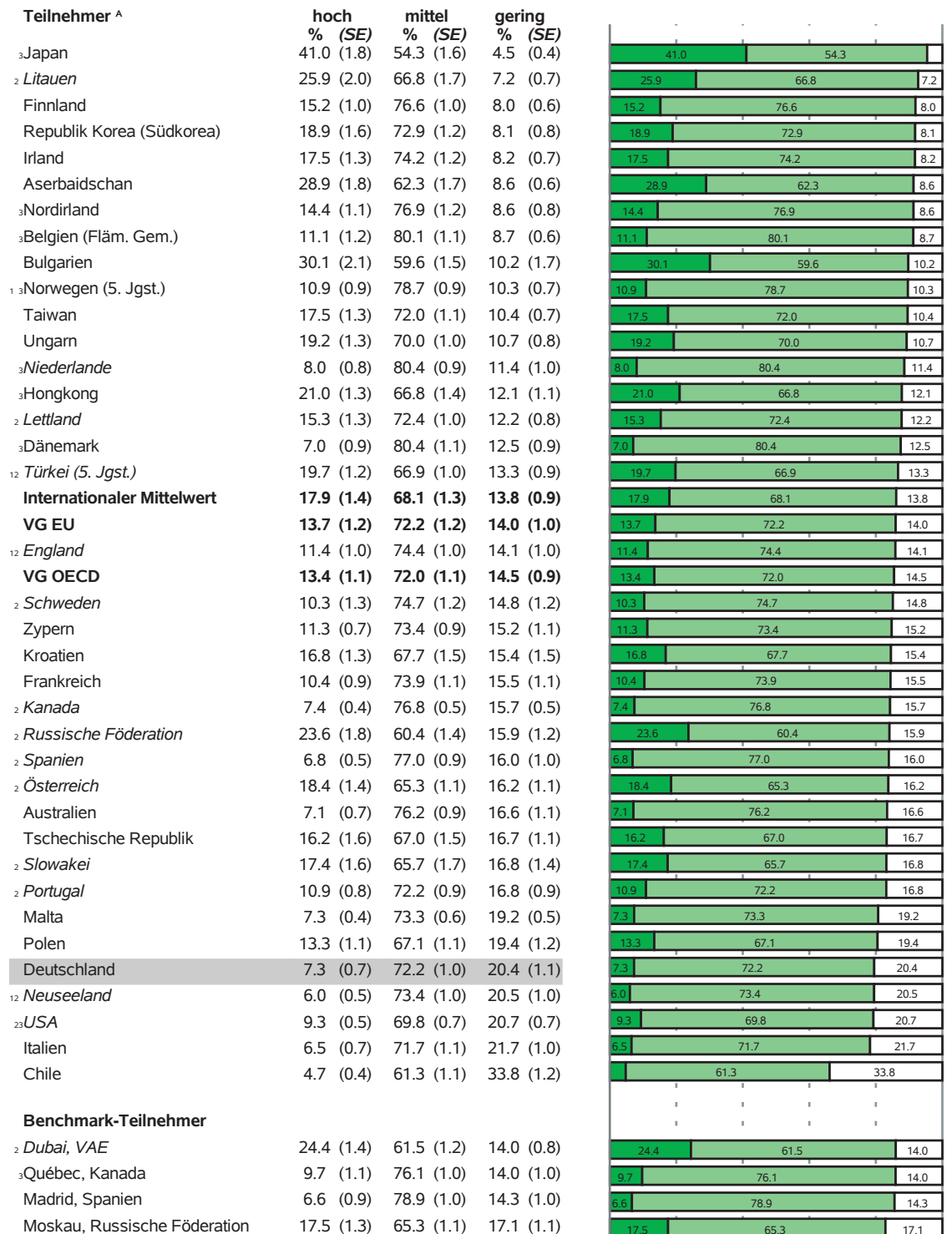
deuten sich ähnlich wie im Mathematikunterricht kleinere Unterschiede nach Leistungsniveau der Schülerinnen und Schüler an: Im Vergleich zu den leistungstärkeren Lernenden nehmen etwas mehr leistungsschwächere Lernende den Sachunterricht in einem höheren Maß als kognitiv fordernd wahr (62.6%). Allerdings berichtet auch jeweils mehr als die Hälfte der Lernenden mit durchschnittlichen Leistungen (57.1%) und überdurchschnittlichen Leistungen (52.8%) von einer hohen kognitiven Aktivierung im Unterricht. In Übereinstimmung mit den Ergebnissen für die konstruktive Unterstützung im Mathematikunterricht wird ebenfalls der naturwissenschaftsbezogene Sachunterricht von Viertklässlerinnen und Viertklässlern mehrheitlich (73.9%) als positiv empfunden. Auch hier zeigt sich, dass der Anteil an Lernenden, die die konstruktive Unterstützung im Unterricht als hoch ausgeprägt einschätzen, bei den leistungstärkeren (78.1%) im Vergleich zu leistungsschwächeren Lernenden (68.1%) größer ausfällt. Schülerinnen und Schüler, die durchschnittliche Leistungen erbringen, schätzen die konstruktive Unterstützung im Sachunterricht im Vergleich zu leistungstärkeren Lernenden seltener als hoch ausgeprägt ein (73.1%).

6.3.3 Wahrgenommene Unterrichtsqualität im Fach Mathematik aus Sicht der Schülerinnen und Schüler im internationalen Vergleich

International wurden bezogen auf die Unterrichtsqualitätsdimension Klassenführung nur wahrgenommene Störungen im Fach Mathematik anhand einer Skala gemessen. Störungen im Unterricht und der Umgang damit sind ein wichtiges Merkmal der Dimension Klassenführung. Die sechs Aussagen erfassen die Häufigkeit störenden Verhaltens während der Unterrichtszeit in Mathematik (vierstufige Skala von 1 = *in jeder oder fast jeder Stunde* bis 4 = *nie*).³ Für die Analysen wurden die Werte rekodiert. Die Antworten der Schülerinnen und Schüler wurden in die Kategorien hoch, mittel und gering gruppiert (siehe Abbildung 6.3). Die Unterrichtsqualität ist als hoch einzustufen, wenn Störungen in keiner oder nur in wenigen Stunden auftreten. Entsprechend ist die Klassenführung als mittel und gering einzustufen, wenn Störungen in manchen Stunden beziehungsweise (fast) allen Stunden vorkommen. Ein Beispielitem lautet „*Mein Lehrer muss uns immer wieder sagen, dass wir uns an die Klassenregeln halten sollen.*“.⁴ Der internationale Vergleich zeigt, dass es in der Wahrnehmung der Häufigkeit von Störungen Unterschiede zwischen den Staaten gibt (siehe Abbildung 6.3). In Deutschland geben rund 72 Prozent der Schülerinnen und Schüler an, dass es in manchen Stunden zu störendem Verhalten im Mathematikunterricht kommt, sodass die Lernzeit gemindert ist. Somit erleben die Lernenden eine mittelhoch ausgeprägte Klassenführung. Ein Fünftel (20.4%) erlebt eine wenig effiziente, also geringe Klassenführung und berichtet somit von Störungen in jeder oder fast jeder Stunde. Ähnliche Angaben wie die Schülerinnen und Schüler aus Deutschland machen Viertklässlerinnen und Viertklässler zum Beispiel in Malta und Neuseeland. Im direkten Vergleich dazu nimmt in Japan lediglich etwas mehr als die Hälfte (54.3%) der Lernenden die Klassenführung als mittelhoch ausgeprägt wahr und erleben so-

3 International war die Antwortreihenfolge umgekehrt (*nie* bis *in jeder oder fast jeder Stunde*).

4 Die internationalen Items unterscheiden sich in Umfang und Formulierung von den nationalen Items zur Klassenführung, die in TIMSS 2015 sowie in TIMSS 2019 für alle Analysen herangezogen wurden.

Abbildung 6.3: Wahrgenommene Störungen im Mathematikunterricht nach hoher, mittlerer und geringer Ausprägung (Anteile der Schülerinnen und Schüler in Prozent)

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

■ % der Schülerinnen und Schüler, die die Klassenführung als hoch ausgeprägt einschätzen
 ■ % der Schülerinnen und Schüler, die die Klassenführung als mittel ausgeprägt einschätzen
 ■ % der Schülerinnen und Schüler, die die Klassenführung als gering ausgeprägt einschätzen

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

¹ = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.² = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.³ = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.^A = In Singapur wurde die Klassenführung nicht erhoben. Somit ist Singapur nicht Teil dieser Darstellung und des berichteten internationalen Mittelwertes.

mit Störungen im Mathematikunterricht in manchen Stunden. Mehr als ein Drittel der Viertklässlerinnen und Viertklässler in Japan erlebt hingegen eine effiziente, also hoch ausgeprägte Klassenführung, bei welcher Störungen in keiner oder nur in wenigen Stunden auftreten (41.0%). Ein fast gegensätzliches Bild zeichnet sich für Chile ab. Gut rund ein Drittel der Schülerinnen und Schüler in Chile berichten von einer gering ausgeprägten Klassenführung, bei welcher Unterrichtsstörungen in jeder oder fast jeder Stunde (33.8%) vorkommen.

Vergleicht man die Angaben der Schülerinnen und Schüler aus Deutschland mit denen aller Teilnehmerstaaten, denen der OECD-Staaten und denen der EU-Staaten zeigt sich folgendes Bild: Die Mehrheit der Schülerinnen und Schüler nimmt die Unterrichtsqualität als mittel ausgeprägt wahr, sodass Unterrichtsstörungen in manchen Stunden auftreten (zwischen 68.1% und 72.2%). In Deutschland ist die Gruppe der Schülerinnen und Schüler, die die Klassenführung als gering wahrnimmt und entsprechend Unterrichtsstörungen in (fast) jeder Stunde erlebt, im Vergleich zu den Vergleichsgruppen (Internationaler Mittelwert, VG_{OECD} und VG_{EU}) um circa 6 Prozent größer. Die Gruppe, die in Deutschland die Klassenführung als hoch einschätzt beziehungsweise von Störungen in keiner oder nur in wenigen Stunden berichtet, ist mit 7.3 Prozent kleiner als in den genannten Vergleichsgruppen (internationaler Mittelwert: 17.9%; OECD-Staaten: 13.4%; EU-Staaten: 13.7%). Diese Unterschiede deuten in eine eher ungünstige Richtung. Insgesamt zeigt sich für Deutschland entsprechend der Häufigkeitsverteilung eine stark ausgeprägte Mitte, die auch international gegeben ist, mit einer vergleichsweise kleinen Gruppe, die keine oder nur sehr wenige Störungen im Unterricht berichtet. Die internationalen Ergebnisse zu Störungen als eine Facette von Klassenführung lassen sich unter anderem aufgrund der Nutzung verschiedener Variablen, der Positionierung im Fragebogen in Verbindung mit anderen Anteilen fehlender Werte und einer nicht identischen Skalenbildung nicht direkt mit den Befunden zu der nationalen Skala zur wahrgenommenen Klassenführung vergleichen.

6.3.4 Unterrichtsqualität im Vergleich zu TIMSS 2015

Für den nationalen Vergleich der Basisdimensionen von Unterrichtsqualität zwischen TIMSS 2015 und TIMSS 2019 zeichnet sich bei Betrachtung der Mittelwerte und Standardfehler der Unterrichtsqualitätsskalen für alle eingeschätzten Dimensionen der Unterrichtsqualität in beiden Fächern, mit Ausnahme der wahrgenommenen Klassenführung in Mathematik, eine bedeutende Veränderung ab (für TIMSS-2015-Werte siehe Lepper, Stang, Rieser & McElvany, 2020).⁵ Nachdem bei TIMSS 2015 im Durchschnitt positive Einschätzungen der Unterrichtsqualität durch die Schülerinnen und Schüler dokumentiert wurden, fallen diese Einschätzungen bei TIMSS 2019 im Mittel niedriger aus (für TIMSS-2019-Werte siehe Mittelwerte und Standardfehler in Abbildung 6.1 und 6.2). Dennoch beurteilen die Lernenden die Unterrichtsqualität mit Blick auf die drei Basisdimensionen im Jahr 2019 weiterhin im Mittel positiv.

⁵ Ein direkter Vergleich mit den Mittelwerten und Standardfehlern, die im Berichtsband von TIMSS 2015 vorgestellt wurden, ist aufgrund einer unterschiedlichen methodischen Herangehensweise nicht möglich (Rieser et al., 2016).

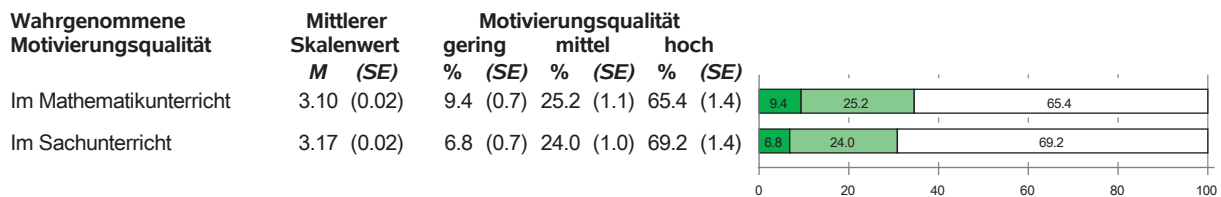
6.4 Einschätzung weiterer Gestaltungsaspekte des Unterrichts

Über die drei Basisdimensionen der Unterrichtsqualität hinaus werden zudem weitere Unterrichtsmerkmale wie die erlebte Motivierungsqualität, fachbezogene Gestaltungsmerkmale des Mathematik- beziehungsweise Sachunterrichts, adaptives Unterrichten oder Differenzierungsmöglichkeiten für leistungsstärkere und leistungsschwächere Lernende als relevant für Lern- und Entwicklungsprozesse von Schülerinnen und Schülern erachtet (Corno, 2008; Hertel, 2014; Klieme & Warwas, 2011; Rakoczy, 2008; Tomlinson, 2014). Aufgrund der Bedeutung der weiteren Gestaltungsmerkmale werden diese nachfolgend genauer betrachtet.

6.4.1 Motivierungsqualität aus Sicht der Schülerinnen und Schüler

Die Motivierungsqualität des Unterrichts gilt als relevantes Merkmal der Unterrichtsqualität (Helmke & Schrader, 2010). Sie fokussiert auf die Förderung der Lern- und Leistungsbereitschaft der Schülerinnen und Schüler hinsichtlich unterrichtlichen Lernens und bildet ab, inwiefern das Interesse der Lernenden im Unterricht geweckt wird und deren Freude und Begeisterung an Lerninhalten gestärkt werden (Kunter & Voss, 2011; Rakoczy, 2008). Dabei spielen insbesondere die Auswahl, die Vermittlung und die Gestaltung der Unterrichtsinhalte eine wichtige Rolle. Konzeptuell kann die Motivierungsqualität als Facette der konstruktiven Unterstützung verstanden werden (Ditton, 2009; Helmke & Schrader, 2010). Der Motivierungsqualität wird Bedeutsamkeit für Lernprozesse beigemessen, da sie positiv mit zentralen motivationalen Merkmalen, beispielsweise der intrinsischen Motivation oder dem Interesse der Lernenden am Unterrichtsfach, verknüpft ist (Kleickmann et al., 2020; Lazarides & Raufelder, 2017; Lepper, Stang & McElvany, 2020; Rakoczy, Klieme & Pauli, 2008; Reeve & Jang, 2006). Hingegen wurde das Verhältnis zwischen Motivierungsqualität und schulischen Leistungen bislang wenig beforscht (z. B. Kleickmann et al., 2020; Lepper et al., 2020). Um herauszufinden, wie gut es Lehrkräften gelingt, den Mathematik- beziehungsweise Sachunterricht motivierend zu gestalten und Schülerinnen und Schüler für die Inhalte des Unterrichts zu interessieren, wurde die wahrgenommene Motivierungsqualität in TIMSS 2019 mittels der Angaben von Lernenden erfasst.

Die wahrgenommene Motivierungsqualität wurde mithilfe von sechs Items erfragt. Zum Beispiel beantworteten Lernende anhand einer vierstufigen Likertskala (1 = *stimme völlig zu* bis 4 = *stimme überhaupt nicht zu*), inwiefern die Lehrkraft im Mathematik- beziehungsweise Sachunterricht interessante Inhalte auswählt. Die Items wurden rekodiert, sodass ein höherer Wert für eine höhere Ausprägung steht. Die Reliabilität der Motivierungsqualität in beiden Domänen ist als gut zu bezeichnen. Die Abbildung 6.4 zeigt, dass die Lernenden den Mathematikunterricht mehrheitlich (65.4%) als sehr motivierend wahrnehmen. Auch für den naturwissenschaftsbezogenen Sachunterricht geben 69.2 Prozent der Lernenden an, von der Lehrkraft stark motiviert zu werden. In beiden Fächern fällt der Anteil an Schülerinnen und Schülern, die den jeweiligen Unterricht als wenig motivierend erleben, entsprechend niedrig aus (Mathematik: 9.4%; Sachunterricht: 6.8%). Basierend auf weiterführenden Analysen kann festgehalten werden, dass sich für die erlebte Motivierungsqualität in Abhängigkeit des Leistungsniveaus der Lernenden für den Mathematik- beziehungsweise Sachunterricht lediglich minimale Unterschiede erkennen las-

Abbildung 6.4: Wahrgenommene Motivierungsqualität im Mathematik- und naturwissenschaftsbezogenen Sachunterricht (Anteile der Schülerinnen und Schüler in Prozent)

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

■ % der Schülerinnen und Schüler, die die Motivierungsqualität als gering ausgeprägt einschätzen
 ■ % der Schülerinnen und Schüler, die die Motivierungsqualität als mittelstark ausgeprägt einschätzen
 ■ % der Schülerinnen und Schüler, die die Motivierungsqualität als hoch ausgeprägt einschätzen

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

sen: Insgesamt erlebt die Mehrheit der Viertklässlerinnen und Viertklässler, unabhängig vom Kompetenzniveau, sowohl den Mathematik- als auch den Sachunterricht an ihren Grundschulen als hoch motivierend (Mathematik: 63.4–67.6%; Sachunterricht: 66.3–70.3%). Lediglich jeweils ein geringer Prozentsatz an Lernenden aller Kompetenzniveaus berichtet von einer als niedrig wahrgenommenen Motivierungsqualität (Mathematik: 7.9–10.6%; Sachunterricht: 5.8–7.2%).

6.4.2 Zusammenhang von wahrgenommener Unterrichtsqualität mit Leistung und Motivation

Die drei Basisdimensionen Klassenführung, kognitive Aktivierung und konstruktive Unterstützung sowie die Facette erlebte Motivierungsqualität prägen Lernprozesse von Schülerinnen und Schülern auf bedeutsame Art und Weise (Fauth et al., 2014a, 2014b; Kunter et al., 2013; Pianta & Hamre, 2009; Praetorius et al., 2018; Rakoczy, 2008). Basierend auf empirischen Befunden sowie einem heuristischen Angebots-Nutzungsmodell, das Zusammenhänge zwischen den drei Basisdimensionen und Schülerinnen- und Schülerleistungen beziehungsweise -motivation abbildet, lässt sich annehmen, dass die Dimensionen der Unterrichtsqualität unterschiedlich wichtig für kognitive und motivationale Lernergebnisse sind (Baumert et al., 2010; Fauth et al., 2014a; Klieme & Rakoczy, 2008; Kunter et al., 2013; Ryan & Deci, 2000). Während die Klassenführung zugleich als bedeutsam für leistungsbezogene und motivationale Faktoren erachtet wird, wird der kognitiven Aktivierung insbesondere Relevanz für die Leistungen von Lernenden und der konstruktiven Unterstützung vorwiegend Bedeutsamkeit für die Motivation von Lernenden beigemessen (Klieme & Rakoczy, 2008).

Im Rahmen von TIMSS 2019 wurde untersucht, ob die von Lernenden wahrgenommene Unterrichtsqualität prädiktiv für die Leistungen und intrinsische Motivation der Lernenden im Mathematik- und Sachunterricht ist. Die Zielkriterien Leistung und intrinsische Motivation in den Fächern Mathematik beziehungsweise Sachunterricht wurden mithilfe von reliablen Instrumenten erhoben (das Skalenhandbuch zu TIMSS 2019 befindet sich zurzeit in Vorbereitung). Aufgrund der vorliegenden geschachtelten Datenstruktur, Schülerinnen und Schüler in Klassen, sowie der Höhe der Intraklassenkorrelationen der Unterrichtsqualitätsdimensionen (ICC; Hox, 2013) wurde für die Vorhersage der Leistung eine mehrebenenanalytische Herangehensweise gewählt. Die Wahrnehmung

der Unterrichtsqualität wurde auf Klassen- und Schülerebene operationalisiert. Auf Schülerebene wurde zusätzlich für relevante Individualmerkmale (Geschlecht der Lernenden, Migrationshintergrund, sozioökonomischer Status, kognitive Fähigkeiten) kontrolliert. Aufgrund von Multikollinearitätsproblemen (Cohen, 1992) der Konstrukte Klassenführung und kognitive Aktivierung, wurde die Klassenführung in den Analysen nicht berücksichtigt. Zur Vorhersage der Schülerinnen- und Schülermotivation konnte der mehrerebenenanalytische Ansatz nicht verfolgt werden, da eine zu geringe Varianz auf Ebene 2 für die intrinsische Motivation vorlag. Daher wurden zwei einzelne lineare Regressionen auf Individualebene berechnet.

Zum Zusammenhang von wahrgenommener Unterrichtsqualität mit Leistung

In Tabelle 6.1 sind die Ergebnisse für die Zusammenhänge zwischen verschiedenen Facetten der Unterrichtsqualität und der Leistung der Lernenden für das jeweilige Unterrichtsfach dargestellt. Für den Mathematikunterricht zeigt sich auf Schülerebene im Modell 1, ohne Aufnahme der Kontrollvariablen, dass die Mathematikleistung statistisch bedeutsam mit der kognitiven Aktivierung und der konstruktiven Unterstützung zusammenhängt. Unter Kontrolle der mög-

Tabelle 6.1: Regression von Leistung in Mathematik und Naturwissenschaften auf verschiedene Basisdimensionen der Unterrichtsqualität, erlebte Motivierungsqualität und individuelle Merkmale von Lernenden (Mehrebenenmodell)

	Mathematik				Naturwissenschaften			
	ICC	Modell 1 b (SE)	Modell 2 b (SE)		ICC	Modell 1 b (SE)	Modell 2 b (SE)	
Konstante		527.3 (7.8)**	-482.5 (9.8)**			522.6 (7.1)**	-634.4 (17.2)**	
Klassenebene								
Kognitive Aktivierung		11.4 (20.7) ^{ns}	-2.3 (20.1) ^{ns}			38.6 (23.6) ^{ns}	4.2 (18.9) ^{ns}	
Konstruktive Unterstützung		-13.0 (20.3) ^{ns}	4.1 (17.2) ^{ns}			-23.7 (12.5) ^{ns}	0.5 (13.6) ^{ns}	
Erlebte Motivierungsqualität		4.3 (17.2) ^{ns}	0.8 (16.3) ^{ns}			-2.5 (17.7) ^{ns}	13.3 (19.2) ^{ns}	
Schülerebene								
Kognitive Aktivierung		-16.6 (5.9)**	-10.6 (6.4) ^{ns}			-18.5 (8.4)*	-7.3 (8.3) ^{ns}	
Konstruktive Unterstützung		28.5 (5.5)**	14.7 (6.4)*			24.6 (7.0)**	11.6 (7.3) ^{ns}	
Erlebte Motivierungsqualität		-7.6 (5.4) ^{ns}	1.0 (5.5) ^{ns}			-3.2 (6.5) ^{ns}	0.3 (7.1) ^{ns}	
Geschlecht ^A			18.0 (6.0)**				11.0 (6.5) ^{ns}	
Migrationshintergrund (ein Elternteil) ^B			4.9 (12.0) ^{ns}				19.8 (14.4) ^{ns}	
Migrationshintergrund (kein Elternteil) ^C			-0.0 (8.6) ^{ns}				30.8 (11.0)**	
Sozioökonomischer Status ^D			0.6 (0.1)**				0.6 (0.1)**	
Kognitive Fähigkeiten			9.7 (1.1)**				11.0 (1.2)**	
Erklärte Varianzanteile								
Zwischen den Klassen	25.7	1.0	1.0		21.3	4.6	4.6	
Innerhalb der Klassen	74.2	4.3	37.7		78.6	1.0	37.0	
Gesamt		3.5	28.3			1.7	30.0	

ICC = Intraklassenkorellation

b = Regressionsgewicht (unstandardisiert)

Signifikanzniveau: ns = nicht signifikant; * = signifikant ($p \leq .05$); ** = signifikant ($p \leq .01$)

A = Geschlecht (0 = Mädchen; 1 = Jungen)

B = Migrationshintergrund nach Geburtsland der Eltern (0 = beide Elternteile im Ausland geboren; 1 = ein Elternteil im Ausland geboren; 0 = kein Elternteil im Ausland geboren)

C = Migrationshintergrund nach Geburtsland der Eltern (0 = beide Elternteile im Ausland geboren; 0 = ein Elternteil im Ausland geboren; 1 = kein Elternteil im Ausland geboren)

D = Berufsstatus: Höchster ISEI (International Socio-Economic Index of Occupational Status) im Haushalt

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

lichen Einflussfaktoren erweist sich nur noch die konstruktive Unterstützung im Unterricht als prädiktiv für die Mathematikleistung der Lernenden (siehe Modell 2). Das bedeutet, dass Schülerinnen und Schüler, die die konstruktive Unterstützung des Unterrichts höher einstufen, im Durchschnitt bessere Mathematikleistungen erzielen. Ein ähnliches Bild wie für die Mathematikleistung zeichnet sich für den Sachunterricht ab. Auf Schülerebene hängen die Facetten kognitive Aktivierung und konstruktive Unterstützung statistisch bedeutsam mit der Leistung der Lernenden zusammen, wenn nicht um weitere Variablen kontrolliert wird (siehe Modell 1). Bei Aufnahme der Kontrollvariablen existieren keine statistisch bedeutsamen Zusammenhänge mehr zwischen der Wahrnehmung der Unterrichtsmerkmale und der Leistung (siehe Modell 2). Auf Klassenebene lassen sich weder für den Mathematik- noch für den Sachunterricht bedeutsame Zusammenhänge zwischen den Facetten der Unterrichtsqualität und den mittleren Fachleistungen finden. Insgesamt werden, unter Kontrolle weiterer Einflussvariablen, lediglich Leistungsunterschiede im Mathematikunterricht innerhalb der Klassen anhand der individuellen Wahrnehmung einer einzelnen Unterrichtsqualitätsfacette von Lernenden erklärt. Deutlich wird, dass individuelle Merkmale sowie Voraussetzungen der Lernenden Relevanz für die Vorhersage der Schülerinnen- und Schülerleistung in beiden Fächern besitzen (siehe Kapitel 8, 9, 10 und 12 in diesem Band) und somit die Vorhersagekraft der erlebten Unterrichtsqualitätsdimensionen überlagern.

Zum Zusammenhang von wahrgenommener Unterrichtsqualität mit Motivation

Bei den beiden durchgeführten Regressionen, eine zur Vorhersage der intrinsischen Motivation in Mathematik, die andere zur Vorhersage der intrinsischen Motivation im naturwissenschaftsbezogenen Sachunterricht, wurden neben der kognitiven Aktivierung, der konstruktiven Unterstützung und der erlebten Motivierungsqualität auch direkt die Kontrollvariablen aufgenommen, die auch zur Vorhersage der Leistung betrachtet wurden. Hinsichtlich des Zusammenhangs zwischen den verschiedenen Facetten der Unterrichtsqualität und der intrinsischen Motivation zeigt sich, dass in beiden Fächern die erlebte konstruktive Unterstützung sowie die erlebte Motivierungsqualität in einer signifikant positiven Relation zu der intrinsischen Motivation stehen (Mathematik: $b_{\text{konstruktiveUnterstützung}} = .28$, $SE = 0.0$, $b_{\text{erlebteMotivierung}} = .44$, $SE = 0.0$; Sachunterricht: $b_{\text{konstruktiveUnterstützung}} = .16$, $SE = 0.0$, $b_{\text{erlebteMotivierung}} = .51$, $SE = 0.0$). Dies bedeutet, dass Kinder, die sich konstruktiv unterstützt beziehungsweise sich durch die Lehrkraft motiviert fühlen, im Mittel auch von einer höher ausgeprägten intrinsischen Motivation im jeweiligen Fach berichten. In Mathematik steht zudem die intrinsische Motivation in einem signifikant negativen Zusammenhang mit der wahrgenommenen kognitiven Aktivierung ($b = -.12$, $SE = 0.0$). Lernende, die sich in Mathematik stärker kognitiv aktiviert und damit möglicherweise stärker herausgefordert fühlen, verspüren im Durchschnitt eine geringer ausgeprägte intrinsische Motivation. Darüber hinaus zeigt sich in Bezug auf das Geschlecht der Lernenden, dass Jungen in der vierten Klassenstufe für Mathematik deutlich stärker intrinsisch motiviert sind als Mädchen ($b = .27$, $SE = 0.0$). Dieser Geschlechterunterschied ist für die Motivation im naturwissenschaftsbezogenen Sachunterricht hingegen nicht zu verzeichnen. Für diesen ergibt sich, dass kognitive Grundfähigkeiten minimal bedeutsam für die Motivation sind ($b = .02$, $SE = 0.0$). Weder Migrationshintergrund noch sozioökonomischer Status stehen in einem statistisch bedeutsamen Zusammenhang mit der intrinsischen

Motivation in den beiden Domänen. Die aufgeklärte Varianz beträgt 28 Prozent für die Motivation im Mathematikunterricht und 38 Prozent für den naturwissenschaftsbezogenen Sachunterricht.

6.4.3 Fachspezifische Unterrichtsqualität aus der Sicht von Lehrkräften

Über die eher generischen Basisdimensionen der Unterrichtsqualität hinaus sind fachspezifische Gestaltungsmerkmale des Unterrichts von Interesse (Helmke & Schrader, 2010; Praetorius et al., 2020), da sie als wichtig für die Herausbildung eines vertieften Verständnisses der Mathematik und naturwissenschaftlicher Phänomene erachtet werden. Ein solches Verständnis bildet eine wichtige Grundlage für die weitere schulische und die sich anschließende berufliche Laufbahn (Mullis & Martin, 2017). Im Fokus der Erforschung fachspezifischer Unterrichtsqualität stehen daher bei TIMSS 2019 Aussagen von Lehrkräften zur fachbezogenen Gestaltung des Unterrichts in den Fächern Mathematik und (naturwissenschaftsbezogener) Sachunterricht.

Konkret nimmt die fachspezifische Unterrichtsqualität Bezug darauf, inwiefern spezifische mathematische beziehungsweise naturwissenschaftliche Inhalte unter Berücksichtigung wichtiger Merkmale von Unterrichtsqualität, beispielsweise Strategien der kognitiven Aktivierung, wie das Anregen von Vergleichen, das Anwenden von Wissen in verschiedenen Kontexten oder die geeignete Sequenzierung von Inhalten, im jeweiligen Fachunterricht vermittelt werden (Mullis & Martin, 2017). Zusätzlich werden für den naturwissenschaftsbezogenen Sachunterricht in der TIMSS-Rahmenkonzeption fünf elementare wissenschaftliche Denk- und Arbeitsweisen, sogenannte *science practices*, aufgelistet (z. B. Beantwortung von Forschungsfragen; Mullis & Martin, 2017). Lernende sollen im Sachunterricht zentrale Praktiken wissenschaftlicher Untersuchungen sowie grundlegende Schritte eines Forschungskreislaufes erlernen, beginnend mit der Hypothesengenerierung bis hin zur Dateninterpretation (Mullis & Martin, 2017). Im Unterricht sind diese Denk- und Arbeitsweisen relevant, um beispielsweise bestehende Ideen der Lernenden zu naturwissenschaftlichen Phänomenen (Präkonzepte) durch Beobachtungen kritisch zu hinterfragen. Im Kontext der Erfassung von Unterrichtsqualität sind Angaben von Lehrkräften zum Einsatz naturwissenschaftlicher Denk- und Arbeitsweisen entsprechend von Interesse, da sie Auskunft darüber geben, inwiefern Lehrkräfte ihre Schülerinnen und Schüler im Sachunterricht zu wissenschaftlichem Denken und Arbeiten anregen (Barzel, Reinthoffer & Schrenk, 2012; Harlen, 2013; Jonen, Möller & Hardy, 2003). Nachfolgend wird berichtet, ob Lehrkräfte Methoden der fachspezifischen Unterrichtsqualität im Mathematik- beziehungsweise Sachunterricht einsetzen.

Zur Erfassung der fachspezifischen Unterrichtsqualität wurden die Items so formuliert, dass sie die verschiedenen kognitiven Anforderungsbereiche (*Reproduzieren, Anwenden, Problemlösen*) in unterschiedlichen Inhaltsbereichen in den Fächern Mathematik und Sachunterricht erfassen (Mathematische Inhaltsbereiche: *Arithmetik, Messen und Geometrie* sowie *Daten*; Naturwissenschaftliche Inhaltsbereiche: *Biologie, Physik/Chemie* und *Geografie*; Mullis & Martin, 2017). Für die Auswertungen der fachspezifischen Gestaltungsmerkmale des Mathematik- und Sachunterrichts wurden jeweils neun Items zu Skalen zusammengefasst (Beispielitem Mathematik: „*Ich gebe den Schülern Aufgaben, bei denen sie Datenquellen nutzen können, um Fragen des Alltags zu beantworten.*“; Beispielitem Sachunterricht: „*Ich lasse Schüler spezifische Vorgänge*

in Ökosystemen anhand von Beispielen erklären.“). Zur Erfassung der naturwissenschaftlichen Denk- und Arbeitsweisen im Sachunterricht wurde jeweils ein Item für jede der fünf zentralen Praktiken (*science practices*) entwickelt (Fragenstellen basierend auf Beobachtungen, Evidenzgenerierung, Arbeit mit Daten, Beantwortung von Forschungsfragen, Evidenzbasiertes Argumentieren; Mullis & Martin, 2017). Entsprechend wurden fünf Items eingesetzt (z. B. „*Ich zeige meinen Schülern, wie man wissenschaftliche Experimente korrekt durchführt, um Hypothesen aus dem Sachunterricht zu testen.“*). Auf einer vierstufigen Likertskala konnten Lehrkräfte den Aussagen zustimmen (1 = *trifft nicht zu* bis 4 = *trifft zu*). Die Reliabilitäten der eingesetzten Skalen lagen in einem guten Bereich.

Für die fachspezifische Unterrichtsqualität im Fach Mathematik zeigt sich, dass mehr als die Hälfte (54.8%) der Schülerinnen und Schüler von Lehrkräften unterrichtet werden, die grundlegende fachspezifische Prinzipien im Unterricht insgesamt eher umsetzen. Weitere 40.3 Prozent werden von Lehrkräften unterrichtet, die die fachspezifischen Gestaltungsaspekte des Mathematikunterrichts eher weniger in ihren Unterricht integrieren. Ein ähnliches Befundmuster besteht für den Sachunterricht. Auch hier werden rund 57 Prozent der Schülerinnen und Schüler von Lehrkräften unterrichtet, die angeben, fachspezifische Merkmale der Unterrichtsqualität in ihrer Unterrichtsgestaltung eher zu berücksichtigen. Gleichzeitig zeigt sich für mehr als ein Drittel der Viertklässlerinnen und Viertklässler (37.6%), dass sie von Lehrkräften unterrichtet werden, bei denen die Umsetzung dieser Gestaltungsmerkmale im naturwissenschaftsbezogenen Sachunterricht eher wenig Berücksichtigung findet. Darüber hinaus wird in etwa die Hälfte der Schülerinnen und Schüler (53.7%) von Lehrkräften unterrichtet, die grundlegende naturwissenschaftliche Denk- und Arbeitsweisen wie Beobachten, Vermuten oder Experimentieren im Sachunterricht eher nutzen. Ungefähr ein Drittel der Schülerinnen und Schüler (33.2%) wird von Lehrkräften unterrichtet, die einschätzen, dass diese Arbeitsweisen in ihrem Unterricht eher nicht vorkommen. Zusammenfassend zeigt sich, dass etwas mehr als die Hälfte der Viertklässlerinnen und Viertklässler von Lehrkräften unterrichtet werden, die im Mathematik- und Sachunterricht zentrale fachspezifische Gestaltungsaspekte und im Sachunterricht zusätzlich naturwissenschaftliche Denk- und Arbeitsweisen berücksichtigen.

6.4.4 Differenzierung im Unterricht aus der Sicht von Lehrkräften

Neben den fachspezifischen Merkmalen der Unterrichtsqualität werden auch Konzepte eines individualisierten und differenzierten Unterrichts als wichtige Bezugsrahmen unterrichtlicher Lehr-Lernprozesse erachtet. Schulklassen sind von Heterogenität geprägt, die sich beispielsweise in Bezug auf die Leistungen von Schülerinnen und Schülern zeigt (Dumont, Neumann, Maaz & Trautwein, 2013; Heinzl, 2008). Für die Gestaltung des Unterrichts ist daher die Berücksichtigung unterschiedlicher individueller Lernvoraussetzungen innerhalb einer Klasse relevant, um bestmögliche Lerngelegenheiten zu schaffen (Corno, 2008; Klieme & Warwas, 2011). Angaben zur adaptiven Unterrichtsgestaltung geben über allgemeine Formen der leistungsbezogenen Differenzierung Auskunft, beispielsweise über die Variation der Lernmethoden oder der zu Verfügung stehenden Bearbeitungszeit (Brühwiler & Vogt, 2020; Hardy et al., 2011). Hingegen werden mit der Differenzierung für stärkere Lernende und Differenzierung für

schwächere Lernende konkrete, auf leistungsstärkere und -schwächere Lernende abgestimmte, Handlungsstrategien erfasst. Im Rahmen der leistungsbezogenen Differenzierung werden Lerninhalte und -angebote an die unterschiedlichen Leistungsniveaus der Lernenden angepasst, sodass verstärkt individualisierte Lerngelegenheiten geschaffen werden (Corno, 2008; Corno & Snow, 1986). Aufgrund der Bedeutung des adaptiven Unterrichts und konkreter leistungsbezogener Differenzierung stellt sich die Frage, wie diese aktuell im Mathematik- und Sachunterricht umgesetzt werden.

Adaptives Unterrichten in Mathematik wurde mit fünf Items (z.B. „*Ich differenziere in meinem Mathematikunterricht in dieser Klasse für unterschiedliche Schülergruppen nach mathematischer Fähigkeit die Aufgabeninhalte.*“) erfasst. Für den Sachunterricht wurden die Items äquivalent formuliert, jedoch um ein weiteres Item ergänzt. Lehrkräfte machten auf einer sechsstufigen Skala Angaben, in wie vielen Stunden sie den Unterricht adaptiv gestalten (1 = *in fast jeder Stunde* bis 6 = *nie*). Zur Messung der leistungsbezogenen Differenzierung für stärkere und schwächere Lernende wurden jeweils sowohl im Mathematik- als auch im Sachunterricht fünf Items eingesetzt (z.B. Differenzierung für Stärkere in Mathematik: „*Stärkere Schüler müssen im Vergleich zu schwächeren Schülern in meiner Klasse Aufgaben mit mehreren Rechenschritten lösen.*“; Differenzierung für Schwächere im Sachunterricht: „*Schwächere Schüler müssen im Vergleich zu stärkeren Schülern in meiner Klasse einfachere Versuche planen und durchführen.*“). Die Lehrkräfte beantworteten die Aussagen auf einer vierstufigen Likertskala (1 = *trifft überhaupt nicht zu* bis 4 = *trifft voll und ganz zu*). Die Reliabilitäten der Skalen zur Erfassung adaptiven Unterrichts und der leistungsbezogenen Differenzierung waren gut.

In Bezug auf adaptives Unterrichten lässt sich erkennen, dass etwa die Hälfte der Schülerinnen und Schüler (52.8%) von Lehrkräften unterrichtet wird, die in mindestens fünf bis acht von zehn Unterrichtsstunden adaptive Unterrichtsmethoden im Mathematikunterricht einsetzen. Gleichzeitig zeigt sich für 31.4 Prozent der Viertklässlerinnen und Viertklässler, dass sie von Lehrkräften unterrichtet werden, die angeben, dass sie etwas seltener (in drei oder vier von zehn Stunden) adaptiv unterrichten. Im Vergleich dazu werden im naturwissenschaftsbezogenen Sachunterricht nur 30.5 Prozent der Schülerinnen und Schüler von Lehrkräften unterrichtet, die mindestens fünf bis acht von zehn Stunden adaptiv gestalten. Der größte Anteil an Schülerinnen und Schülern (38.0%) wird von Lehrkräften unterrichtet, die im Durchschnitt drei oder vier von zehn Schulstunden adaptiv unterrichten. Hinsichtlich der leistungsbezogenen Differenzierung im Unterricht geht aus Abbildung 6.5 hervor, dass die Mehrheit der Schülerinnen und Schüler von Lehrkräften unterrichtet wird, die

Abbildung 6.5: Leistungsbezogene Differenzierung im Mathematikunterricht aus Sicht der Lehrkräfte

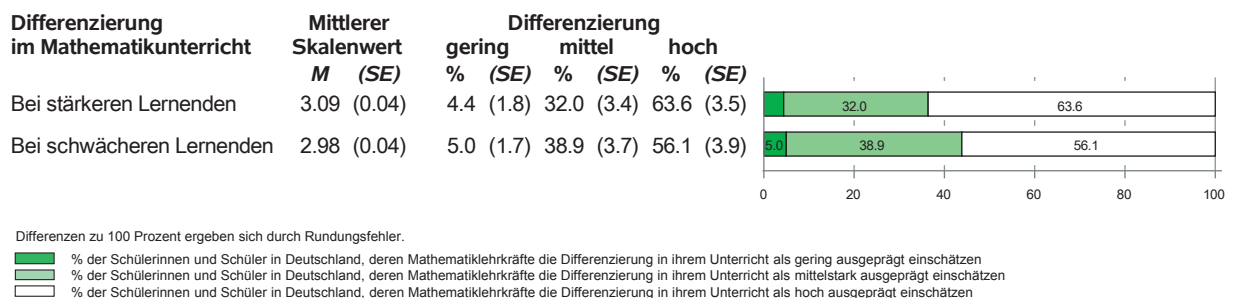
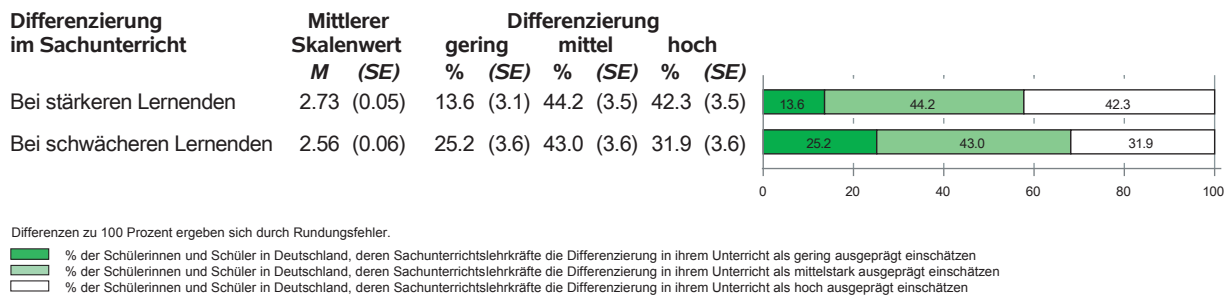


Abbildung 6.6: Leistungsbezogene Differenzierung im naturwissenschaftsbezogenen Sachunterricht aus Sicht der Lehrkräfte

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

häufig differenzierte Lernangebote im Mathematikunterricht sowohl für leistungstärkere (63.6%) als auch für leistungsschwächere Lernende (56.1%) schaffen. Für den Sachunterricht zeigt sich ein etwas anderes Bild. Im Vergleich zum Mathematikunterricht zeigt sich für den naturwissenschaftsbezogenen Sachunterricht, dass etwas weniger als die Hälfte der Lernenden (42.3%) von Sachunterrichtslehrkräften unterrichtet wird, die die Differenzierung in ihrem Unterricht für leistungstärkere Lernende als hoch ausgeprägt einschätzen. Geringer ist hingegen der Anteil an Lernenden (31.9%), deren Lehrkräfte die Differenzierung in ihrem naturwissenschaftsbezogenen Sachunterricht für leistungsschwächere Lernende als hoch ausgeprägt wahrnehmen. Im Vergleich zur leistungsbezogenen Differenzierung im Mathematikunterricht scheint die Differenzierung für Lernende unterschiedlicher Leistungsniveaus im Sachunterricht nicht nur weniger stark ausgeprägt, sondern auch weniger zwischen den Leistungsgruppen ausbalanciert zu sein (siehe Abbildung 6.6).

6.5 Zusammenfassung

Die Qualität von Unterricht, die sowohl generische als auch stärker fachspezifische Merkmale umfasst, ist von grundlegender Bedeutung für das erfolgreiche Lernen von Schülerinnen und Schülern (Einsiedler, 2017; Kleickmann et al., 2020; Kunter & Voss, 2011; Ohle & McElvany, 2016). Die Klassenführung, die konstruktive Unterstützung, zu der die emotionale und die strukturierende Unterstützung gehören, sowie die kognitive Aktivierung werden sowohl in dem Basisdimensionenmodell als auch in weiteren Konzeptionen als zentrale Elemente der Unterrichtsqualität beschrieben. Um Einblicke in die Qualität des Mathematik- und des naturwissenschaftsbezogenen Sachunterrichts zu erhalten, wurden in TIMSS 2019 die Schülerinnen und Schüler nach ihren Einschätzungen der drei Basisdimensionen der Unterrichtsqualität gefragt. Um darüber hinaus differenziertere Einschätzungen zur Unterrichtsqualität zu erhalten, wurde in TIMSS 2019 zusätzlich auf weitere relevante Unterrichtsmerkmale fokussiert. Anhand neu entwickelter Skalen wurde die Erfassung der Unterrichtsqualität nicht nur durch die schülerseitig wahrgenommene Motivierungsqualität, sondern auch durch Angaben von Lehrkräften zu weiteren zentralen und zum Teil fachspezifischen Gestaltungsmerkmalen des Mathematik- beziehungsweise naturwissenschaftsbezogenen Sachunterrichts erweitert.

Die Ergebnisse in TIMSS 2019 zeigen, dass die Lernenden die Klassenführung, die kognitive Aktivierung und die konstruktive Unterstützung über beide Unterrichtsfächer hinweg positiv einschätzen. Diese Befunde stehen in Einklang mit vorherigen Studienergebnissen, die ebenfalls von einer günstigen Einschätzung verschiedener Dimensionen berichteten (z.B. Fauth et al., 2014a; Kleickmann et al., 2020). Gleichzeitig deuten sich einige Unterschiede in der Wahrnehmung der Basisdimensionen in Abhängigkeit des Leistungsniveaus der Lernenden an (vgl. auch Rieser et al., 2016). Allerdings fallen diese in Abhängigkeit der betrachteten Dimension von Unterrichtsqualität prozentual gesehen verschieden groß aus. So schätzen leistungsstarke Schülerinnen und Schüler die Klassenführung und konstruktive Unterstützung etwas höher ein als die leistungsschwächeren Schülerinnen und Schüler, während sich bei der kognitiven Aktivierung tendenziell das umgekehrte Bild ergibt. Die Unterschiede sind aber größtenteils als eher klein zu bezeichnen. Perspektivisch sollten die Wahrnehmungsunterschiede sowie mögliche zugrunde liegende Faktoren daher differenzierter betrachtet werden (vgl. auch Igler, Ohle-Peters & McElvany, 2019).

Neben den drei zentralen Merkmalen von Unterrichtsqualität wird auch die Motivierungsqualität für beide Unterrichtsfächer im Mittel als hoch ausgeprägt eingestuft. Die überwiegend positive Einschätzung der wahrgenommenen Motivierungsqualität steht in Einklang mit bestehenden Befunden für Grundschullernende (z.B. Igler et al., 2019; Lepper et al., 2020). Hinsichtlich der erlebten Motivierungsqualität in Abhängigkeit des Leistungsniveaus zeigen sich in TIMSS 2019 lediglich sehr geringe prozentuale Unterschiede, sodass diese von der Mehrheit der Lernenden, unabhängig ihres Leistungsstandes, als positiv und hoch ausgeprägt empfunden wird. In Bezug auf die Zusammenhänge zwischen zentralen Unterrichtsqualitätsdimensionen und leistungsbezogenen Lernergebnissen lässt sich resümieren, dass unter Kontrolle relevanter weiterer Einflussvariablen, entgegen theoretisch sowie empirisch abgeleiteten Erwartungen, lediglich nur eine statistisch bedeutsame Relation zwischen erlebter konstruktiver Unterstützung und der Leistung in Mathematik zu finden ist. Hinsichtlich des Zusammenhangs zwischen verschiedenen Facetten der Unterrichtsqualität und der Schülerinnen- und Schülermotivation ist festzuhalten, dass die Befunde auf individueller Ebene der Lernenden weitestgehend in Einklang mit theoretischen Annahmen sowie empirischen Arbeiten, fokussierend auf unterschiedliche Domänen und Altersstufen, stehen (z.B. Kleickmann et al., 2020; Klieme & Rakoczy, 2008; Stang, Schwabe & McElvany, in Druck). Insgesamt bieten die Befunde einen wichtigen ersten Überblick und stellen eine Grundlage für weiterführende Analysen dar. Dabei wäre zum Beispiel die Betrachtung von Klassenkompositionseffekten von Interesse.

Hinsichtlich der wahrgenommenen Störungen im Mathematikunterricht, als einem Teilaspekt der Klassenführung in TIMSS, zeigt sich im internationalen Vergleich, dass die prozentuale Verteilung für Deutschland in den drei Gruppen, gering, mittel und hoch, geringfügig von der Verteilung der anderen Teilnehmerstaaten abweicht. Die Unterschiede ergeben sich aus einem größeren Anteil an Schülerinnen und Schülern in Deutschland, die Störungen als häufig vorkommend beschreiben, und einem kleineren Anteil an Lernenden, die Störungen als selten vorkommend wahrnehmen. Im Vergleich der Basisdimensionen der Unterrichtsqualität zwischen TIMSS 2015 und TIMSS 2019 lassen sich einige Veränderungen beobachten. Diese deuten insgesamt auf eine weniger positive Wahrnehmung der Schülerinnen und Schüler in TIMSS 2019

hin. Das Ergebnis lässt sich mit den in TIMSS 2019 leicht gesunkenen Werten der positiven Einstellungen gegenüber dem Mathematik- und dem naturwissenschaftsbezogenen Sachunterricht in Einklang bringen (siehe Kapitel 3 und 4 in diesem Band). Gleichwohl sind die mittleren Ausprägungen für kognitive Aktivierung, konstruktive Unterstützung und mit einigen Abstrichen auch für Klassenführung immer noch im positiven Bereich und decken sich damit mit Ergebnissen verschiedener Studien aus unterschiedlichen Jahren, die auf ähnliche Domänen und Altersstufen fokussierten (z.B. Fauth et al., 2014b; Kleickmann et al., 2020; Wagner et al., 2016).

Mit Blick auf die anhand neuer Skalen untersuchten verschiedenen fachspezifischen Gestaltungsaspekte im Unterricht lässt sich ein eher gemischtes Fazit ziehen. Während ein Großteil der Schülerinnen und Schüler in TIMSS 2019 von Lehrkräften unterrichtet wird, die bestätigen, fachdidaktisch bedeutsame Gestaltungsprinzipien sowie naturwissenschaftliche Denk- und Arbeitsweisen in ihrem Unterricht anzuwenden, lässt sich ungefähr ein Drittel der Schülerinnen und Schüler identifizieren, die von Lehrkräften unterrichtet werden, die dies nicht für die eigene Unterrichtsgestaltung bestätigen können. Etwas optimistischer sind die Befunde zur Differenzierung im Mathematikunterricht. Diesbezüglich zeigt sich, dass die Schülerinnen und Schüler von Lehrkräften unterrichtet werden, die häufig adaptive Unterrichtsformen nutzen und sowohl für leistungsstärkere als auch leistungsschwächere Lernende differenzierte Lernangebote anbieten, was in Anbetracht heterogener Klassenkontexte von Relevanz ist (Dumont, 2019; Gräsel, Decristan & König, 2017; Hertel, 2014; Parsons et al., 2018). Eine Erklärungsmöglichkeit für Unterschiede hinsichtlich der leistungsbezogenen Differenzierung im Mathematik- und Sachunterricht könnte beispielsweise in der Anzahl an Angeboten fachspezifischer Aus- und Fortbildungsmöglichkeiten für den Mathematik- und Sachunterricht liegen. Zudem könnte es einigen Lehrkräften leichter fallen, mithilfe von Aufgaben im Fach Mathematik zu differenzieren, als Lernende zum Beispiel naturwissenschaftliche Versuche auf unterschiedlichen Niveaus durchführen zu lassen.

Anhand der Befundmuster zur Gestaltung des Unterrichts in TIMSS 2019 lassen sich Forschungsbedarfe aufzeigen sowie erste mögliche Implikationen für die schulische Bildungspraxis und Lehramtsausbildung ableiten. In zukünftigen TIMSS-Zyklen wäre eine Fortführung der differenzierten Erfassung von Unterrichtsqualität aus Perspektive der Schülerinnen und Schüler und der Lehrkräfte wünschenswert, um ein umfassenderes Verständnis der Ausgestaltung und Wahrnehmung von Lehr-Lerngelegenheiten zu erhalten sowie die Entwicklung der Unterrichtsqualität im Trend verfolgen zu können. Die vorliegenden Daten und Ergebnisse bieten eine gute Grundlage für weiterführende Analysen, welche beispielsweise auf differenzielle Effekte der Unterrichtsqualität etwa für Kinder mit unterschiedlichen Ausgangsbedingungen fokussieren könnten. Anhand eines detaillierteren Einblicks hinsichtlich möglicher Unterschiede in der Wahrnehmung der Unterrichtsqualität zwischen verschiedenen Schülerinnen- und Schülergruppen ließen sich konkretere Handlungsansätze für die Praxis ableiten. Vor dem Hintergrund einer zunehmenden Heterogenität von Schülerinnen und Schülern, die sich durch gesellschaftliche Entwicklungen und Änderungen im Schulsystem ergibt (Autorengruppe Bildungsberichterstattung, 2020), erscheinen Unterrichtsqualitätsmerkmale wie Klassenführung und Differenzierungsansätze als zentrale Ansatzpunkte, um auch Schülerinnen und Schülern mit ungünstigeren Voraussetzungen eine erfolgreiche Teilnahme am Unterricht zu ermöglichen. Dies setzt voraus, dass angehende wie praktizierende Lehrkräfte ge-

nügend Lerngelegenheiten haben, sich mit generischen und fachspezifischen Unterrichtsqualitätsmerkmalen auseinanderzusetzen, um so den Unterricht langfristig weiterzuentwickeln.

Literatur

- Autorengruppe Bildungsberichterstattung. (Hrsg.). (2020). *Bildung in Deutschland 2020. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt*. Bielefeld: wbv. Verfügbar unter:
https://www.bildungsbericht.de/static_pdfs/bildungsbericht-2020.pdf
- Barzel, B., Reinhoffer, B. & Schrenk, M. (2012). Das Experimentieren im Unterricht. In W. Rieß, M. Wirtz, B. Barzel & A. Schulz (Hrsg.), *Experimentieren im mathematisch-naturwissenschaftlichen Unterricht. Schüler lernen wissenschaftlich denken und arbeiten* (S. 103–128). Münster: Waxmann.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- Brühwiler, C. & Vogt, F. (2020). Adaptive teaching competency. Effects on quality of instruction and learning outcomes. *Journal for Educational Research Online*, 12(1), 119–142.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, 43, 161–173. <https://doi.org/10.1080/00461520802178466>
- Corno, L. & Snow, R. E. (1986). Adapting teaching to individual differences in learners. In M. C. Wittrock (Hrsg.), *Third handbook of research on teaching* (S. 605–629). Washington, DC: American Educational Research Association.
- Decristan, J., Kunter, M., Fauth, B., Büttner, G., Hardy, I. & Hertel, S. (2016). What role does instructional quality play for elementary school children's science competence? A focus on students at risk. *Journal for Educational Research Online*, 8(1), 66–89.
- Dietrich, J., Dicke, A. L., Kracke, B. & Noack, P. (2015). Teacher support and its influence on students' intrinsic value and effort: Dimensional comparison effects across subjects. *Learning and Instruction*, 39, 45–54. <https://doi.org/10.1016/j.learninstruc.2015.05.007>
- Ditton, H. (2009). Unterrichtsqualität. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 235–243). Bad Heilbrunn: Klinkhardt.
- Dumont, H. (2019). Neuer Schlauch für alten Wein? Eine konzeptuelle Betrachtung von individueller Förderung im Unterricht. *Zeitschrift für Erziehungswissenschaft*, 22(2), 249–277. <https://doi.org/10.1007/s11618-018-0840-0>
- Dumont, H., Neumann, M., Maaz, K. & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen. Internationale und nationale Befunde. *Psychologie in Erziehung und Unterricht*, 60(3), 163–183. <https://doi.org/10.2378/peu2013.art14d>
- Einsiedler, W. (2017). Von Erziehungs- und Unterrichtsstilen zur Unterrichtsqualität. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge* (S. 267–287). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-15083-9_12
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014a). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9. <https://doi.org/10.1016/j.learninstruc.2013.07.001>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014b). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, 28(3), 127–137. <https://doi.org/10.1024/1010-0652/a000129>
- Gräsel, C., Decristan, J. & König, J. (2017). Adaptiver Umgang mit Heterogenität im Unterricht. *Unterrichtswissenschaft*, 45(4), 195–206.
- Hardy, I., Hertel, S., Kunter, M., Klieme, E., Warwas, J., Büttner, G. & Lühken, A. (2011). Adaptive Lerngelegenheiten in der Grundschule: Merkmale, methodisch-didakti-

- sche Schwerpunktsetzung und erforderliche Lehrerkompetenzen. In W. Helsper & R. Tippelt (Hrsg.), *Pädagogische Professionalität* (Zeitschrift für Pädagogik, 57. Beiheft, S. 819–833).
- Harlen, W. (2013). Inquiry-based learning in science and mathematics. *Review of Science, Mathematics and ICT Education*, 7(2), 9–33.
- Heinzel, F. (2008). Umgang mit Heterogenität in der Grundschule. In J. Ramseger & M. Wagener (Hrsg.), *Chancenungleichheit in der Grundschule. Ursachen und Wege aus der Krise* (S. 133–138). Wiesbaden: VS Verlag für Sozialwissenschaften.
https://doi.org/10.1007/978-3-531-91108-3_20
- Helmke, A. & Schrader, F.-W. (2010). Merkmale der Unterrichtsqualität: Potenzial, Reichweite und Grenzen. In B. Schaal & F. Huber (Hrsg.), *Qualitätssicherung im Bildungswesen: Auftrag und Anspruch der bayerischen Qualitätsagentur* (S. 69–108). Münster: Waxmann.
- Hertel, S. (2014). Adaptive Lerngelegenheiten in der Grundschule: Merkmale, methodisch-didaktische Schwerpunktsetzungen und erforderliche Lehrerkompetenzen. In B. Kopp, S. Martschinke, M. Munser-Kiefer, M. Haider, E. M. Kirschhock, G. Ranger & G. Renner (Hrsg.), *Individuelle Förderung und Lernen in der Gemeinschaft* (S. 19–34). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-04479-4_2
- Hochweber, J., Hosenfeldt, I. & Klieme, E. (2014). Classroom composition, classroom management, and the relationship between student attributes and grades. *Journal of Educational Psychology*, 106(1), 289–300. <https://doi.org/10.1037/a0033829>
- Hox, J. J. (2013). Multilevel regression and multilevel structural equation modeling. In T. D. Little (Hrsg.), *The oxford handbook of quantitative methods. Volume 2: Statistical analysis* (S. 281–294). New York, NY: Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199934898.013.0014>
- Igler, J., Ohle-Peters, A. & McElvany, N. (2019). Mit den Augen eines Grundschulkindes – Individuelle Prädiktoren für divergierende Schülereinschätzungen von Unterrichtsqualität. *Zeitschrift für Pädagogische Psychologie*, 33, 191–205.
<https://doi.org/10.1024/1010-0652/a000243>
- Jonen, A., Möller, K. & Hardy, I. (2003). Lernen als Veränderung von Konzepten – am Beispiel einer Untersuchung zum naturwissenschaftlichen Lernen in der Grundschule. In D. Cech & H. J. Schwier (Hrsg.), *Lernwege und Aneignungsformen im Sachunterricht* (S. 93–108). Bad Heilbrunn: Klinkhardt.
- Kleickmann, T., Steffensky, M. & Praetorius, A.-K. (2020). Quality of teaching in science education. More than three basic dimensions? *Zeitschrift für Pädagogik*, 66. Beiheft, 37–53.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts ‚Pythagoras‘. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–146). Münster: Waxmann.
- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54(2), 222–237.
- Klieme, E. & Warwas, J. (2011). Konzepte der individuellen Förderung. *Zeitschrift für Pädagogik*, 57(6), 805–818.
- Kuger, S. (2016). Curriculum and learning time in international school achievement studies. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Hrsg.), *Assessing contexts of learning. An international perspective* (S. 395–422). Berlin: Springer.
https://doi.org/10.1007/978-3-319-45357-6_16
- Kunter, M., Klusmann, U., Baumert, S., Richter, D., Voss, T. & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820.
<https://doi.org/10.1037/a0032583>
- Kunter, M. & Trautwein, U. (2013). *Psychologie des Unterrichts*. Stuttgart: UTB.
- Kunter, M. & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV* (S. 83–113). Münster: Waxmann.

- Lazarides, R. & Raufelder, D. (2017). Longitudinal effects of student-perceived classroom support on motivation – A latent change model. *Frontiers in Psychology*, 8, 1–11. <https://doi.org/10.3389/fpsyg.2017.00417>
- Lepper, C., Stang, J. & McElvany, N. (2020). *Zur Bedeutung wahrgenommener Motivierungsqualität im Unterricht für motivationale Merkmale von Lernenden*. Manuskript eingereicht zur Publikation.
- Lepper, C., Stang, J., Rieser, S. & McElvany, N. (2020). *Wie nehmen Grundschulkinder die Qualität ihres Unterrichts wahr? Ergebnisse aus TIMSS 2015*. Verfügbar unter: http://www.ifs.tu-dortmund.de/cms/de/Praxis/Wie-nehmen-Grundschulkinder-die-Qualitaet-ihres-Unterrichts-wahr_-Ergebnisse-aus-TIMSS-2015/Praxisportal_UQ_Okt_2020.pdf
- Lipowsky, F., Rakoczy, K., Drollinger-Vetter, B., Klieme, E., Reusser, K. & Pauli, C. (2009). Quality of geometry instruction and its short-term impact on students' understanding of Pythagorean Theorem. *Learning and Instruction*, 19(6), 527–537. <https://doi.org/10.1016/j.learninstruc.2008.11.001>
- Lotz, M. (2016). *Kognitive Aktivierung im Leseunterricht in der Grundschule. Eine Videostudie zur Gestaltung und Qualität von Leseübungen im ersten Schuljahr*. Wiesbaden: Springer Fachmedien. <https://doi.org/10.1007/978-3-658-10436-8>
- Mullis, I. V. S. & Martin, M. O. (Hrsg.). (2017). *TIMSS 2019 assessment frameworks*. Chestnut Hill: TIMSS & PIRLS International Study Center, Boston College. Verfügbar unter: <http://timssandpirls.bc.edu/timss2019/frameworks/>
- Ohle, A. & McElvany, N. (2016). Erfassung von Unterrichtsqualität in der Grundschule: Kognitiver Anspruch, Strukturierung und Motivierungsqualität. In N. McElvany, W. Bos, H. G. Holtappels, M. Gebauer & F. Schwabe (Hrsg.), *Bedingungen und Effekte guten Unterrichts* (Dortmunder Symposium der Empirischen Bildungsforschung, Bd. 1, S. 117–134). Münster: Waxmann.
- Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G. & Allen, M. (2018). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research*, 88(2), 205–242. <https://doi.org/10.3102/0034654317743198>
- Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119. <https://doi.org/10.3102/0013189X09332374>
- Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B. & Nehring, A. (2020). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswissenschaft*, 48(3), 1–38. <https://doi.org/10.1007/s42010-020-00082-8>
- Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM Mathematics Education*, 50(3), 407–426. <https://doi.org/10.1007/s11858-018-0918-4>
- Rakoczy, K. (2008). *Motivationsunterstützung im Mathematikunterricht. Unterricht aus der Perspektive von Lernenden und Beobachtern*. Münster: Waxmann.
- Rakoczy, K., Klieme, E. & Pauli, C. (2008). Die Bedeutung der wahrgenommenen Unterstützung motivationsrelevanter Bedürfnisse und des Alltagsbezugs im Mathematikunterricht für die selbstbestimmte Motivation. *Zeitschrift für Pädagogische Psychologie*, 22(1), 25–35. <https://doi.org/10.1024/1010-0652.22.1.25>
- Reeve, J. & Jang, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology*, 98(1), 209–218. <https://doi.org/10.1037/0022-0663.98.1.209>
- Rieser, S., Stahns, R., Walzebug, A. & Wendt, H. (2016). Einblicke in die Gestaltung des Mathematik- und Sachunterrichts. In H. Wendt, W. Bos, C. Selzer, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015.– Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 205–224). Münster: Waxmann.
- Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and wellbeing. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Schiepe-Tiska, A., Reiss, K., Obersteiner, A., Heine, J.-H., Seidel, T. & Prenzel, M. (2013). Mathematikunterricht in Deutschland: Befunde aus PISA 2012. In M. Prenzel, C. Sälzer, E. Klieme & O. Köller (Hrsg.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (S. 123–154). Münster: Waxmann.

- Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499. <https://doi.org/10.3102/0034654307310317>
- Stang, J., Schwabe, F. & McElvany, N. (in Druck). Bedeutung des wahrgenommenen motivierenden Umgangs für die intrinsische Lese- und Mathematikmotivation in der Sekundarstufe. *Zeitschrift für Erziehungswissenschaft*.
- Tomlinson, C. A. (2014). *Differentiated classroom: Responding to the needs of all learners*. Alexandria: ASCD.
- Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B. & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, 108, 705–721. <https://doi.org/10.1037/edu0000075>
- Wendt, H., Bos, W., Goy, M. & Jusufi, D. (Hrsg.). (2017). *TIMSS 2015. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente und Arbeit mit den Datensätzen*. Waxmann: Münster.

Kapitel 7

Privater Nachhilfeunterricht und Lehrkräftefortbildung am Ende der Grundschulzeit

Karin Guill und Heike Wendt

7.1 Einleitung

Neben dem regulären Schulunterricht erhält ein großer Anteil von Schülerinnen und Schülern zusätzlich privaten Nachhilfeunterricht. Diesen Umstand kann man als Zeichen eines hohen Engagements für gute Schulleistungen begrüßen, aber auch kritisch als Vertrauensverlust in das reguläre Schulsystem bewerten: Demnach würde dem Schulsystem nicht zugetraut, jede Schülerin und jeden Schüler optimal zu fördern (Bray, 2020; Klemm & Hollenbach-Biele, 2016). Vor diesem Hintergrund ist es sinnvoll, die Verbreitung von Nachhilfeunterricht an zentralen Punkten der Schullaufbahn im Blick zu behalten, um insbesondere Trends frühzeitig zu erkennen und gegebenenfalls bildungspolitisch zu reagieren, beispielsweise durch den Ausbau schulischer Förderangebote. Komplementär dazu schließt das Kapitel daher mit einem kursorischen Blick darauf, in welchem Umfang Lehrkräfte Fortbildungen nutzen, die der Weiterentwicklung der eigenen professionellen Kompetenz zur Förderung von Schülerinnen und Schülern dienen.

7.2 Verbreitung von Nachhilfeunterricht

Im Rahmen der Berichterstattung zu der *Trends in International Mathematics and Science Study* (TIMSS) 2015 wurde erstmals ein Kapitel der Nutzung von Nachhilfeunterricht in Deutschland und im internationalen Vergleich gewidmet. Auch wenn es Hinweise auf eine wachsende Nachfrage nach außerschulischem Nachhilfeunterricht in Deutschland gibt (Hille, Spieß & Staneva, 2016), konnte für das Ende der Grundschulzeit nur eine relativ geringe Nutzung konstatiert werden (Guill & Wendt, 2016). Nicht einmal jedes achte Kind erhielt zu diesem Zeitpunkt Nachhilfeunterricht in Mathematik. Sachunterricht spielte nahezu keine

Rolle als Nachhilfefach. Grundlage für das Verständnis von Nachhilfe in diesem Kapitel ist folgende Definition:

Nachhilfeunterricht

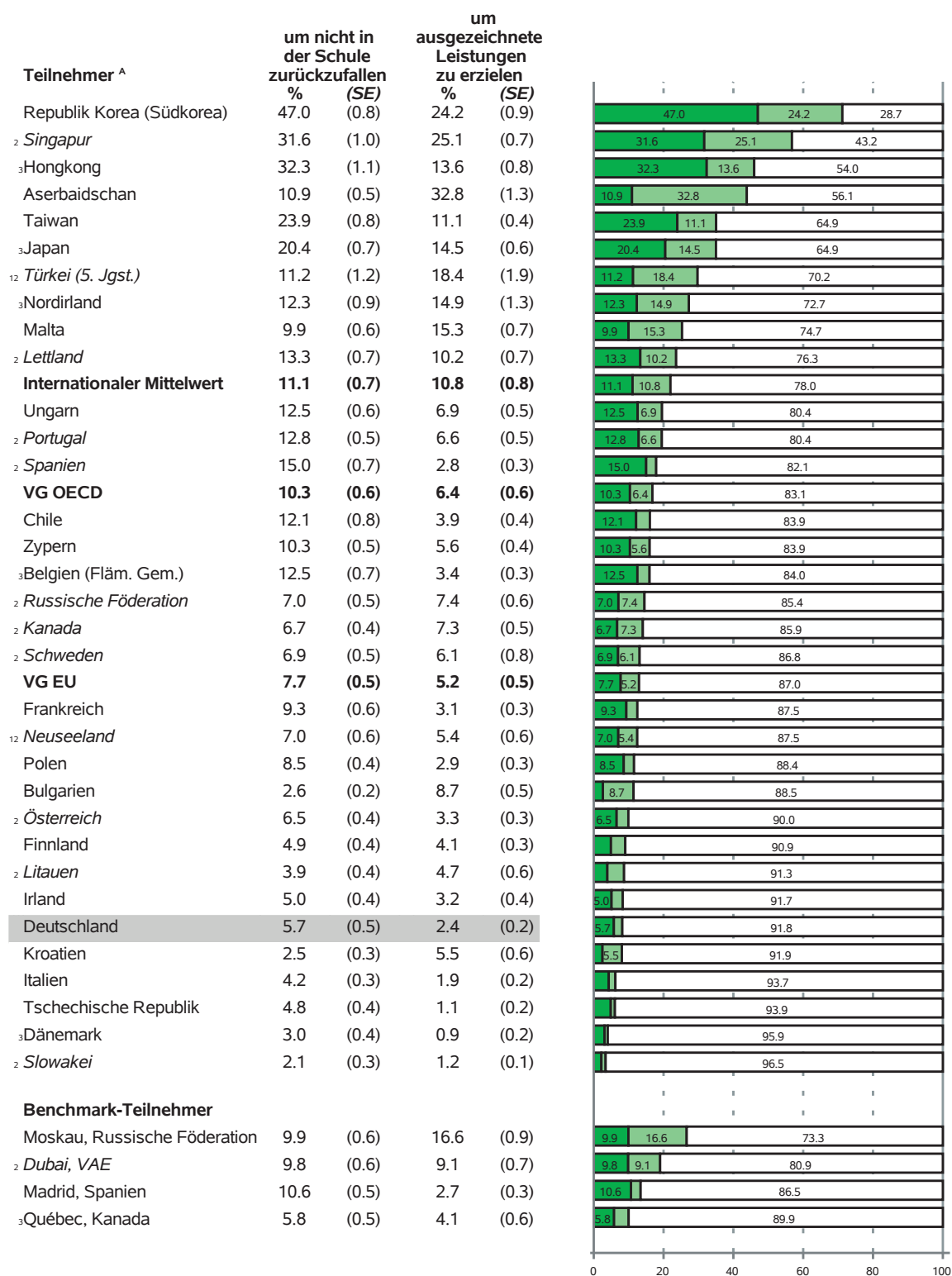
- findet sowohl außerfamiliär als auch außerhalb des Schulunterrichts statt,
- wird regelmäßig und häufig vorübergehend genutzt,
- wird durch Lehrkräfte, Studierende, (ältere) Schülerinnen und Schüler oder andere Personen erteilt,
- ist kostenpflichtig und
- zielt auf die Erfolgssicherung im Unterricht ab (Krüger, 1977).

Mit TIMSS 2019 ist es möglich, anhand von Fragen im internationalen Teil des Elternfragebogens, die Teilnahme an Nachhilfeunterricht in Mathematik und den Naturwissenschaften im Vergleich zu 2015 in den teilnehmenden Staaten zu beschreiben. Dabei muss nach wie vor berücksichtigt werden, dass in den verschiedenen Staaten sehr unterschiedliche Formen von Nachhilfeunterricht zu finden sind. Die Spanne reicht vom in Deutschland eher üblichen Einzel- oder Kleingruppenunterricht bis hin zu Nachhilfeunterricht in der auch Klassenstärke überschreitenden Großgruppe (Bray, 2009).

Zusätzliche Fragen in der nationalen Ergänzung der Elternfragebögen in der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) 2006, in TIMSS 2011, TIMSS 2015 und TIMSS 2019 ermöglichen es zudem, in Deutschland Trends in der Nutzung von Nachhilfeunterricht in ausgewählten Fächern zu identifizieren. Die jeweiligen Formulierungen der Fragen im nationalen sowie im internationalen Teil des Elternfragebogens weichen jedoch voneinander ab, sodass die Angaben zu den einzelnen Fragen nicht direkt miteinander verglichen werden können.¹ Im Folgenden wird zunächst die Nachhilfenutzung im internationalen Vergleich dargestellt und anschließend im Trend seit IGLU 2006.

Im internationalen Teil des Elternfragebogens in TIMSS 2015 und 2019 wird zwischen außerschulischem Nachhilfeunterricht zum Ausgleich von Leistungsdefiziten und Nachhilfeunterricht zum Erreichen von schulischen Spitzenleistungen unterschieden. In den zwölf Monaten vor der TIMSS-Datenerhebung erhielten in Deutschland in 2019 insgesamt 8.2 Prozent der Schülerinnen und Schüler Nachhilfe in Mathematik: 2.4 Prozent, um Spitzenleistungen zu erzielen und 5.7 Prozent zum Ausgleich von Leistungsdefiziten, was mit dem Anteil von Nachhilfeschülerinnen und -schülern in TIMSS 2015 vergleichbar ist (siehe Abbildung 7.1). Knapp die Hälfte dieser Schülerinnen und Schüler besuchte den Nachhilfeunterricht schon länger als acht Monate; das heißt, er hatte bereits im vorhergehenden Schuljahr begonnen. Die Verbreitung von Nachhilfeunterricht liegt in Deutschland in 2019 im Mittel unterhalb der Verbreitung in den teilnehmenden EU-Staaten, in denen 13.0 Prozent der Kinder Nachhilfeunterricht erhielten, und auch unter der in der Vergleichsgruppe der OECD-Staaten, in der 16.8 Prozent der Schülerinnen und Schüler Nachhilfeunterricht in Mathematik nutzten. Damit hat sich die Verbreitung von Nachhilfeunterricht in beiden Gruppen im Vergleich zu TIMSS 2015 nicht substanziell verändert. Warnungen vor einer wachsenden Verbreitung von Nachhilfeunterricht in Europa (Bray,

¹ Im internationalen Teil des Elternfragebogens wird seit TIMSS 2015 die Nutzung von außerschulischem Nachhilfeunterricht am Ende der Grundschulzeit erfragt. In der nationalen Ergänzung des Elternfragebogens wird seit IGLU 2006 nach der außerschulischen Förderung in Mathematik, im Lesen und in Rechtschreibung in den letzten Jahren gefragt.

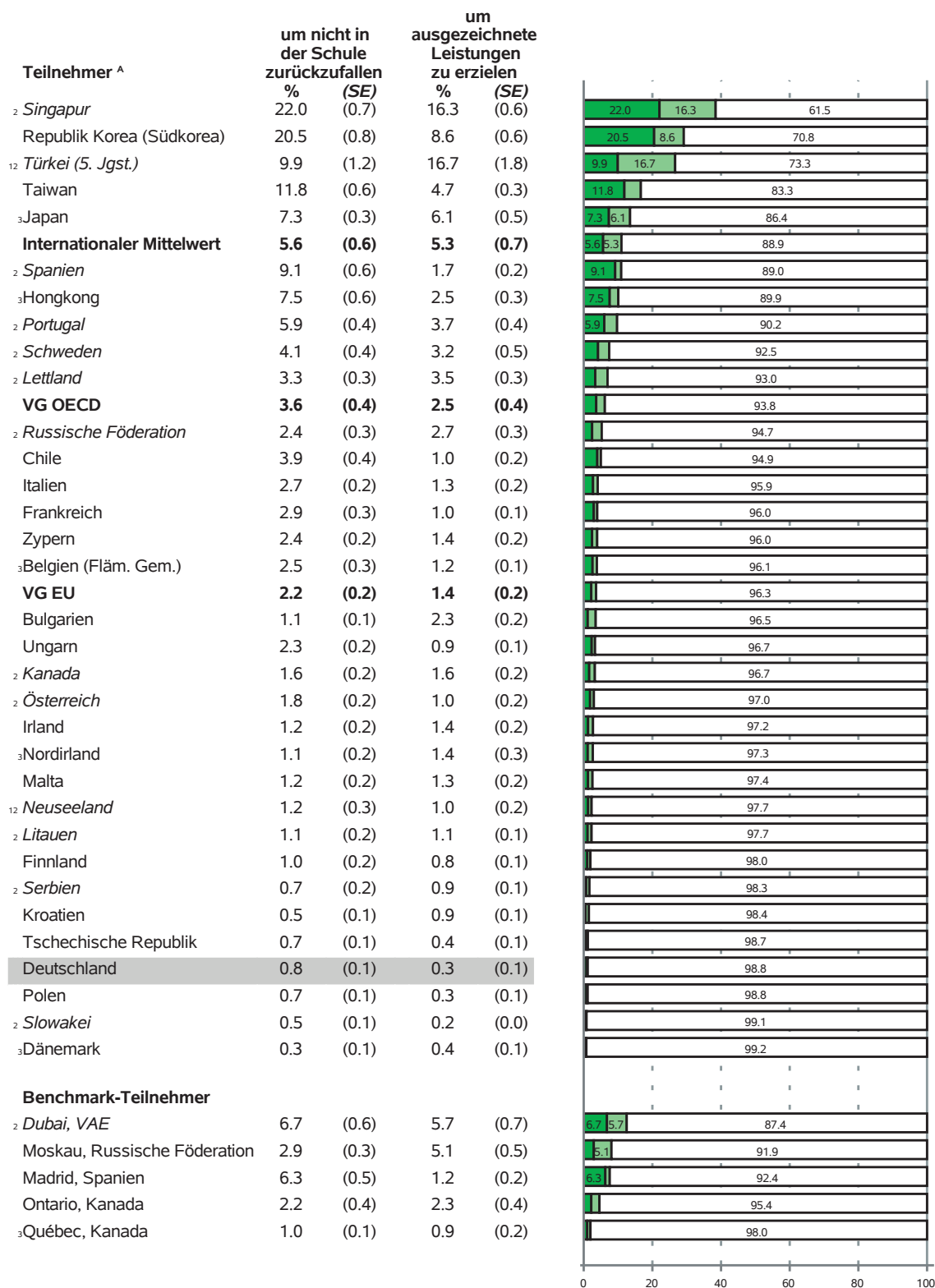
Abbildung 7.1: Privater Nachhilfeunterricht in Mathematik im internationalen Vergleich (Anteil der Schülerinnen und Schüler nach Elternangaben in Prozent)

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

■ Nachhilfe, um nicht in der Schule zurückzufallen
 ■ Nachhilfe, um ausgezeichnete Leistungen zu erzielen
 □ Keine Nachhilfe

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

¹ = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.² = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.³ = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.^A = In Australien, England, den Niederlanden, Norwegen und den USA wurde der Nachhilfeunterricht nicht erhoben. Somit sind diese Staaten nicht Teil dieser Darstellung und der berichteten Mittelwerte (International, VG EU, VG OECD).

Abbildung 7.2: Privater Nachhilfeunterricht in Sachunterricht (Anteil der Schülerinnen und Schüler nach Elternangaben in Prozent)

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

■ Nachhilfe, um nicht in der Schule zurückzufallen
 ■ Nachhilfe, um ausgezeichnete Leistungen zu erzielen
 □ Keine Nachhilfe

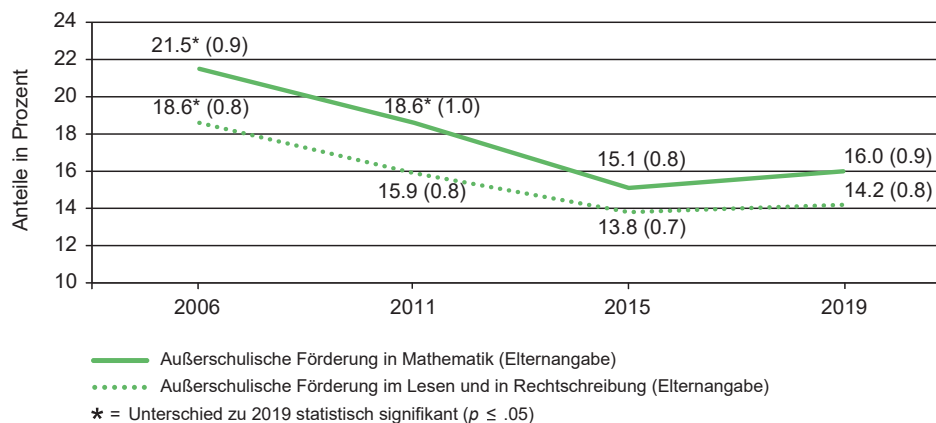
Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

¹ = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.² = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.³ = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.^A = In Australien, England, den Niederlanden, Norwegen und den USA wurde der Nachhilfeunterricht nicht erhoben. Somit sind diese Staaten nicht Teil dieser Darstellung und der berichteten Mittelwerte (International, VG EU, VG OECD).

2020) lassen sich zumindest für das Ende der Grundschulzeit nicht bestätigen. Spitzenreiter in Hinblick auf die Verbreitung von Nachhilfeunterricht sind die Staaten Republik Korea, Singapur, Hongkong und Aserbaidshan, in denen am Ende der Grundschulzeit mehr als die Hälfte beziehungsweise fast die Hälfte der Kinder privaten Nachhilfeunterricht nutzte. Die hohe Verbreitung von Nachhilfeunterricht in vielen südostasiatischen Staaten (Bray, 2009) spiegelt sich damit auch in den TIMSS-Daten wider.

Im Sachunterricht erhielten am Ende der Grundschulzeit in TIMSS 2019 nahezu keine Schülerinnen und Schüler in Deutschland Nachhilfeunterricht, um Spitzenleistungen zu erzielen. Knapp 1 Prozent nutzte Nachhilfeunterricht zum Ausgleich von Leistungsdefiziten, sodass sich insgesamt ein Anteil von 1.2 Prozent der Schülerinnen und Schüler ergibt (siehe Abbildung 7.2). Privater Nachhilfeunterricht im Fach Sachunterricht ist nur in wenigen Staaten wie Singapur, der Republik Korea und der Türkei sehr verbreitet. Im Mittel der OECD-Staaten erhielten nur 6.2 Prozent der Schülerinnen und Schüler Nachhilfeunterricht im Sachunterricht, im Mittel der EU-Staaten nur 3.6 Prozent. Da Nachhilfeunterricht im Sachunterricht in Deutschland in TIMSS 2019 wie schon in TIMSS 2015 nahezu keine Rolle spielt, liegt der Fokus im Folgenden auf Nachhilfeunterricht im Fach Mathematik.

Abbildung 7.3: Außerschulische Förderung im Vergleich von IGLU 2006 und TIMSS 2011 bis TIMSS 2019 (Anteil der Schülerinnen und Schüler nach Elternangaben in Prozent)



IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Mit Blick auf internationale Trends wird seit vielen Jahren vor einer zunehmenden Verbreitung von Nachhilfeunterricht gewarnt. Wenn Nachhilfeunterricht vor allem von Schülerinnen und Schülern aus privilegierten sozialen Lagen genutzt wird, könnte er den Zusammenhang zwischen sozialer Herkunft und Bildungserfolg verstärken (Bray, 2020). Auch für Deutschland konnte gezeigt werden, dass im Trend seit Anfang der 2000er-Jahre immer mehr Schülerinnen und Schüler im Laufe ihrer Schulzeit privaten Nachhilfeunterricht in Anspruch nehmen (Hille et al., 2016). Dieser Trend ließ sich jedoch für das Ende der Grundschulzeit zuletzt nicht bestätigen (Guill & Wendt, 2016). Im Gegenteil: Nach Angaben der Erziehungsberechtigten in IGLU 2006 erhielten 21.5 Prozent der Schülerinnen und Schüler eine außerschulische Förderung in Mathematik und 18.6 Prozent in Lesen und Rechtschreibung, während dies in TIMSS 2015 nur 15.1 Prozent für Mathematik und 13.8 Prozent für Lesen und Rechtschreibung

angaben. In TIMSS 2019 haben sich die Nachhilfequoten auf diesem Niveau stabilisiert: 16.0 Prozent der Erziehungsberechtigten berichten von außerschulischer Förderung in Mathematik und 14.2 Prozent von außerschulischer Förderung in Lesen und Rechtschreibung (siehe Abbildung 7.3).

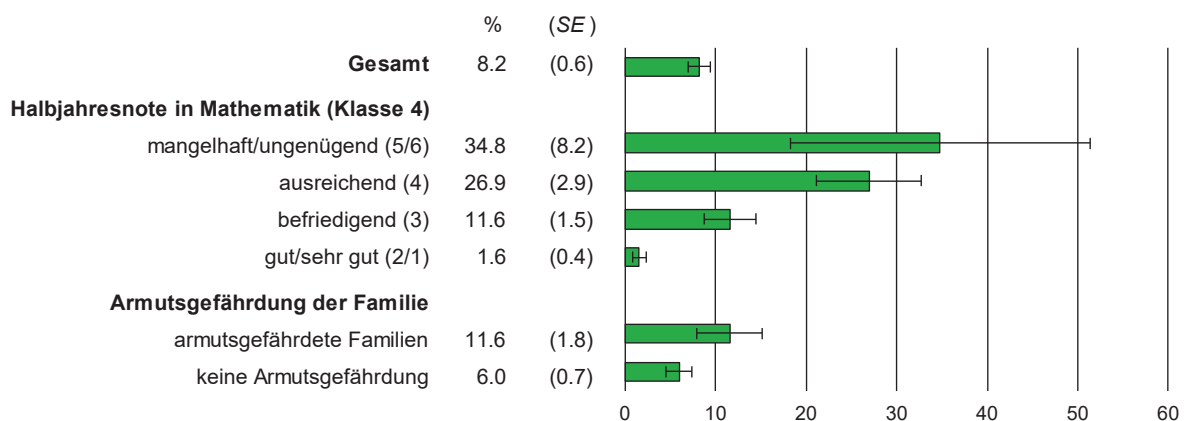
Ähnlich wie in TIMSS 2015 bleibt anzunehmen, dass der Ausbau inner-schulischer Förderangebote, der inzwischen geringer Anteil von Ländern der Bundesrepublik Deutschland mit verbindlichen Übergangsempfehlungen (Guill & Lintorf, 2019) und der Ausbau von Schulen mit mehreren Bildungsgängen (siehe Kapitel 11 in diesem Band) Erziehungsberechtigte von der subjektiv empfundenen Notwendigkeit entlastet, schon in der Grundschule in privaten Nachhilfeunterricht zu investieren, um ihrem Kind den Weg zum Abitur offenzuhalten.

7.3 Nutzung von privatem Nachhilfeunterricht in verschiedenen Schülerinnen- und Schülergruppen

Die beiden Fragen, wer Nachhilfeunterricht bekommt und ob Nachhilfeunterricht auch dort ankommt, wo er am meisten benötigt wird, sind zentral für einen differenzierten Blick auf die Verbreitung von Nachhilfeunterricht in Deutschland, da sie aus unterschiedlichen Perspektiven Fragen der Bildungsgerechtigkeit betreffen. Weil Nachhilfeunterricht in Deutschland vor allem in den Hauptfächern Deutsch und Mathematik genutzt wird, liegt auch der Fokus dieses Abschnitts auf Nachhilfeunterricht in Mathematik, während Nachhilfeunterricht im Fach Sachunterricht nicht weiter berücksichtigt wird.

Nachhilfeunterricht zielt nach Krüger (1977), wie zu Beginn des Kapitels angeführt, auf die Erfolgssicherung im Unterricht ab. Demnach wäre er vor allem bei leistungsschwachen Schülerinnen und Schülern zu erwarten. Tatsächlich zeigt sich, dass vor allem Schülerinnen und Schüler mit mangelhaften und schlechteren Noten in Mathematik (34.3%) sowie Schülerinnen und Schüler mit ausreichenden Noten (26.9%) Nachhilfeunterricht erhalten. Sind die Noten befriedigend oder gut beziehungsweise sehr gut, nutzen nur 11.6 Prozent beziehungsweise 1.6 Prozent Nachhilfeunterricht (siehe Abbildung 7.4).² Ein ähnliches Muster

Abbildung 7.4: Privater Nachhilfeunterricht in Mathematik differenziert nach Halbjahresnoten und Armutsrisiko der Schülerinnen und Schüler (Anteile der Schülerinnen und Schüler in Prozent)



IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

² Abweichungen zu Abbildung 7.1 ergeben sich durch Rundungsfehler.

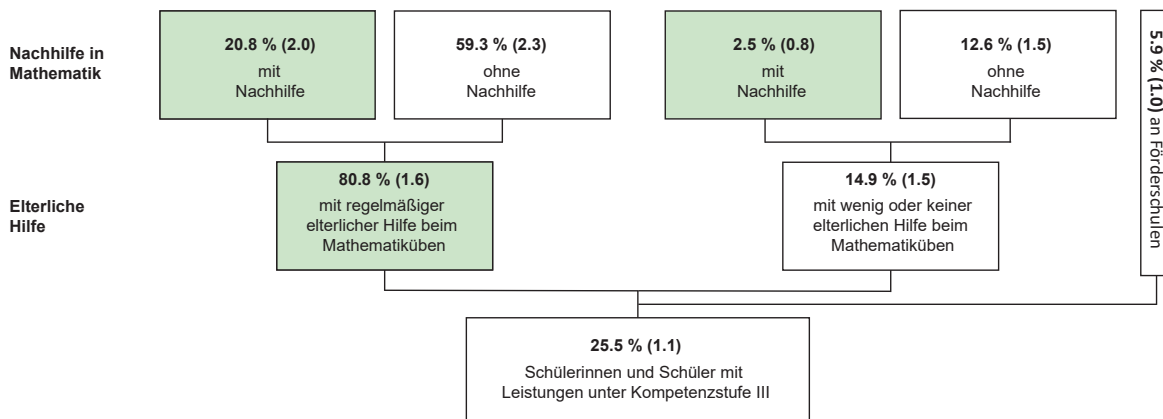
hatte sich bereits in TIMSS 2015 gezeigt: Nachhilfe wird in erster Linie von Schülerinnen und Schülern mit ausreichenden und mangelhaften Schulleistungen genutzt (Guill & Wendt, 2016). Die genauen Quoten, welcher Anteil der Schülerinnen und Schüler in den einzelnen Untergruppen Nachhilfeunterricht nutzt, sind zwischen TIMSS 2015 und TIMSS 2019 wegen Abweichungen in der Datenaufbereitung nicht vergleichbar (siehe Kapitel 2 in diesem Band).

Da Nachhilfeunterricht in der Regel mit Kosten für die Familien verbunden ist, gilt es zu prüfen, ob dieser auch für weniger wohlhabende Familien zugänglich ist. Tatsächlich zeigt sich in TIMSS 2019, dass in armutsgefährdeten Familien mit einem Anteil von 11.6 Prozent sogar signifikant mehr Kinder Nachhilfeunterricht nutzen als in Familien ohne Armutsgefährdung mit einem Anteil von 6.0 Prozent. Allerdings sind leistungsschwächere Kinder in den armutsgefährdeten Familien überrepräsentiert (siehe Kapitel 9 in diesem Band), sodass nicht ausgeschlossen werden kann, dass in nicht armutsgefährdeten Familien bei geringeren Leistungsdefiziten mit Nachhilfe reagiert wird als in armutsgefährdeten Familien. Umso wichtiger ist es, im Blick zu behalten, dass gerade diese Kinder durch schulische Förderangebote erreicht werden, zumal vermutet werden könnte, dass Kinder aus armutsgefährdeten Familien kostengünstigere und damit potenziell weniger wirksame Nachhilfeangebote nutzen.

Nach diesem Blick darauf, in welchen Schülerinnen- und Schülergruppen Nachhilfeunterricht besonders verbreitet ist, soll umgekehrt auch betrachtet werden, ob alle Schülerinnen und Schüler mit besonderem Unterstützungsbedarf auch zusätzliche Unterstützung erhalten. In Deutschland erreicht mehr als jedes vierte Grundschulkind (25.5%) lediglich mathematische Kompetenzen unterhalb der Kompetenzstufe III (siehe Kapitel 3 in diesem Band) und hat damit einen erheblichen Unterstützungsbedarf, um nach dem Übergang in die Sekundarstufe I erfolgreich weiterzulernen. Erziehungsberechtigte, Schule und Nachhilfeunterricht könnten diese Hilfe theoretisch leisten.

Ähnlich wie in TIMSS 2015 kann auch in TIMSS 2019 ein großer Anteil (80.8%) der leistungsschwächeren Schülerinnen und Schüler auf elterliche Hilfe beim Mathematiküben zurückgreifen (siehe Abbildung 7.5). Ein Fünftel (20.8%) erhält zusätzlich privaten Nachhilfeunterricht in Mathematik. Ein kleiner Anteil von 2.5 Prozent erhält zwar wenig oder keine elterliche Hilfe, dafür aber privaten Nachhilfeunterricht. Rund ein Achtel (12.6%) erhält dagegen weder elterliche Hilfe beim Mathematiküben noch privaten Nachhilfeunterricht. Diese Schülerinnen und Schüler sind besonders auf schulische Fördermaßnahmen, sei es in Form von zusätzlichen Förderstunden oder im Rahmen der Binnendifferenzierung, angewiesen. Inwieweit dies spezifisch für Mathematik auch erfolgt, wurde in TIMSS 2019 im Gegensatz zu TIMSS 2015 nicht erfasst. 5.9 Prozent der leistungsschwachen Schülerinnen und Schüler besuchen aktuell Förderschulen. Mit Blick auf nicht immer optimale Formen der Unterstützung von Lernprozessen im Elternhaus und im Nachhilfeunterricht (Guill, 2020), bleibt zu wünschen, dass die außerschulische Unterstützung leistungsschwacher Schülerinnen und Schüler im Sinne von Bildungspartnerschaften von Elternhaus und Schule (Wild, 2020) von schulischer Seite begleitet wird.

Abbildung 7.5: Außerschulische Unterstützung von Schülerinnen und Schülern mit Mathematikleistungen unterhalb der Kompetenzstufe III (Anteil der Schülerinnen und Schüler in Prozent und deren Standardfehler in Klammern, bezogen auf alle Schülerinnen und Schüler mit entsprechenden Leistungen)



Lesebeispiel: 20.8 Prozent (Kasten links oben) der Schülerinnen und Schüler mit Leistungen unter Kompetenzstufe III erhalten elterliche Hilfe beim Mathematiküben und außerschulische Nachhilfe in Mathematik.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

7.4 Wissen der Lehrkräfte um den Besuch von privatem Nachhilfeunterricht

In Studien zur Verbreitung von Nachhilfeunterricht wird von Zeit zu Zeit diskutiert, ob Schülerinnen und Schüler und ihre Erziehungsberechtigten offen mit dem Thema umgehen oder lieber geheim halten, dass das Kind Nachhilfe erhält (Rudolph, 2002). Für eine Geheimhaltung spreche, dass es mit Scham besetzt sei, gute Schulleistungen nur mithilfe von Nachhilfe zu erbringen. Erziehungsberechtigte und Kinder könnten befürchten, dass die Leistungen des Kindes kritischer beurteilt werden, wenn sie mithilfe von Nachhilfeunterricht erreicht werden, oder dass höhere Erwartungen an das Kind gestellt werden. Demgegenüber könnten Lehrkräfte, die über den Nachhilfeunterricht informiert sind, mit den Nachhilfelehrkräften kooperieren, diagnostisches Wissen um Leistungsdefizite austauschen und gezielte Hinweise für die Förderung im Nachhilfeunterricht geben. In Hinblick auf die Empfehlung für die weitere Schullaufbahn lässt sich Nachhilfeunterricht positiv als zusätzliche Ressource und Unterstützungsbereitschaft des Elternhauses oder negativ als mangelnde Kompetenz oder fehlende Selbstständigkeit des Kindes werten. Vor diesem Hintergrund wurde in TIMSS 2019 im Rahmen der Schülerteilnahmeliste für jedes Kind erfragt, ob es nach Wissen der Lehrkraft in den Hauptfächern Mathematik und Deutsch/Lesen/Rechtschreibung in den letzten zwölf Monaten Nachhilfeunterricht erhalten habe. Die Frage ist angelehnt an den internationalen Teil des Elternfragebogens, sodass ein direkter Vergleich der Angaben möglich ist.

In fast 40 Prozent der Fälle war es den Lehrkräften nach eigenen Angaben nicht bekannt, ob ein Kind Nachhilfeunterricht in Mathematik oder in Deutsch erhält (siehe Tabelle 7.1). Erhielt ein Kind nach Elternangabe Nachhilfeunterricht in Mathematik, wussten die Lehrkräfte in nicht einmal einem Drittel dieser Fälle (28 %) um diesen Nachhilfeunterricht. Bei fast der Hälfte (45 %) dieser Kinder ging die Lehrkraft davon aus, dass es keinen privaten Nachhilfeunterricht erhalte, bei knapp einem Viertel (24 %) konnte die Lehrkraft dazu keine Angabe machen. Für das Fach Deutsch zeigt sich ein ähnliches Muster.

Erhielt ein Kind nach Elternangaben keinen Nachhilfeunterricht in Mathematik oder Deutsch, gingen die Lehrkräfte in wenigen Fällen (1 %) trotzdem davon aus, dass Nachhilfeunterricht stattfindet. Hier muss offenbleiben, ob die Lehrkräfte sich irren oder die Erziehungsberechtigten den Nachhilfeunterricht im TIMSS-Elternfragebogen nicht angeben wollten. In fast 60 Prozent der Fälle gaben die Lehrkräfte übereinstimmend mit den Erziehungsberechtigten an, dass kein privater Nachhilfeunterricht genutzt werde, in fast 40 Prozent der Fälle wussten sie nach eigenen Angaben nicht, ob das Kind Nachhilfeunterricht erhält. Auch hier zeigt sich für das Fach Deutsch ein vergleichbares Muster.

Zusammenfassend lässt sich sagen, dass privater Nachhilfeunterricht am Ende der Grundschulzeit in Deutschland zu einem sehr großen Anteil tatsächlich Privatsache ist. Das muss nicht zwingend schlecht sein, heißt aber auch, dass in diesen Fällen zwischen schulischen Lehrkräften und Nachhilfelehrkräften kein Austausch und keine Abstimmung stattfindet, wie die betroffenen Kinder optimal zu fördern sind. In Hinblick auf die ernüchternden Befunde zur Wirksamkeit von privatem Nachhilfeunterricht (Guill, Lüdtke & Köller, 2020; Park, Buchmann, Choi & Merry, 2016) muss das als vertane Chance bewertet werden.

Tabelle 7.1: Übereinstimmung der Angaben von Erziehungsberechtigten und Lehrkräften zur Nutzung von privatem Nachhilfeunterricht in Mathematik und Deutsch (Lesen und Rechtschreibung)

Elternangabe \ Lehrkräfteangabe	Mathematik					Deutsch				
	Ja, Nachhilfe (%)	Keine Nachhilfe (%)	Nicht bekannt (%)	Keine Angabe (%)	Summe (n)	Ja, Nachhilfe (%)	Keine Nachhilfe (%)	Nicht bekannt (%)	Keine Angabe (%)	Summe (n)
Ja, Nachhilfe (%)	28	45	24	3	181	20	52	26	2	277
Nein, keine Nachhilfe (%)	1	59	37	3	2008	1	58	38	3	1951
Kein Fragebogen ausgefüllt/ Keine Angabe (%)	6	53	40	1	1248	6	53	40	1	1209
Gesamt	4	56	38	2	3437	4	56	38	2	3437

Lesebeispiel: Wenn Kinder nach Angaben ihrer Erziehungsberechtigten Nachhilfe in Mathematik erhalten, sind Lehrkräfte in 28 Prozent dieser Fälle informiert, in 45 Prozent der Fälle gehen sie davon aus, dass das Kind keine Nachhilfe erhält, in 24 Prozent der Fälle ist es ihnen nicht bekannt und 3 Prozent haben dazu keine Angabe gemacht.

7.5 Teilnahme der Lehrkräfte an Fortbildungen

Zum Verständnis des Lehrerberufs gehört in Deutschland die ständige Erweiterung professioneller Kompetenzen (KMK, 2014). Den formalen Lerngelegenheiten, wie der Teilnahme an institutionell angebotenen Fortbildungsveranstaltungen, kommt eine besondere Bedeutung zu, da zum einen ihr Besuch per Rechtsverordnung beziehungsweise Schulgesetz geregelt ist und zum anderen die Fortbildungsforschung zeigt, dass der Besuch Veränderungen im Wissen und Handeln von Lehrkräften und schließlich Wirkungen auf Schülerinnen und Schüler haben kann (Fussangel, Rürup & Gräsel, 2016). Angesichts der zunehmenden Heterogenität der Schülerschaft in Deutschland und der Herausforderungen, die sich damit an ein adaptives Unterrichten ergeben, wäre es wünschenswert, dass sich Lehrkräfte aktiv und kontinuierlich fortbilden. In Tabelle 7.2 ist für TIMSS 2019 dargestellt, wie viel Prozent der Schülerinnen und Schüler von Mathematiklehrkräften unterrichtet wurden, die nach eigenen

Tabelle 7.2: Fortbildungsteilnahme der Mathematiklehrkräfte (international) nach Inhalten (Anteile der Schülerinnen und Schüler mit entsprechenden Fortbildungen der Mathematiklehrkräfte in Prozent)

Teilnehmer	Mathematische Inhalte		Mathematikdidaktik		Lehrplan zum Mathematikunterricht		Stärkung kritischen Denkens oder Problemlösens		Integration von Informationstechnologien		Leistungs-feststellung im Mathematikunterricht		Eingehen auf die individuellen Bedürfnisse der Schülerinnen und Schüler	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)
² Singapur	73.5	(2.4)	81.8	(2.6)	69.3	(2.3)	61.9	(2.6)	61.1	(2.3)	67.3	(2.5)	45.9	(2.5)
³ Hongkong	72.4	(3.2)	72.1	(3.3)	72.3	(3.9)	57.6	(4.3)	76.5	(3.9)	40.4	(4.3)	56.3	(4.1)
² ³ USA	66.0	(2.3)	65.4	(2.6)	69.9	(2.2)	58.5	(2.5)	47.0	(2.3)	48.3	(2.5)	55.5	(2.5)
Polen	73.5	(3.5)	55.4	(3.6)	69.4	(3.6)	46.1	(4.0)	62.6	(3.5)	38.1	(4.2)	54.1	(3.5)
Malta	59.9	(0.4)	61.6	(0.4)	54.7	(0.4)	62.2	(0.4)	40.5	(0.4)	67.2	(0.4)	45.4	(0.4)
² Kanada	62.2	(2.0)	67.5	(1.8)	46.9	(2.8)	58.4	(2.1)	37.1	(3.0)	51.8	(2.8)	52.0	(2.0)
¹ ² England	65.6	(5.9)	62.6	(6.9)	57.2	(6.8)	63.3	(5.4)	17.4	(4.3)	46.7	(6.8)	42.0	(6.8)
² Russische Föderation	37.4	(3.8)	43.2	(3.8)	49.6	(3.7)	50.8	(3.5)	55.3	(3.3)	54.7	(3.7)	51.0	(4.2)
Australien	55.7	(3.6)	54.4	(3.4)	52.5	(3.4)	55.1	(3.6)	29.7	(3.3)	42.8	(3.1)	51.5	(3.8)
Taiwan	52.6	(3.7)	46.2	(4.0)	58.6	(4.1)	45.3	(3.7)	40.5	(3.7)	43.6	(3.6)	53.6	(4.0)
¹ ² Neuseeland	59.0	(3.6)	60.9	(3.3)	49.8	(3.6)	57.6	(3.2)	25.6	(2.8)	36.4	(3.4)	45.7	(3.1)
Aserbaidschan	48.6	(3.9)	38.6	(3.6)	60.5	(3.7)	43.1	(3.7)	33.4	(3.7)	56.6	(3.5)	39.3	(3.6)
² Lettland	42.0	(3.6)	37.2	(3.4)	34.4	(3.6)	67.4	(3.5)	48.2	(3.6)	36.7	(3.4)	42.2	(3.6)
³ Nordirland	50.4	(4.2)	57.0	(3.9)	39.0	(4.2)	55.8	(4.1)	36.8	(4.2)	29.6	(4.2)	33.4	(4.2)
Internationaler Mittelwert	45.5	(3.4)	45.0	(3.5)	41.4	(3.5)	43.5	(3.4)	34.6	(3.3)	36.8	(3.3)	42.6	(3.5)
Kroatien	49.2	(3.3)	34.5	(3.9)	35.8	(3.0)	43.8	(3.3)	42.6	(3.3)	35.7	(3.4)	42.6	(3.6)
Zypern	66.7	(3.7)	50.9	(4.4)	63.9	(3.8)	32.3	(3.6)	25.8	(2.9)	25.1	(2.4)	14.8	(2.6)
² Litauen	24.0	(2.9)	28.6	(3.5)	24.2	(3.6)	53.4	(3.5)	55.3	(3.5)	33.9	(3.7)	45.1	(4.1)
² Spanien	37.4	(3.0)	45.4	(3.0)	27.8	(3.4)	45.2	(3.4)	28.2	(3.3)	23.7	(2.8)	48.5	(3.7)
Tschechische Republik	31.5	(3.6)	43.4	(3.8)	14.4	(2.7)	46.9	(3.7)	31.1	(3.4)	27.8	(3.7)	51.2	(3.1)
Frankreich	64.3	(3.4)	68.2	(3.4)	35.6	(3.6)	38.6	(3.1)	12.5	(2.3)	5.9	(1.6)	20.6	(2.5)
Republik Korea (Südkorea)	30.0	(4.3)	28.1	(3.7)	38.7	(4.1)	38.1	(3.8)	14.7	(2.9)	38.0	(4.0)	49.2	(4.0)
Italien	32.9	(3.6)	47.0	(3.5)	29.5	(3.5)	26.6	(3.4)	32.9	(3.0)	24.9	(3.3)	41.6	(3.9)
VG OECD	39.0	(3.5)	39.8	(3.5)	32.1	(3.5)	36.4	(3.4)	25.9	(3.0)	25.0	(3.3)	37.2	(3.6)
VG EU	38.3	(3.4)	37.8	(3.5)	31.8	(3.5)	35.2	(3.4)	27.3	(3.0)	24.6	(3.2)	35.5	(3.6)
³ Japan	54.5	(3.6)	58.1	(2.9)	22.7	(2.7)	28.5	(3.0)	11.3	(2.5)	20.9	(3.1)	32.7	(3.3)
² Österreich	49.7	(3.0)	40.7	(2.9)	15.8	(2.4)	33.5	(3.4)	13.4	(1.8)	21.7	(2.7)	50.5	(2.8)
Chile	42.5	(4.3)	33.3	(4.2)	29.6	(3.9)	36.3	(4.3)	17.2	(3.3)	24.1	(3.8)	41.8	(3.6)
² Portugal	40.5	(3.3)	37.4	(3.5)	30.2	(3.0)	33.1	(3.1)	24.5	(3.0)	14.1	(2.2)	37.0	(3.4)
Irland	35.7	(3.5)	32.9	(3.4)	23.4	(3.4)	34.4	(3.5)	23.3	(3.1)	18.6	(2.7)	31.2	(3.4)
³ Niederlande	39.4	(4.6)	33.8	(4.6)	23.7	(5.0)	26.0	(4.7)	4.3	(1.4)	21.8	(4.2)	46.0	(5.1)
Bulgarien	29.5	(3.3)	25.0	(3.0)	29.3	(3.9)	19.3	(3.0)	25.9	(3.4)	25.5	(3.3)	24.6	(3.5)
² Slowakei	14.5	(2.4)	22.7	(2.7)	42.3	(2.6)	20.8	(2.3)	36.5	(2.9)	14.8	(2.2)	22.1	(2.6)
² Schweden	21.6	(3.1)	22.5	(3.1)	23.1	(3.8)	17.4	(3.0)	29.4	(4.0)	24.4	(3.9)	33.1	(4.5)
Deutschland	37.2	(3.3)	27.8	(3.4)	26.6	(3.3)	20.3	(3.2)	8.3	(1.9)	17.8	(2.8)	29.3	(3.2)
¹ ² Türkei (5. Jgst.)	22.2	(3.2)	27.1	(3.0)	31.7	(3.0)	16.8	(3.3)	32.4	(3.4)	15.1	(2.9)	13.5	(2.6)
³ Belgien (Fläm. Gem.)	16.6	(2.7)	18.6	(2.9)	35.1	(3.7)	14.8	(2.9)	7.8	(1.8)	4.5	(1.7)	34.1	(3.5)
³ Dänemark	26.1	(3.5)	21.7	(3.1)	13.4	(2.7)	14.5	(2.9)	20.0	(3.0)	16.7	(2.7)	15.1	(3.2)
¹ ³ Norwegen (5. Jgst.)	23.3	(3.9)	24.5	(4.0)	7.7	(2.4)	18.7	(3.4)	17.5	(3.8)	12.4	(3.1)	13.9	(3.2)
Ungarn	4.6	(1.4)	16.0	(2.4)	3.7	(1.4)	20.8	(2.3)	11.9	(1.9)	5.2	(1.3)	29.6	(3.2)
² Finnland	7.4	(1.9)	16.5	(2.6)	11.7	(2.0)	8.0	(1.9)	15.1	(2.2)	9.2	(1.8)	16.9	(2.6)
Benchmark-Teilnehmer														
² Dubai, VAE	68.9	(2.1)	68.5	(2.5)	72.5	(1.2)	84.4	(2.0)	77.1	(1.7)	69.4	(2.3)	81.1	(2.1)
Moskau, Russische Föderation	25.7	(3.5)	26.8	(3.8)	36.8	(4.1)	36.4	(4.3)	61.6	(4.1)	34.7	(4.3)	40.1	(3.9)
Madrid, Spanien	38.6	(4.0)	43.6	(4.0)	23.1	(3.3)	41.3	(3.9)	32.7	(4.5)	22.0	(3.7)	49.8	(4.1)
³ Québec, Kanada	38.2	(4.0)	49.9	(4.4)	19.8	(3.3)	33.8	(4.3)	22.2	(3.9)	44.0	(4.1)	26.5	(3.8)

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

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

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

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

Angaben in den letzten zwei Jahren Fortbildungsveranstaltungen in verschiedenen Bereichen des Mathematikunterrichts besucht haben. Im internationalen Vergleich zeigt sich zunächst, dass in Deutschland dieser Anteil zumindest nominell in allen Bereichen unterhalb des Niveaus des europäischen Durchschnitts liegt. In Bezug auf *Mathematische Inhalte* ist dieser Anteil nahe beim europäischen Niveau. Die Abweichung ist für Fortbildungen im Bereich *Integration von Informationstechnologien* mit 8 Prozent der Schülerinnen und Schüler in Deutschland versus 27 Prozent der Schülerinnen und Schüler im EU-Durchschnitt am größten. Hier bildet Deutschland zusammen mit den Niederlanden (4.3%) und der Flämischen Gemeinschaft in Belgien (7.8%) im internationalen Vergleich das Schlusslicht. Nur rund jede dritte bis fünfte Schülerin beziehungsweise jeder dritte bis fünfte Schüler wurde in Deutschland von Lehrkräften unterrichtet, die Angebote zur *Mathematikdidaktik* (27.8%), zum *Eingehen auf die individuellen Bedürfnisse der Schülerinnen und Schüler* (29.3%), zur *Leistungsfeststellung*

im Mathematikunterricht (17.8%) und zur Stärkung kritischen Denkens oder Problemlösens (20.3%) – die besonders für einen adaptiven Mathematikunterricht relevant sind – in den letzten zwei Jahren auf dem Fortbildungsplan hatten.

Für die Sachunterrichtslehrkräfte zeigt sich ein vergleichbares Bild (Tabelle 7.3). Der Anteil an Schülerinnen und Schüler, die von Lehrkräften unterrichtet wurden, die in den letzten zwei Jahren Fortbildungen im Bereich des naturwissenschaftlichen Sachunterrichts besucht haben, liegt in fast allen Bereichen nominell unterhalb des europäischen Durchschnitts. Die einzige Ausnahme stellt der Bereich der *Naturwissenschaftliche Inhalte* (32.6%) dar. Für Veranstaltungen zur *Didaktik des naturwissenschaftlichen Sachunterrichts* (22.1%), zum *Lehrplan zum naturwissenschaftlichen Sachunterrichts* (20.7%), zur *Leistungsfeststellung im naturwissenschaftlichen Sachunterricht* (11.9%) und zum *Eingehen auf die individuellen Bedürfnisse der Schülerinnen und Schüler* (18.9%) sind die Abweichungen zum europäischen Niveau relativ gering. Veranstaltungen zur

Tabelle 7.3: Fortbildungsteilnahme der Sachunterrichtslehrkräfte (international) nach Inhalten (Anteile der Schülerinnen und Schüler mit entsprechenden Fortbildungen der Sachunterrichtslehrkräfte in Prozent)

Teilnehmer	Naturwissen- schaftliche Inhalte		Didaktik des naturwiss. Sach- unterrichts		Lehrplan zum naturwiss. Sach- unterricht		Stärkung kritischen Denkens oder Problem- lösens		Integration von Informations- technologien		Leistungs- feststellung im naturwiss. Sach- unterricht		Eingehen auf die individuellen Bedürfnisse der Schülerinnen und Schüler		Naturwiss. Sachunterricht fächer- übergreifend gestalten	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)
² Singapur	67.3	(2.6)	80.5	(2.0)	63.0	(2.6)	48.6	(2.5)	57.7	(2.4)	70.9	(2.5)	35.8	(2.6)	29.6	(2.6)
³ Hongkong	52.5	(4.1)	46.7	(4.5)	53.7	(4.8)	60.7	(4.7)	65.7	(4.9)	34.3	(4.3)	48.1	(4.1)	58.1	(4.4)
Taiwan	61.2	(3.8)	40.6	(3.8)	60.7	(4.0)	44.3	(3.6)	53.3	(4.0)	28.5	(4.0)	50.2	(3.4)	44.6	(3.5)
Polen	63.8	(3.9)	32.5	(3.9)	56.7	(4.2)	45.3	(3.9)	55.6	(3.7)	34.3	(3.8)	51.9	(4.2)	28.4	(3.7)
² Russische Föderation	32.5	(3.4)	36.3	(3.5)	47.6	(3.8)	45.7	(3.4)	50.1	(3.6)	48.2	(3.5)	47.4	(3.9)	49.9	(3.5)
Lettland	36.6	(3.9)	32.7	(3.6)	31.8	(3.8)	61.9	(3.9)	44.3	(3.8)	27.6	(3.4)	41.1	(4.3)	40.3	(3.6)
Kroatien	56.6	(3.5)	41.3	(3.5)	34.0	(3.4)	38.1	(3.3)	42.5	(4.0)	32.1	(3.1)	34.6	(3.3)	33.9	(3.0)
Republik Korea (Südkorea)	31.4	(3.8)	35.0	(3.5)	45.6	(3.6)	31.9	(3.5)	27.4	(3.7)	32.8	(3.5)	32.5	(3.5)	35.3	(3.8)
^{1 2} England	42.7	(7.6)	44.4	(7.6)	49.9	(8.0)	37.8	(7.7)	17.3	(6.4)	28.6	(7.0)	22.8	(6.6)	19.2	(6.3)
Internationaler Mittelwert	35.0	(3.4)	32.7	(3.3)	34.1	(3.3)	36.3	(3.5)	31.5	(3.4)	28.3	(3.2)	33.2	(3.4)	31.1	(3.4)
Zypern	54.9	(5.1)	46.6	(4.7)	53.9	(4.9)	37.7	(4.0)	23.1	(4.2)	15.4	(3.4)	11.7	(3.2)	17.2	(3.6)
^{2 3} USA	39.8	(2.2)	33.7	(2.2)	42.7	(2.4)	37.4	(2.1)	28.9	(2.2)	20.5	(1.7)	29.0	(2.3)	26.0	(2.0)
Australien	34.8	(3.6)	30.9	(3.5)	35.3	(3.3)	42.5	(3.9)	34.4	(3.8)	17.7	(2.7)	29.9	(3.5)	30.1	(3.2)
² Litauen	28.4	(3.6)	25.8	(3.1)	21.8	(3.3)	44.3	(3.9)	34.7	(3.4)	19.5	(2.7)	37.5	(3.7)	36.7	(3.8)
Malta	31.2	(0.3)	33.0	(0.3)	25.0	(0.3)	32.9	(0.3)	18.8	(0.3)	30.4	(0.3)	23.1	(0.3)	25.9	(0.3)
² Serbien	28.3	(3.4)	16.7	(2.9)	21.0	(3.4)	35.1	(3.7)	25.9	(3.6)	19.4	(3.2)	36.3	(3.8)	33.5	(4.0)
^{1 2} Türkei (5. Jgst.)	27.6	(3.4)	28.0	(3.6)	35.5	(3.5)	19.9	(3.0)	31.4	(3.8)	25.1	(3.5)	18.5	(3.2)	20.2	(2.9)
² Spanien	16.3	(2.4)	25.1	(3.2)	16.1	(2.5)	32.9	(3.1)	28.7	(3.3)	19.0	(2.7)	40.3	(3.0)	23.4	(3.4)
^{1 2} Neuseeland	28.8	(3.0)	29.9	(2.8)	28.2	(2.7)	33.6	(3.0)	19.7	(2.4)	13.3	(2.1)	21.4	(2.6)	25.2	(2.9)
² Slowakei	23.7	(3.1)	24.5	(3.1)	44.5	(3.6)	18.5	(2.5)	35.2	(2.6)	11.6	(2.3)	17.1	(2.6)	19.8	(2.7)
VG EU	26.5	(3.4)	22.7	(3.3)	23.2	(3.4)	26.3	(3.5)	21.9	(3.3)	14.8	(2.9)	22.9	(3.3)	20.3	(3.3)
Irland	25.3	(3.6)	22.0	(3.0)	20.8	(3.1)	28.7	(3.8)	21.7	(3.3)	12.2	(2.4)	18.8	(2.8)	28.2	(3.5)
VG OECD	25.0	(3.3)	22.1	(3.2)	22.9	(3.2)	25.8	(3.5)	21.3	(3.2)	14.2	(2.8)	23.5	(3.4)	20.3	(3.2)
³ Nordirland	20.5	(3.3)	25.0	(3.5)	16.6	(3.6)	30.9	(4.2)	25.4	(3.7)	5.3	(2.1)	14.0	(3.1)	26.6	(3.8)
Chile	19.5	(3.5)	12.4	(2.8)	15.1	(3.2)	26.8	(4.4)	16.3	(3.2)	11.4	(3.0)	35.8	(4.8)	23.1	(3.9)
Italien	12.7	(2.5)	18.2	(2.7)	18.3	(3.0)	22.3	(3.5)	19.2	(3.0)	18.0	(3.2)	27.5	(3.4)	19.5	(2.9)
² Portugal	23.5	(2.6)	25.5	(2.8)	15.3	(2.4)	23.8	(3.1)	15.1	(2.4)	9.2	(1.9)	24.0	(3.1)	16.6	(2.4)
³ Japan	40.4	(3.9)	37.2	(3.9)	17.9	(2.9)	11.5	(2.4)	10.7	(2.3)	12.6	(2.8)	13.9	(3.0)	3.9	(1.6)
Tschechische Republik	21.6	(2.9)	14.0	(2.5)	5.8	(1.6)	26.4	(3.3)	13.9	(2.2)	10.4	(2.2)	35.1	(3.4)	19.7	(2.8)
Deutschland	32.6	(3.3)	22.1	(2.9)	20.7	(2.9)	14.3	(3.1)	12.9	(2.6)	11.9	(2.5)	18.9	(3.0)	12.8	(2.8)
² Österreich	40.4	(3.1)	18.6	(2.3)	14.8	(2.3)	24.2	(3.1)	8.7	(1.9)	5.5	(1.9)	15.1	(2.5)	13.9	(2.4)
² Kanada	15.5	(1.6)	13.5	(1.7)	15.1	(1.5)	24.4	(2.3)	17.9	(1.6)	9.0	(1.7)	22.6	(2.1)	19.8	(2.0)
Belgien (Fläm. Gem.)	16.9	(3.0)	11.4	(2.4)	36.6	(3.4)	6.2	(1.8)	9.9	(2.3)	5.7	(1.9)	14.0	(2.7)	21.4	(3.4)
² Schweden	17.3	(3.4)	13.3	(2.8)	10.4	(2.4)	9.5	(2.6)	18.9	(3.4)	15.4	(3.1)	18.7	(3.6)	17.2	(3.3)
Bulgarien	15.4	(2.3)	14.4	(2.8)	19.6	(2.8)	9.2	(2.1)	18.5	(2.7)	14.0	(2.6)	11.5	(2.6)	14.5	(2.6)
³ Dänemark	23.7	(3.8)	15.8	(3.3)	12.9	(3.2)	17.6	(3.5)	16.2	(3.9)	8.2	(2.3)	7.1	(2.2)	10.2	(2.8)
³ Niederlande	8.4	(3.2)	9.0	(3.3)	8.2	(3.2)	23.7	(5.1)	16.1	(4.6)	8.2	(3.4)	17.1	(4.5)	15.4	(4.5)
Frankreich	13.6	(2.4)	15.4	(2.4)	13.2	(2.7)	13.9	(2.4)	9.8	(2.4)	4.2	(1.3)	11.2	(2.2)	11.9	(2.6)
Ungarn	4.4	(1.4)	8.1	(1.8)	1.2	(0.6)	24.1	(2.6)	8.4	(2.4)	2.1	(1.0)	26.3	(3.4)	7.5	(1.7)
Finnland	5.2	(1.5)	7.3	(1.9)	8.1	(1.6)	6.8	(1.9)	10.9	(2.1)	6.8	(1.6)	10.3	(1.8)	7.3	(1.8)
^{1 3} Norwegen (5. Jgst.)	10.5	(2.7)	8.4	(2.4)	5.1	(1.9)	6.6	(2.2)	7.8	(2.5)	5.0	(2.0)	8.8	(2.8)	8.9	(2.3)
Benchmark-Teilnehmer																
² Dubai, VAE	71.4	(1.9)	67.2	(2.2)	74.4	(1.9)	82.2	(2.0)	77.6	(2.0)	75.2	(1.9)	81.0	(2.0)	74.8	(1.4)
Moskau, Russische Föderation	18.0	(3.2)	20.5	(3.4)	38.3	(4.3)	34.6	(4.4)	50.0	(4.4)	27.9	(3.6)	31.0	(4.2)	41.7	(4.2)
Madrid, Spanien	22.2	(2.9)	28.2	(3.7)	18.7	(2.9)	36.3	(3.7)	33.7	(4.3)	15.1	(2.5)	44.5	(3.9)	22.1	(3.4)
Ontario, Kanada	9.6	(2.6)	8.8	(2.3)	18.9	(3.0)	34.7	(4.8)	20.5	(3.2)	8.5	(3.1)	33.9	(4.3)	23.0	(3.6)
³ Québec, Kanada	18.5	(3.4)	16.7	(3.6)	3.0	(1.5)	4.9	(1.7)	12.4	(3.0)	6.9	(2.4)	4.6	(1.5)	9.2	(2.7)

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

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

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

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

Stärkung kritischen Denkens oder Problemlösens und zur *fächerübergreifenden Gestaltung des naturwissenschaftlichen Sachunterrichts* wurden von Lehrkräften von rund 14 beziehungsweise 13 Prozent der Schülerinnen und Schüler besucht. Mit rund 13 Prozent liegt der Anteil in *Integration von Informationstechnologien* etwas höher als in der Mathematik, aber auch hier deutlich unter dem europäischen Durchschnitt (21.9%).

7.6 Zusammenfassung

Hinsichtlich der Verbreitung von privatem Nachhilfeunterricht lässt sich in TIMSS 2019 ein ähnliches Fazit wie schon in TIMSS 2015 ziehen: In Deutschland erhält nicht einmal jedes zehnte Kind am Ende der Grundschulzeit privaten Nachhilfeunterricht in Mathematik. Im Fach Sachunterricht spielt der Nachhilfeunterricht nahezu keine Rolle. Im Trendvergleich zeigt sich seit IGLU 2006, dass die Nutzung außerschulischer Förderung in der Grundschule in Mathematik sowie im Lesen und in Rechtschreibung aus Elternperspektive leicht abgenommen hat, im Vergleich von TIMSS 2015 und 2019 aber stabil geblieben ist.

Privater Nachhilfeunterricht in Mathematik wird in allererster Linie von Schülerinnen und Schülern mit mangelhaften oder nur ausreichenden Mathematiknoten genutzt, aber immerhin auch von gut 10 Prozent der Schülerinnen und Schüler mit befriedigenden Leistungen. Das Aufwachsen in einer armutsgefährdeten Familie geht nicht mit einer geringeren Chance einher, überhaupt privaten Nachhilfeunterricht zu erhalten. Offenbleiben muss aber, ob Schülerinnen und Schüler aus armutsgefährdeten Familien weniger intensiven oder weniger effektiven Nachhilfeunterricht erhalten.

Die Schülerinnen und Schüler mit vergleichsweise schwachen Leistungen in Mathematik (unterhalb von Kompetenzstufe III) erhielten weit überwiegend von ihren Erziehungsberechtigten Hilfe beim Mathematiklernen, während nur gut ein Fünftel privaten Nachhilfeunterricht in Mathematik in Anspruch nahm. Damit erhielt ein Achtel dieser Kinder keine private Unterstützung, sodass wenigstens für diese Schülerinnen- und Schülergruppe schulische Unterstützungsangebote dringend geboten sind.

Privater Nachhilfeunterricht ist am Ende der Grundschulzeit weit überwiegend tatsächlich Privatsache: Wenn Schülerinnen und Schüler privaten Nachhilfeunterricht erhalten, ist nur etwa ein Viertel der Lehrkräfte über diesen Unterricht informiert, sodass die Chancen, die sich aus einer direkten Abstimmung der Förderung zwischen Nachhilfe- und Fachlehrkraft ergeben könnten, nicht genutzt werden.

In Bezug auf das Fortbildungsverhalten von Lehrkräften lässt sich feststellen, dass Grundschullehrerinnen und -lehrer durchaus regelmäßig fachspezifische Fortbildungsveranstaltungen wahrnehmen. Im internationalen Vergleich liegt die Teilnahmeintensität allerdings eher unterhalb des Niveaus des europäischen Durchschnitts. Zudem zeigt sich ein stärkeres Interesse an fachlichen, didaktischen und lehrplanbezogenen Fortbildungsangeboten vor jenen, die diagnostische Aspekte in den Fokus rücken. Fortbildungen zu Aspekten der individuellen Förderung von Schülerinnen und Schülern werden stärker von Mathematik- als von Sachunterrichtslehrkräften nachgefragt. Bemerkenswert ist, dass Lehrkräfte an Grundschulen in Deutschland im internationalen Vergleich deutlich seltener Veranstaltungen besuchen, die den Einsatz digitaler Medien im Unterricht fo-

kussieren. Zwar kann mit den vorliegenden Daten nicht beantwortet werden, worin die Gründe liegen – trotzdem ist dieser Befund angesichts der wachsenden Bedeutung von digitalen Medien in allen Gesellschaftsbereichen und ihres Potenzials zur Förderung von Schülerinnen und Schülern als kritisch anzusehen.

Literatur

- Bray, M. (2009). *Confronting the shadow education system. What government policies for what private tutoring?* Paris: UNESCO/IIEP.
- Bray, M. (2020). Shadow education in Europe. Growing prevalence, underlying forces, and policy implications. *ECNU Review of Education*, 64. <http://doi.org/10.1177/2096531119890142>
- Fussangel, K., Rürup, M. & Gräsel, C. (2016). Lehrerfortbildung als Unterstützungssystem. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (2. Aufl., S. 327–354). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-531-18942-0_13
- Guill, K. (2020). Qualität der Hausaufgabenhilfe in Elternhaus, Schule und Nachhilfeunterricht. [Themenheft] *Psychologie in Erziehung und Unterricht*, 67(4).
- Guill, K. & Lintorf, K. (2019). Private tutoring when stakes are high: Insights from the transition from primary to secondary school in Germany. *International Journal of Educational Development*, 65, 172–182. <http://doi.org/10.1016/j.ijedudev.2018.08.001>
- Guill, K., Lüdtke, O. & Köller, O. (2020). Assessing the instructional quality of private tutoring and effects on student outcomes: Analyses from the German National Educational Panel Study. *British Journal of Educational Psychology*, 90, 282–300. <http://doi.org/10.1111/bjep.12281>
- Guill, K. & Wendt, H. (2016). Außerschulischer Nachhilfeunterricht am Ende der Grundschulzeit. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.). *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 247–256). Münster: Waxmann.
- Hille, A., Spieß, C. K. & Staneva, M. (2016). More and more students, especially those from middle-income households, are using private tutoring. *DIW Economic Bulletin*, 6. Verfügbar unter: https://www.diw.de/documents/publikationen/73/diw_01.c.526666.de/diw_econ_bull_2016-06.pdf
- Klemm, K. & Hollenbach-Biele, N. (2016). *Nachhilfeunterricht in Deutschland: Ausmaß – Wirkung – Kosten*. Gütersloh: Bertelsmann Stiftung.
- KMK (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland). (2014). *Standards für die Lehrerbildung: Bildungswissenschaften*. Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 12.06.2014. Bonn: KMK.
- Krüger, R. (1977). Nachhilfe – Chance oder Skandal? 17 Antworten auf Fragen zu einem vernachlässigten Problem. *Die Deutsche Schule*, 69(9), 545–558.
- Park, H., Buchmann, C., Choi, J. & Merry, J. J. (2016). Learning beyond the school walls: Trends and implications. *Annual Review of Sociology*, 42(1), 231–252. <http://doi.org/10.1146/annurev-soc-081715-074341>
- Rudolph, M. (2002). *Nachhilfe – gekaufte Bildung? Empirische Untersuchung zur Kritik der außerschulischen Lernbegleitung. Eine Erhebung bei Eltern, LehrerInnen und Nachhilfeinstituten*. Bad Heilbrunn: Klinkhardt.
- Wild, E. (2020). Hausarbeiten – ein überkommenes oder hoch aktuelles Konzept? Ein Diskussionsbeitrag. *Psychologie in Erziehung und Unterricht*, 67(4), 314–325. <http://doi.org/10.2378/peu2020.art31d>

Kapitel 8

Geschlechterunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen

Sonja Nonte, Ricarda Steinmayr und Luise A. Scholz

8.1 Einleitung

Vor dem Hintergrund der gesellschaftlichen Bedeutung von Fächern in den Bereichen Mathematik, Informatik, Naturwissenschaften und Technik, den sogenannten MINT-Fächern, spielen mathematische und naturwissenschaftliche Kompetenzen von Mädchen und Jungen eine große Rolle. Da es aber nach wie vor mehr Männer als Frauen und einen wachsenden Fachkräftemangel in MINT-Berufen gibt (Bundesagentur für Arbeit, 2019), sind Geschlechterunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen schon seit langer Zeit Gegenstand wissenschaftlicher Untersuchungen. Dabei zeigt sich konsistent, dass Kompetenzunterschiede zwischen Mädchen und Jungen in Mathematik und den Naturwissenschaften im Durchschnitt eher gering sind (Reiss, Weis, Klieme & Köller, 2019), sodass Kompetenzunterschiede als alleinige Erklärungsvariable für das geschlechterspezifische Wahlverhalten etwa von schulischen Fächern oder des Berufes im MINT-Bereich nicht ausreichend sind.

Als weitere erklärungs mächtige Variablen für Geschlechterunterschiede in Leistungen haben sich sowohl die Selbsteinschätzung der eigenen Fähigkeiten in der jeweiligen Domäne als auch der Wert, der diesem Bereich beigemessen wird, erwiesen (z.B. Lazarides & Lauermann, 2019). Somit sind diese beiden Aspekte neben den reinen Kompetenzunterschieden ebenfalls Gegenstand der Forschung zu Geschlechterunterschieden im Bereich Mathematik und Naturwissenschaften. Hier zeigen sich sowohl im Bereich Mathematik als auch in den Naturwissenschaften deutlichere Geschlechterunterschiede als in den Kompetenzen (z.B. Steinmayr & Spinath, 2010; Wendt, Steinmayr & Kasper, 2016). Trotz ihrer leichten Kompetenz- und (stärkeren) Motivationsvorteile im MINT-Bereich schneiden Jungen in fast allen anderen Leistungsindikatoren schulischer Bildung (z.B. Lesekompetenzen, Noten, Abiturquote) schlechter ab (siehe Abschnitt 8.2).

Es existieren verschiedene Theorien und Annahmen darüber, wie diese unterschiedlichen Befunde erklärt werden können. Während einige populäre Annahmen, wie die „Feminisierung der Schule“ (Helbig, 2010, S. 96) in Form von überwiegend weiblichen Lehrkräften und das damit einhergehende Fehlen männlicher Vorbilder für die Jungen, mittlerweile als empirisch widerlegt gelten (Hannover & Kessels, 2011; Marsh, Martin & Cheng, 2008; Spinath, Eckert & Steinmayr, 2014), sind Erklärungsansätze, die sowohl individuelle als auch soziale Aspekte und Prozesse anführen, sehr viel besser belegt. Beispielsweise erklärt das von Kessels, Heyder, Latsch und Hannover (2014) entwickelte *Interest as Identity Regulation Model* (IIRM; siehe Abschnitt 8.3) sowohl die positiver ausgeprägten Selbstkonzepte und Einstellungen von Jungen im MINT-Bereich als auch deren schlechtere Leistungen im schulischen Bereich insgesamt. Somit tragen solche Modelle auch dazu bei, die Ursachen und Gründe für die niedrigeren Selbstkonzepte und subjektiven Werte von Mädchen im Bereich Mathematik und Naturwissenschaften unabhängig von ihren tatsächlichen Kompetenzen zu beleuchten, die das Wahlverhalten von Mädchen und Frauen im MINT-Bereich negativ beeinflussen können.

Im Folgenden werden in diesem Kapitel zunächst Geschlechterunterschiede im Bildungserfolg allgemein sowie in schulischen Kompetenzen beleuchtet, mit Fokus auf mathematische und naturwissenschaftliche Kompetenzen. Anschließend werden verschiedene Erklärungsansätze mit einem Fokus auf sozial-kognitive Theorien erläutert, die die Rolle von Fähigkeitsselbstkonzepten und Einstellungen für geschlechterrollenspezifische Wahlentscheidungen beschreiben. Dann folgt eine Darstellung der aktuellen Ergebnisse zu Geschlechterunterschieden in mathematischen und naturwissenschaftlichen Kompetenzen sowie Trendanalysen, die eine Entwicklung der geschlechterspezifischen Ergebnisse in den bei der *Trends in International Mathematics and Science Study* (TIMSS) gemessenen Kompetenzen seit 2007 dokumentieren. Aufgrund ihrer hohen Bedeutung für geschlechterspezifische Wahl- und Karriereentscheidungen werden anschließend die Geschlechterunterschiede in der Einschätzung der eigenen Fähigkeiten in Mathematik und Naturwissenschaften sowie die diesem Bereich beigemessenen Werte beziehungsweise Einstellungen beschrieben. Deren Entwicklung ist ebenfalls Gegenstand des Kapitels, das abschließend die Ergebnisse zusammenfasst.

8.2 Geschlechterunterschiede in schulischen Leistungen

Während in Mathematik und Naturwissenschaften Geschlechterunterschiede überwiegend zugunsten von Jungen beziehungsweise Männern ausfallen, übertreffen Mädchen beziehungsweise Frauen diese in fast allen anderen schulischen Bereichen (Spinath et al., 2014; Statistisches Bundesamt, 2019). Mädchen besuchen häufiger als Jungen das Gymnasium, während Jungen an der Haupt- und Förderschule überrepräsentiert sind. Mehr Mädchen schließen die Schule mit der Allgemeinen (Fach-)Hochschulreife ab, während mehr Jungen die Schule ohne Schulabschluss verlassen. Ebenso haben Mädchen in fast allen Staaten bessere Noten als Jungen, was ebenso für die Gesamt- wie für die Fachnoten gilt (O’Dea, Lagisz, Jennions & Nakagawa, 2018; Voyer & Voyer, 2014) und so – teilweise – die Geschlechterunterschiede in der schulischen Bildungsbeteiligung und den Bildungserfolgen erklärt. Während sich in den weiterführenden Schulen teilweise starke Geschlechterunterschiede im MINT-Bereich zeigen, sind die Unterschiede in der Grundschule, oftmals zugunsten der

Jungen, nicht signifikant oder sehr gering (z.B. Helbig, 2010; Steinmayr, Michels & Weidinger, 2017; Wendt, Bos et al., 2016). Hingegen erzielen Mädchen im Sach- und Deutschunterricht sowohl in der Grundschule als auch in den weiterführenden Schulen bessere Noten, wobei diese Geschlechterunterschiede deutlicher ausfallen als die Vorteile der Jungen in der Mathematiknote (Steinmayr et al., 2017; Wendt, Bos et al., 2016). Dieser Vorteil von Mädchen im sprachlichen Bereich spiegelt sich auch in der Lesekompetenz wider. In internationalen Schulleistungsuntersuchungen schneiden Mädchen in der Grundschule besser ab als Jungen, wobei dieser Unterschied (Cohens $d < .2$) die Größe eines kleinen praktischen Effektes erreichte (Effektstärke Cohens d als Maß für die praktische Relevanz des Effektes; kleiner Effekt: $.2 \leq d < .5$; mittlerer Effekt: $.5 \leq d < .8$; großer Effekt: $d \geq .8$.) (McElvany, Kessels, Schwabe & Kasper, 2017). Geschlechterunterschiede in der Lesekompetenz bei Grundschülerinnen und Grundschülern scheinen im zeitlichen Trend relativ stabil zu sein, was auch den Ergebnissen des IQB-Bildungstrends 2016 entspricht (Stanat, Schipolowski, Rjosk, Weirich & Haag, 2017). An der weiterführenden Schule hingegen scheinen sich die Leistungsunterschiede in der Lesekompetenz zugunsten der Mädchen zu vergrößern. Sie sind zudem größer als die Geschlechterunterschiede in den Domänen Mathematik und Naturwissenschaften.

Die in standardisierten Schulleistungstests gemessenen mathematischen Kompetenzen von Mädchen liegen im Grundschulalter im Mittel nur geringfügig unter denen der Jungen, wobei sich diese Unterschiede von TIMSS 2007 bis TIMSS 2015 in der Grundschule in Deutschland auf 5 Punkte verringerten (z.B. Bergold, Wendt, Kasper & Steinmayr, 2017; Wendt, Bos et al., 2016). Jedoch war letzteres Ergebnis alleinig auf eine Leistungsverschlechterung der Jungen zurückzuführen. Ein ähnliches Bild zeigt sich bei TIMSS 2007 bis TIMSS 2015 in den naturwissenschaftlichen Kompetenzen bei Grundschülerinnen und Grundschülern. In der letzten TIMS-Studie war der Unterschied nicht mehr statistisch signifikant. Die Reduktion des Geschlechterunterschieds bei TIMSS 2015 konnte zum einen durch etwas bessere Leistungen der Mädchen und zum anderen durch etwas schlechtere Leistungen der Jungen erklärt werden und war im Vergleich zu TIMSS 2007 signifikant (Wendt, Bos et al., 2016). Ein ähnliches Bild zeigt sich auch bei internationalen Schulleistungsuntersuchungen an weiterführenden Schulen in Deutschland. Auch hier reduzierte sich der leichte Vorsprung der Jungen in den mathematischen und naturwissenschaftlichen Kompetenzen in den letzten Jahren (OECD, 2016; Reinhold, Reiss, Diedrich, Hofer & Heinze, 2019; Schiepe-Tiska, Rönnebecke & Neumann, 2019). Zusammenfassend zeigen sich somit die bei anderen Leistungsindikatoren gefundenen Geschlechterunterschiede zugunsten der Mädchen nur in der Lesekompetenz, wobei diese in der Grundschule geringer ausfallen als an der weiterführenden Schule. Hingegen waren die in früheren internationalen Schulleistungsuntersuchungen gefundenen Geschlechterunterschiede in Mathematik und den Naturwissenschaften zugunsten der Jungen in den aktuellsten Studien nur noch sehr gering oder nicht mehr signifikant.

Bei den bisher berichteten Geschlechterunterschieden handelte es sich ausschließlich um Unterschiede im Mittelwert. Für einen differenzierteren Einblick in die Geschlechterunterschiede in den Kompetenzen und mögliche Erklärungen für eine Über- oder Unterrepräsentation von Gruppen müssen neben den Mittelwertunterschieden noch Geschlechterunterschiede in der Streuung der Testwerte betrachtet werden. Wie bei Wendt, Bos et al. (2016) dargestellt, kann ein Unterschied im Mittelwert zwischen den beiden Gruppen dadurch zustan-

de kommen, dass sich Mädchen und Jungen auf allen Kompetenzstufen unterscheiden oder aber dadurch, dass viel mehr Jungen als Mädchen die ausgeprägt niedrigsten Kompetenzen erreichen, während die Geschlechterverteilung im oberen Kompetenzbereich gleich verteilt ist. Darüber hinaus kann es auch sein, dass sich kein Unterschied im Mittelwert, aber durchaus in der Verteilung der Testwerte zwischen zwei Gruppen zeigt. Sowohl in Intelligenztests als auch in Kompetenztests in den Domänen Lesen, Mathematik und Naturwissenschaften finden sich größere Unterschiede in der Gruppe der Jungen als in der Gruppe der Mädchen (Arden & Plomin, 2006; Baye & Monseur, 2016; Gray et al., 2019; Lindberg, Hyde, Petersen & Linn, 2010; Reilly, Neumann & Andrews, 2015). Dieses Phänomen wird in der Literatur als *greater male variation hypothesis* beschrieben (z. B. Gray et al., 2019). Da Jungen in Mathematik im Durchschnitt leicht höhere Mittelwerte haben, zeigt sich, dass Jungen in den oberen Kompetenzbereichen in dieser Domäne überrepräsentiert sind, während sich in den mittleren und unteren Kompetenzbereichen keine Geschlechterunterschiede oder eine leichte Überrepräsentation der Mädchen finden lässt (z. B. Hedges & Nowell, 1995; Machin & Pekkarinen, 2008; Nowell & Hedges, 1998). Je strenger das *Cut-Off-Kriterium* für die oberen Testwerte gewählt wird (also die besten 5%, 2% oder 1% betrachtet werden), desto mehr verschiebt sich das Zahlenverhältnis zugunsten der Jungen. Ähnliche Ergebnisse finden sich für Naturwissenschaften (z. B. Hedges & Nowell, 1995; Nowell & Hedges, 1998; Reilly et al., 2015). Das beschriebene Muster zeigt sich auch in TIMSS 2011 und TIMSS 2015, wo Jungen in Deutschland in den beiden oberen mathematischen und naturwissenschaftlichen Kompetenzstufen, Mädchen dagegen in der mittleren und den beiden unteren Kompetenzstufen überrepräsentiert waren. Ein ähnliches Befundmuster findet sich auch bei Geschlechterunterschieden in Schulnoten, bei denen die Unterschiede (Varianz) in der Gruppe der Jungen größer war als in der Gruppe der Mädchen (O'Dea et al., 2018). In MINT-Fächern waren die Unterschiede in der Varianz zwischen den beiden Gruppen jedoch kleiner als bei Nicht-MINT-Fächern. Während in MINT-Fächern 94.2 Prozent der Noten zwischen den beiden Gruppen überlappten, war das nur für 88.2 Prozent der Noten in den Nicht-MINT-Fächern der Fall. Bei den Schülerinnen und Schülern, die zu den besten 10 Prozent in den jeweiligen Fachnoten gehörten, waren in den MINT-Fächern ähnlich viele Jungen wie Mädchen vertreten, während hier die Mädchen in den Nicht-MINT-Fächern und in der Gesamtnote überrepräsentiert waren (O'Dea et al., 2018).

Für eine Einordnung der Ergebnisse muss darauf hingewiesen werden, dass die Unterschiede in den Leistungen innerhalb der Geschlechtergruppen in der Regel größer sind als die Leistungsunterschiede zwischen Mädchen und Jungen, wie sich auch an den oben berichteten großen Überlappungen der Verteilung der gruppenspezifischen Werten zeigte. Darüber hinaus vergrößern (z. B. numerische Kompetenzen) beziehungsweise verkleinern (z. B. Lesekompetenz) sich Geschlechterunterschiede häufig mit zunehmender Selektivität der Stichproben (z. B. Studien mit Studierenden) (Hyde, 2005), da im Laufe der akademischen Ausbildung nur diejenigen im System verbleiben, die über entsprechend gute Kompetenzen und eine ausreichend hohe Motivation verfügen und somit eher das obere Ende der Verteilung repräsentieren (z. B. Baye & Monseur, 2016).

8.3 Erklärungsansätze zu Geschlechterunterschieden in schulischen Leistungen

Zur Erklärung von Geschlechterunterschieden in unterschiedlichen Domänen wurde in der Literatur bereits eine Vielzahl an Ursachen diskutiert. Eine vollständige Darstellung und differenzierte Diskussion der verschiedenen Erklärungsansätze von Geschlechterunterschieden im Leistungsbereich ist im vorliegenden Band nicht möglich. Zur Erklärung von Geschlechterunterschieden im MINT-Bereich werden in der Literatur biologische Ansätze (Gene, Hormone, hirnpfysiologische Merkmale), Fähigkeits- und Persönlichkeitsunterschiede in verschiedenen Bereichen (Intelligenz, logisches Denken, Problemlösen, bereichsspezifische Fähigkeit wie mathematische oder figurale Begabungen, Konzentrationsfähigkeit, Selbstdisziplin, soziale Kompetenz, Kommunikationsfähigkeit etc.), Einstellungen, Interessen und andere motivationale Konstrukte sowie ihr Verhalten in und für die Schule herangezogen. Auch das distale (z.B. Gesellschaft) und das proximale Umfeld (z.B. Erziehungsberechtigte, Freunde) und damit verbundene vermeintlich geschlechtertypische Entwicklungen (z.B. des Geschlechterrollenselbstkonzepts), Stereotype sowie der damit verbundene gruppenspezifische *Stereotype Threat* werden als Gründe diskutiert (z.B. Ceci, Williams & Barnett, 2009; Kessels & Heyder, 2018). Unter *Stereotype Threat* versteht man generalisierte Überzeugungen über die Eigenschaften der Mitglieder einer Gruppe, die sich auf die Informationsverarbeitung bei sozialen Interaktionen auswirken können (Fiske & Neuberg, 1990; Holder & Kessels, 2017). Da bei der Erklärung von Leistungsunterschieden motivationale Variablen eine große Rolle spielen (Lauermann, Meißner & Steinmayr, 2020; Steinmayr & Spinath, 2009; Steinmayr, Weidinger, Schwinger & Spinath, 2019) und diese darüber hinaus einen großen Beitrag zur Erklärung von Geschlechterunterschieden leisten (Spinath et al., 2014; Steinmayr & Spinath, 2008), sollen die motivationalen Variablen im Folgenden näher beleuchtet werden.

Theoretisch lassen sich motivationale Variablen gut in das vielfach empirisch bewährte Erwartungs-Wert-Modell von Eccles und Wigfield (2002) einordnen. In der Erwartungs-Wert-Theorie werden spezifische Annahmen dazu gemacht, wie Geschlechterunterschiede in Leistungsindikatoren zustande kommen (u. a. Wigfield, Muenks & Rosenzweig, 2016). Als proximale Variablen zur Erklärung von Leistungsunterschieden dienen Erfolgserwartungen und Werte. Erfolgserwartungen werden häufig über das Fähigkeitsselbstkonzept erfasst (Retelsdorf, Köller & Möller, 2011), da das Fähigkeitsselbstkonzept empirisch nicht von Erfolgserwartungen zu trennen ist (Eccles & Wigfield, 1995). Somit sind die in TIMSS bei allen Erhebungen erfassten Fähigkeitsselbstkonzepte als Erfolgserwartungsvariablen zu interpretieren. Fähigkeitsselbstkonzepte sind die subjektiven Bewertungen beziehungsweise kognitiven Repräsentationen eigener Fähigkeiten in schulischen Leistungsbereichen (Schöne, Dickhäuser, Spinath & Stiensmeier-Pelster, 2012). Der einem Bereich oder einer Domäne beigemessene Wert setzt sich zusammen aus dem intrinsischen Wert, der persönlichen Wichtigkeit, dem beurteilten Nutzen einer Aufgabe für zukünftige Pläne und Ziele und, negativ, den mit einer Tätigkeit verbundenen Kosten (Steinmayr & Spinath, 2010). Den höchsten Zusammenhang mit der Schulleistung weist die intrinsische Werte-Komponente auf (z.B. Steinmayr & Spinath, 2010), die konzeptuell identisch mit intrinsischer Motivation ist. Unter intrinsischer Motivation versteht man, dass beispielsweise Schülerinnen und Schüler den Mathematik- oder den Sachunterricht mögen, das jeweilige Fach interessant finden und eine

Auseinandersetzung mit dem Fach eher zu positiven Emotionen und in der Folge zu einer tieferen Bearbeitung der entsprechenden Inhalte führt (Wigfield et al., 2015). Hierbei wird davon ausgegangen, dass Interessen oder intrinsische Motivation relativ überdauernde Merkmale sind. Aufgrund ihrer hohen Bedeutung für das Leistungsverhalten von Schülerinnen und Schülern wurde die intrinsische Motivation schon bei TIMSS 2007, 2011 und 2015 unter dem Begriff *fachspezifische positive Einstellungen* erfasst und wird auch in TIMSS 2019 wieder erhoben.

Dem Eccles-Modell zufolge werden Werte und Fähigkeitsselbstkonzepte der Schülerinnen und Schüler direkt von der persönlichen Einschätzung der Einstellungen und Erwartungen anderer Personen (z.B. Wahrnehmung von Geschlechterrollen und Stereotypen) sowie von den subjektiven Interpretationen vorheriger Leistungsergebnisse der Person (z.B. Kausalattributionen: „Da hatte ich nur Glück.“ vs. „Das konnte ich richtig gut.“) beeinflusst. Erstere hängen direkt von dem kulturellen Milieu ab, in dem die Person lebt (z.B. soziale Herkunft, gesellschaftliche Stereotype), letztere ergeben sich aus früheren Leistungserfahrungen. Zudem hängen sowohl die Interpretationen vorheriger Leistungsergebnisse als auch die persönlichen Einschätzungen der Einstellungen und Erwartungen anderer Personen von den Überzeugungen sowie dem Verhalten wichtiger Bezugspersonen (z.B. Erziehungsberechtigte und Lehrkräfte) und den Eigenschaften einer Person ab, darunter auch das Geschlecht. Tatsächlich tragen sowohl Fähigkeitsselbstkonzepte als auch Werte beziehungsweise Einstellungen zur Erklärung von Geschlechterunterschieden in schulischen Leistungsindikatoren bei (Steinmayr & Spinath, 2008). Jungen berichteten bei TIMSS 2015 etwas positivere Einstellungen in Bezug auf Mathematik als Mädchen. Die Effektstärke ist zwar größer als für den Geschlechterunterschied in den mathematischen Kompetenzen, hat sich aber seit 2007 signifikant verringert, was allein auf eine positivere Einstellung der Mädchen im Trend zurückzuführen war (Wendt, Bos et al., 2016). In Studien mit Schülerinnen und Schülern der gymnasialen Oberstufe zeigen sich keine Geschlechterunterschiede im mathematischen Interesse (Steinmayr & Spinath, 2010). Trotz der geringen objektiven Kompetenzunterschiede war das mathematische Fähigkeitsselbstkonzept der Jungen auch bei TIMSS 2015 deutlich höher als das der Mädchen ($d = .36$), ein Befund, der sich auch bei älteren Schülerinnen und Schülern zeigt (z.B. Steinmayr & Spinath, 2008). Im Trend zeigte sich eine negative Entwicklung der mathematischen Fähigkeitsselbstkonzepte der Mädchen. Für die Naturwissenschaften zeigten sich diese Unterschiede nicht; sowohl bei den Einstellungen als auch beim Fähigkeitsselbstkonzept in Bezug auf den Sachunterricht unterschieden sich Mädchen und Jungen nicht. Zudem war kein klarer Trend in die eine oder andere Richtung zu erkennen. Bei älteren Schülerinnen und Schülern in Deutschland finden sich teilweise große Geschlechterunterschiede für das Fähigkeitsselbstkonzept und kleinere für subjektive Werte im MINT-Bereich (Hannover & Kessels, 2002; Steinmayr & Spinath, 2010).

Warum die Geschlechterunterschiede in den Einstellungen und Fähigkeitsselbstkonzepten von den in den objektiv gemessenen Kompetenzen abweichen können, verdeutlicht eine aktuelle Studie von Steinmayr, Weidinger, Heyder und Bergold (2019). Anhand einer repräsentativen Stichprobe von 738 Viertklässlerinnen und Viertklässlern in Nordrhein-Westfalen zeigt sich, dass Geschlechterunterschiede im mathematischen Fähigkeitsselbstkonzept nicht durch entsprechende Unterschiede in mathematischen Kompetenzen erklärt werden können. Sie können jedoch auf Geschlechterunterschiede zugunsten der

Jungen in den Einstellungen von Erziehungsberechtigten und Lehrkräften bezüglich der Begabung der Kinder in Mathematik und in Mathematiknoten zurückgeführt werden. Ein weiterer Einflussfaktor nach dem Eccles-Modell sind gesellschaftliche Stereotype, die sich ebenfalls auf unsere Werte und Fähigkeits-selbstkonzepte auswirken. Das *Interest as Identity Regulation Model* (Kessels et al., 2014) postuliert, dass eine Diskrepanz zwischen der Stereotypisierung eines Fachs (z.B. Mathematik als eher männliche Domäne) und der eigenen Geschlechterrollenidentität das Interesse und Engagement für dieses Fach negativ beeinflussen. Darüber hinaus spielen dem Modell zufolge soziale Informationen wie Stereotype bei der Entwicklung des eigenen Selbstkonzepts eine große Rolle. Von Schülerinnen und Schülern geteilte Stereotype in Bezug auf Mathematik umfassen größere Begabungen, Fähigkeiten und auch größeres Interesse von Jungen als von Mädchen und finden sich schon bei Kindern im Grundschulalter (Steffens, Jelenec & Noack, 2010). Diese Stereotype werden bei der Einschätzung der eigenen Fähigkeiten in Abhängigkeit von der eigenen Geschlechterrollenidentität übernommen, sodass Mädchen sich unabhängig von ihren eigentlichen Kompetenzen als weniger fähig und interessiert in Mathematik beurteilen, als Jungen das auch äußern.

Eine weitere Erklärung für die vor allem im mathematischen Fähigkeitsselbstkonzept ausgeprägten Geschlechterunterschiede liefert das *Internal/External Frame of Reference Model* von Marsh (1986). Demzufolge wird das Fähigkeitsselbstkonzept nicht nur durch soziale Vergleiche beispielsweise in Form von Noten („In Mathe bin ich besser als Peter“), sondern auch durch interne Vergleichsprozesse („In Mathe bin ich schlechter als in Deutsch“) beeinflusst. Letztere wirken sich negativ auf die Fähigkeitsselbstkonzepte in der Domäne aus, die man selber als die schwächere einschätzt. Da Mädchen vor allem im muttersprachlichen Bereich sowohl in Form von Noten als auch in objektiven Kompetenztests gute Leistungen erbringen, ist bei ihnen die Wahrscheinlichkeit höher, dass sie in diesem Bereich ihre Leistungen besser einschätzen als beispielsweise im Bereich Mathematik, sodass ihr Fähigkeitsselbstkonzept in Mathematik dadurch geringer wird. Umgekehrt haben Jungen im internalen Vergleich mit höherer Wahrscheinlichkeit häufiger bessere Leistungen in Mathematik als in Deutsch, was sich dann wiederum negativ auf ihr Fähigkeitsselbstkonzept in Deutsch auswirkt (Heyder, Kessels & Steinmayr, 2017; Parker, Nagy, Trautwein & Lüdtke, 2014).

Die somit empirisch und theoretisch belegten Geschlechterunterschiede in Fähigkeitsselbstkonzepten und Werten im MINT-Bereich tragen auch zur Erklärung von Geschlechterunterschieden in schulischen Leistungen und Wahlen bei. Das Fähigkeitsselbstkonzept ist auch nach Kontrolle vorangegangener Leistung ein guter Prädiktor von verschiedenen Leistungsindikatoren wie Noten (z.B. Steinmayr & Spinath, 2009) oder standardisierten Schulleistungstests (z.B. Marsh, Trautwein, Lüdtke, Köller & Baumert, 2005) und trug in verschiedenen Studien zur Erklärung von Geschlechterunterschieden in Noten und Schulleistungstests in Mathematik und anderen Fächern bei (Brehl, Wendt & Bos, 2012; Eccles, 1994; Steinmayr & Spinath, 2008). Ähnlich wie das Fähigkeitsselbstkonzept erklären Maße zur Erfassung der intrinsischen Motivation Geschlechterunterschiede in der Schulleistung (z.B. Steinmayr & Spinath, 2008). Während die Effekte von intrinsischer Motivation auf die Schulleistung eher gering sind (Weidinger, Spinath & Steinmayr, 2020; Weidinger, Steinmayr & Spinath, 2017), zeigen sich größere Effekte auf die Kurswahl (z.B. Steinmayr & Spinath, 2010; Wang, Degol & Ye, 2015). Lazarides und Lauermaun (2019) zeig-

ten dies auch an einer Stichprobe von Schülerinnen und Schülern in Deutschland sowohl für das mathematische Fähigkeitsselbstkonzept als auch für persönliche Werte. Beide Prädiktoren erklärten, warum weniger Mädchen als Jungen Karrierepläne im MINT-Bereich hatten. Da sowohl Kompetenzen als auch Fähigkeitsselbstkonzepte und Werte in Bezug auf ein Fach bereits zum Ende der Grundschulzeit relativ stabil sind (Weidinger, Steinmayr & Spinath, 2019), sind diese Ergebnisse vor allem vor dem Hintergrund des andauernden beziehungsweise sich in Deutschland vergrößernden Fachkräftemangels und dem gleichbleibend geringen Anteil von Frauen im MINT-Bereich relevant (Bundesagentur für Arbeit, 2019; Halpern, 2014). Somit wurden auch bei TIMSS 2019 die intrinsische Motivation als erklärungsmächtigste Werte-Variable in Form von fachspezifisch positiven Einstellungen gegenüber Mathematik und Naturwissenschaften sowie das Fähigkeitsselbstkonzept in diesen Bereichen erfasst. Dadurch ist es beispielsweise möglich zu überprüfen, ob der negative Trend im mathematischen Fähigkeitsselbstkonzept bei den Mädchen anhält oder ob dieser im Laufe der Zeit unter anderem durch schulische und außerschulische Förderangebote verringert oder gar gestoppt werden kann.

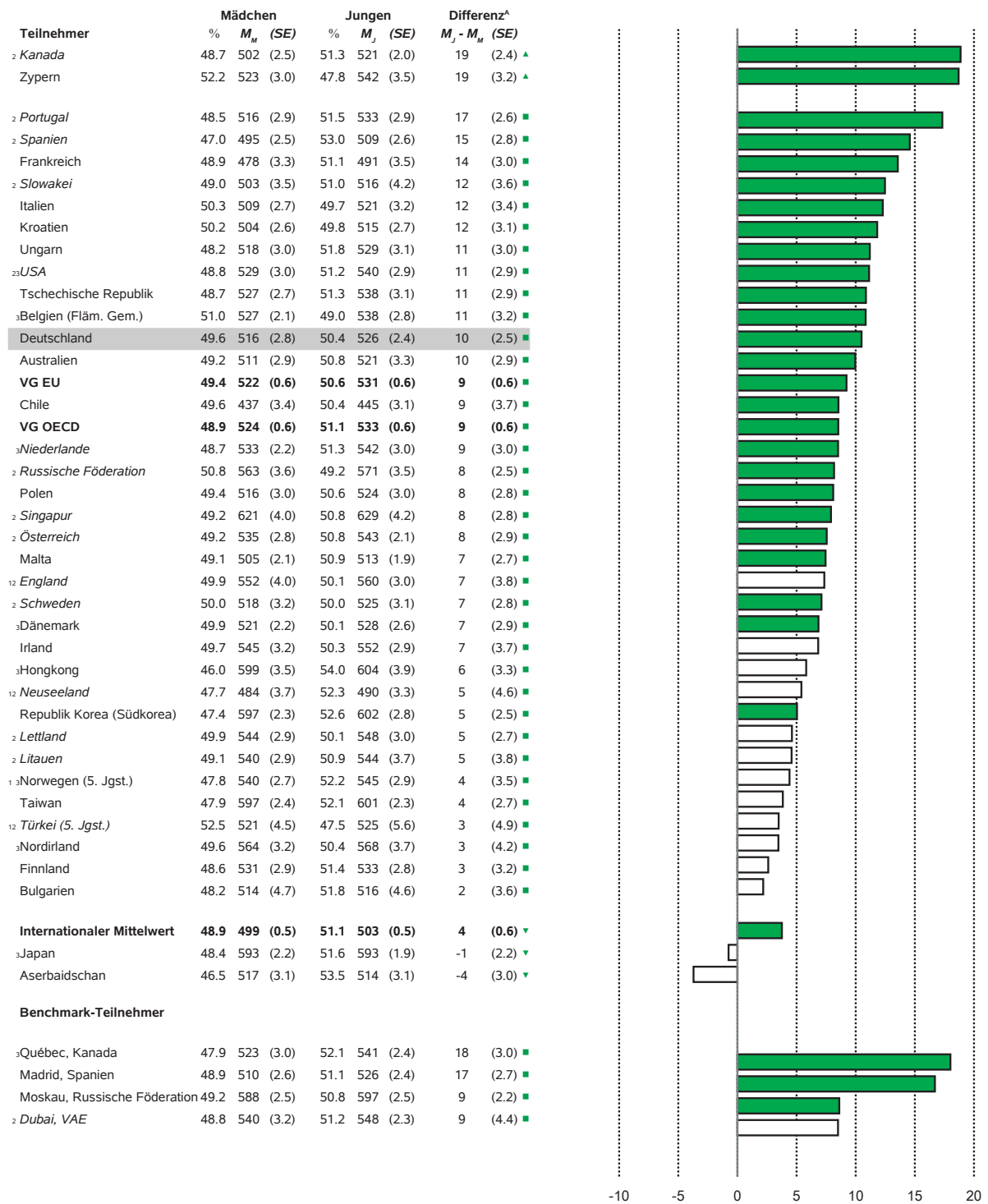
8.4 Ergebnisse zu geschlechterspezifischen Leistungsunterschieden in Mathematik

In diesem und den folgenden Abschnitten werden die Ergebnisse zu den geschlechterspezifischen Unterschieden von Mädchen und Jungen im Hinblick auf ihre mathematischen Leistungen, die Mathematiknoten sowie hinsichtlich der Einstellungen und Motive im Fach Mathematik dokumentiert. Die Struktur ist an die bisherigen nationalen TIMSS-Berichtsbände angelehnt und weist nur geringfügige Veränderungen auf (vgl. u. a. Wendt, Steinmayr & Kasper, 2016). Zunächst erfolgt die Darstellung der Ergebnisse für die Domäne Mathematik und anschließend für Naturwissenschaften.

8.4.1 Geschlechterspezifische Leistungsunterschiede auf der Gesamtskala Mathematik

In Abbildung 8.1 sind die geschlechterbezogenen Leistungsunterschiede für die Gesamtskala Mathematik dargestellt. In der ersten Spalte werden die Teilnehmerstaaten aufgelistet. In der zweiten Spalte wird der prozentuale Anteil (%) an Mädchen und Jungen für jeden Teilnehmerstaat angegeben. Zusätzlich kann der durchschnittliche Skalenmittelwert für die Gesamtskala Mathematik für Mädchen (M_M) und Jungen (M_J) sowie der jeweils zugehörige Standardfehler (SE) der Tabelle entnommen werden. In der letzten Spalte wird die Differenz in den Leistungen von Mädchen und Jungen auf der Gesamtskala Mathematik ($M_J - M_M$) sowie der zugehörige Standardfehler (SE) berichtet. Eine visuelle Einordnung der geschlechterbezogenen Unterschiede ermöglichen die rechts neben der Tabelle dargestellten Balken. Die grün eingefärbten Balken weisen auf einen signifikanten Geschlechterunterschied für die Gesamtskala Mathematik hin. Die Sortierung der Teilnehmerstaaten wurde auf der Grundlage der Mittelwertunterschiede der Mathematikleistung von Mädchen und Jungen vorgenommen. Staaten, in denen der Vorsprung der Jungen gegenüber den Mädchen am größten ist, finden sich weiter oben in der Tabelle, Staaten, in denen der Vorsprung gering ist oder

Abbildung 8.1: Testleistung nach Geschlecht – Gesamtskala Mathematik



□ Keine statistisch signifikanten Unterschiede zwischen Mädchen und Jungen ($p > .05$)

■ Statistisch signifikante Unterschiede zwischen Mädchen und Jungen ($p \leq .05$)

▲ = Differenzwert im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)

▼ = Differenzwert im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)

■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)

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

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

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

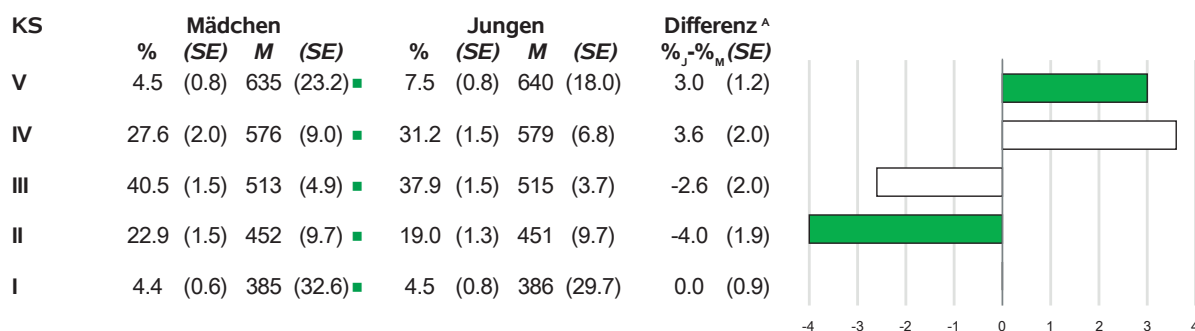
3 = Die Teilnahmequoten auf Schul- und/oder Schülerschülerbene erreichen nicht die internationalen Vorgaben.

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

Mädchen einen Vorsprung in den erreichten Leistungspunkten gegenüber den Jungen haben, finden sich weiter unten in der Liste.

Anhand Abbildung 8.1 wird deutlich, dass in einer überwiegenden Anzahl der Teilnehmerstaaten statistisch signifikante Geschlechterunterschiede bestehen. Am deutlichsten fallen diese Unterschiede in Kanada und Zypern aus. Dort erzielen Jungen durchschnittlich 19 Leistungspunkte mehr auf der Gesamtskala Mathematik als Mädchen. In 14 Staaten (ohne *Benchmark*-Teilnehmer) können keine Leistungsunterschiede zwischen Mädchen und Jungen beobachtet werden. Dies sind unter anderem England, Irland und Finnland; in Japan und in Aserbaidshan weisen Mädchen tendenziell bessere mathematische Leistungen auf als Jungen. Dieser Mittelwertunterschied in den Differenzen ist statistisch jedoch nicht signifikant. Im internationalen Durchschnitt liegt die Differenz der Testleistungen von Mädchen und Jungen ($M_J - M_M$) bei 4 Punkten und ist statistisch signifikant. In den Vergleichsgruppen (VG_{EU} und VG_{OECD}) beträgt die mittlere Leistungs-differenz jeweils 9 Punkte, wobei diese Unterschiede ebenfalls statistisch signifikant sind. In Deutschland kann im Durchschnitt ein Leistungsunterschied von 10 Punkten zwischen Mädchen und Jungen zugunsten der Jungen beobachtet werden. Auch dieser Unterschied ist signifikant und befindet sich leicht oberhalb der Geschlechterdifferenzen in den Vergleichsgruppen. Diese Gesamtübersicht über die von Mädchen und Jungen erreichten Leistungspunkte auf der Gesamtskala Mathematik ermöglicht einen ersten Einblick, wie stark die Geschlechterunterschiede in Deutschland im Vergleich zu denen in anderen Staaten ausgeprägt sind. Von besonderer Bedeutung für das Ableiten von Handlungsempfehlungen ist zudem die differenzierte Betrachtung der Verteilung von Mädchen und Jungen auf den Kompetenzstufen sowie die detaillierte Darstellung der Leistungen nach Inhalts- und kognitiven Anforderungsbereichen. Eine differenzierte Betrachtung zu geschlechterspezifischen Unterschieden unter Berücksichtigung der erreichten Kompetenzstufen ermöglicht Abbildung 8.2.

Abbildung 8.2: Mittlere Testleistungen von Mädchen und Jungen und prozentuale Verteilungen auf die Kompetenzstufen – Mathematik



□ Keine statistisch signifikanten Unterschiede zwischen Mädchen und Jungen ($p > .05$); Linksseitige Ausprägung = Vorsprung der Mädchen; Rechtsseitige Ausprägung = Vorsprung der Jungen
 ■ Statistisch signifikante Unterschiede zwischen Mädchen und Jungen ($p \leq .05$); Linksseitige Ausprägung = Vorsprung der Mädchen; Rechtsseitige Ausprägung = Vorsprung der Jungen
 ■ = Unterschied in den mittleren Leistungswerten zu Jungen nicht statistisch signifikant ($p > .05$)
 A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Signifikante Geschlechterunterschiede für die Gesamtskala Mathematik können für die Kompetenzstufen (KS) II und V beobachtet werden. Der Anteil an Mädchen auf Kompetenzstufe II ist mit 22.9 Prozent signifikant höher als der Anteil an Jungen auf dieser Kompetenzstufe, der insgesamt 19 Prozent beträgt. Der Anteil an Mädchen auf der höchsten Kompetenzstufe V ist mit 4.5 Prozent hingegen signifikant geringer als der Anteil an Jungen (7.5%). Keine Unterschiede können für die Kompetenzstufe I sowie lediglich nominelle, nicht signifikante Effekte für die Kompetenzstufen IV zugunsten der Jungen und für Kompetenzstufe III zugunsten der Mädchen beobachtet werden. Anhand dieser deskriptiven Befunde wird deutlich, dass Jungen insbesondere auf den höheren Kompetenzstufen und Mädchen auf den niedrigeren Kompetenzstufen überrepräsentiert sind. In TIMSS 2015 waren diese Unterschiede durchgehend nicht signifikant.

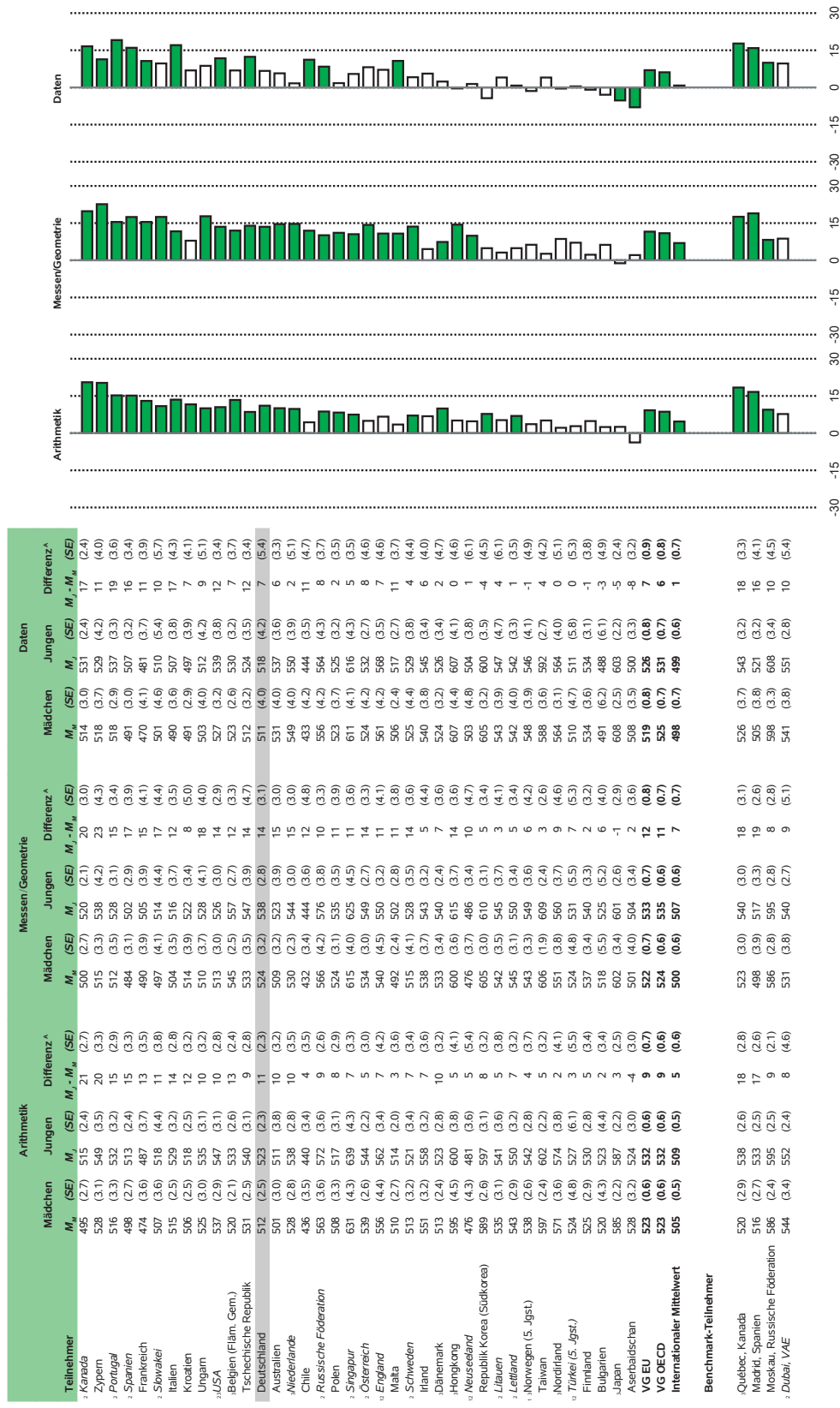
8.4.2 Geschlechterspezifische Leistungsunterschiede nach mathematischen Inhaltsbereichen

Um ein differenzierteres Bild geschlechterbezogener Unterschiede in Mathematik zeichnen zu können, sieht die TIMSS-Rahmenkonzeption vor, die drei Inhaltsbereiche *Arithmetik*, *Messen und Geometrie* sowie *Daten* getrennt voneinander zu betrachten (siehe Kapitel 3 in diesem Band). Dies ermöglicht es unter anderem, Potenziale und Herausforderungen nach Geschlecht zu identifizieren, um Hinweise für mögliche Veränderungen in der schulischen Praxis zu erhalten.

In Abbildung 8.3 sind erneut die Mittelwerte für Jungen (M_J) und Mädchen (M_M), die jeweiligen Standardfehler (SE) sowie die Differenz der Leistungen von Mädchen und Jungen ($M_J - M_M$) für jeden Teilnehmerstaat angegeben. Die Leistungswerte werden getrennt für die Inhaltsbereiche *Arithmetik*, *Messen und Geometrie* und *Daten* berichtet. Im Inhaltsbereich *Arithmetik* erreichen Jungen in allen Teilnehmerstaaten, mit Ausnahme von Aserbaidschan, höhere Testwerte als Mädchen – in mehr als der Hälfte der Teilnehmerstaaten ist dieser Unterschied signifikant. Unter den Staaten mit den größten Geschlechterunterschieden im Bereich *Arithmetik* befinden sich Kanada mit 21 Punkten und Zypern mit 20 Punkten Unterschied zugunsten der Jungen. Im internationalen Durchschnitt liegt die Differenz bei 5 Punkten. In den Vergleichsgruppen OECD und EU liegt die mittlere Geschlechterdifferenz bei 9 Punkten. Beide Werte sind signifikant. In Deutschland liegt der durchschnittliche Punktwert von Mädchen im Inhaltsbereich *Arithmetik* bei 512 Punkten, der der Jungen bei 523 Punkten. Der Unterschied von 11 Skalenwertpunkten ist statistisch signifikant und liegt über dem Durchschnitt der Vergleichsgruppen EU und OECD.

Für den Inhaltsbereich *Messen und Geometrie* können für alle Teilnehmerstaaten (mit Ausnahme von Japan) Geschlechterunterschiede zugunsten der Jungen beobachtet werden. In 25 Staaten (ohne *Benchmark*-Teilnehmer) ist der Unterschied signifikant. Diese reichen von 23 Leistungspunkten Differenz in Zypern und 20 Punkten Differenz in Kanada (signifikant) bis zu 2 Punkten Differenz in Finnland (nicht signifikant). Der internationale Unterschied liegt bei 7 Punkten Differenz zugunsten der Jungen, die mittlere Differenz in der Vergleichsgruppe OECD bei 11 Punkten und in der Vergleichsgruppe EU bei 12 Punkten. Alle diese Unterschiede sind statistisch signifikant und fallen zugunsten der Jungen aus. In Deutschland erreichen Jungen eine im Mittel um 14 Punkte höhere Testleistung im Inhaltsbereich *Messen und Geometrie* als

Abbildung 8.3: Testleistungen in Inhaltsbereichen nach Geschlecht im internationalen Vergleich – Mathematik



□ Keine statistisch signifikanten Unterschiede zwischen Mädchen und Jungen ($p > .05$); Linksschiefe Ausprägung = Vorsprung der Mädchen;
 ■ Statistisch signifikante Unterschiede zwischen Mädchen und Jungen ($p \leq .05$); Linksschiefe Ausprägung = Vorsprung der Mädchen;
 □ Statistisch signifikante Unterschiede zwischen Jungen und Mädchen ($p \leq .05$); Linksschiefe Ausprägung = Vorsprung der Jungen;
 ■ Statistisch signifikante Unterschiede zwischen Jungen und Mädchen ($p \leq .05$); Linksschiefe Ausprägung = Vorsprung der Jungen;
 Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss (für die Besonderheiten von Chile siehe Anhang).

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Mädchen. Damit liegt der Unterschied zwischen Mädchen und Jungen über dem Durchschnittswert der Vergleichsgruppen OECD und EU. Im Bereich *Messen und Geometrie* fallen die beobachteten Geschlechterunterschiede allgemein am deutlichsten aus.

Die Ergebnisse für Geschlechterunterschiede im Inhaltsbereich *Daten* können ebenfalls Abbildung 8.3 entnommen werden. Insgesamt fallen die Geschlechterunterschiede hier geringer aus. Hier können für insgesamt elf Staaten signifikante Unterschiede zugunsten der Jungen sowie in zwei Staaten signifikante Unterschiede zugunsten der Mädchen beobachtet werden. Am größten fällt der Geschlechterunterschied in Portugal aus. Dort erreichen Jungen im Mittel eine um 19 Punkte höhere Testleistung als Mädchen. In Japan und in Aserbaidschan erreichen Mädchen hingegen im Durchschnitt 5 Leistungspunkte mehr als Jungen. In 25 Teilnehmerstaaten (ohne *Benchmark*-Teilnehmer) liegen keine signifikanten Unterschiede vor. Der internationale Durchschnittswert für Geschlechterunterschiede im Inhaltsbereich *Daten* liegt bei einem Punkt und ist statistisch signifikant. In der Vergleichsgruppe EU liegt der mittlere Differenzwert bei 7 Punkten und in der Vergleichsgruppe OECD bei 6 Punkten, beide Werte fallen statistisch signifikant zugunsten der Jungen aus. In Deutschland liegt die Differenz für den Inhaltsbereich *Daten* bei 7 Leistungspunkten zugunsten der Jungen und ist signifikant. Dieser Wert entspricht der mittleren Differenz der Vergleichsgruppen. In diesem Inhaltsbereich zeichnen sich die vergleichsweise geringsten und am wenigsten konsistenten Geschlechterunterschiede ab.

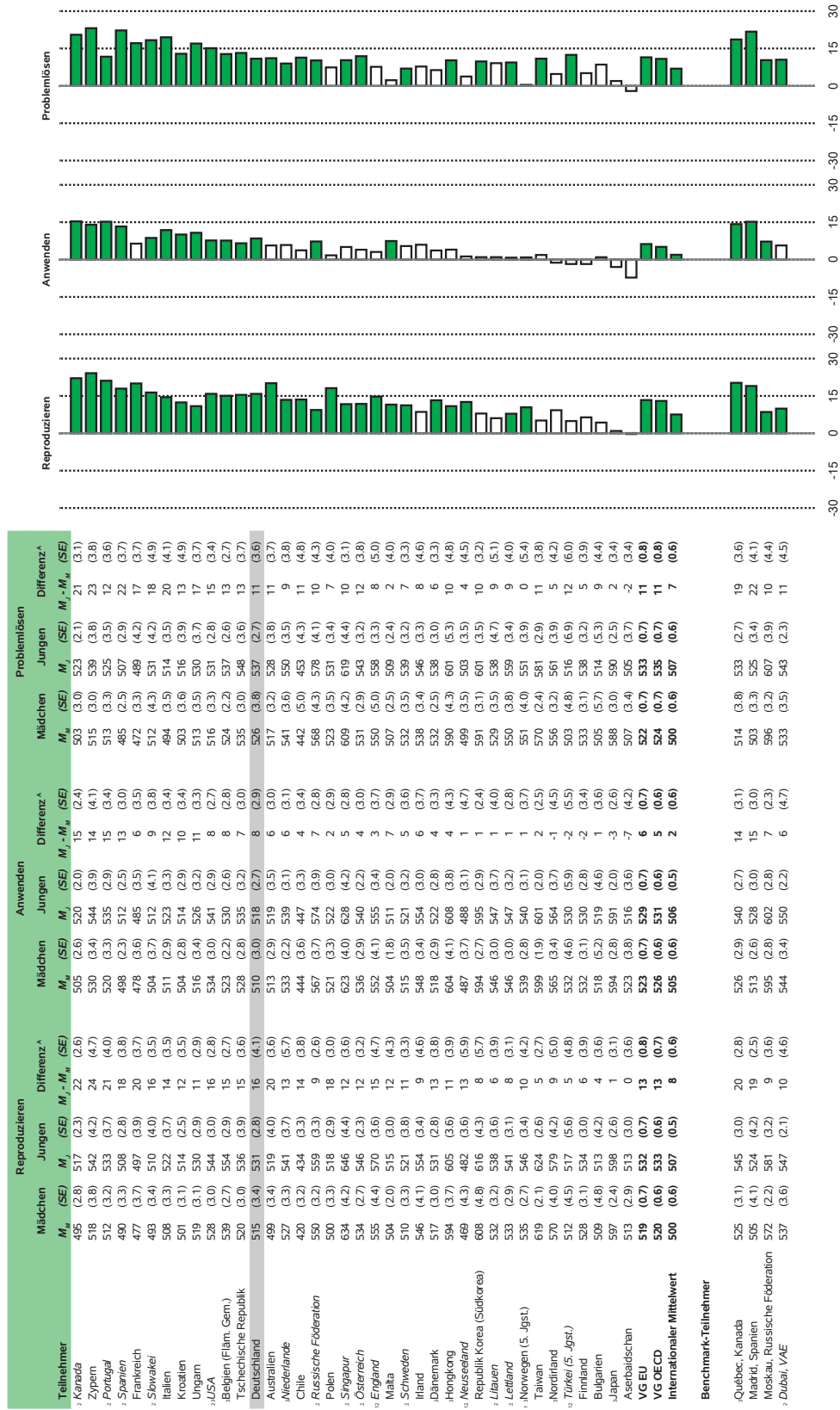
8.4.3 Geschlechterspezifische Unterschiede nach kognitiven Anforderungsbereichen in Mathematik

Die Rahmenkonzeption von TIMSS sieht neben der Differenzierung nach Inhaltsbereichen auch eine Differenzierung nach den drei Anforderungsbereichen *Reproduzieren*, *Anwenden* und *Problemlösen* vor. Die TIMSS-Testaufgaben können jeweils einem dieser Bereiche zugeordnet werden. Die Ergebnisse für Geschlechterunterschiede nach Anforderungsbereichen für Deutschland sowie den anderen Teilnehmerstaaten sind in Abbildung 8.4 dargestellt.

Im Bereich *Reproduzieren* können mit wenigen Ausnahmen signifikante Unterschiede zugunsten der Jungen in den Teilnehmerstaaten beobachtet werden. In allen Teilnehmerstaaten übersteigen die Testleistungen der Jungen diejenigen der Mädchen, in neun Staaten können die beobachteten Unterschiede jedoch nicht gegen den Zufall abgesichert werden. Die größten Unterschiede nach Geschlecht finden sich erneut in Kanada und Zypern. In Zypern beträgt die mittlere Differenz der Leistungen von Jungen gegenüber denen der Mädchen 24 Punkte, in Kanada liegt dieser Wert bei 22 Punkten zugunsten der Jungen. Am geringsten ausgeprägt ist die mittlere Differenz in Japan (1 Punkt) und in Aserbaidschan (0 Punkte). Hier ist der Unterschied nicht signifikant. International liegt die mittlere Differenz zugunsten der Jungen bei 8 Punkten. In der Vergleichsgruppe OECD und EU liegt die mittlere Differenz bei jeweils 13 Punkten. Die berichteten Unterschiede sind signifikant. In Deutschland erreichen Mädchen im Bereich *Reproduzieren* durchschnittlich eine um 16 Punkte geringere Testleistungen als Jungen. Dieser Unterschied ist signifikant und liegt oberhalb der mittleren Differenz der Vergleichsgruppen OECD und EU.

Für den Anforderungsbereich *Anwenden* liegen für 14 Teilnehmerstaaten (ohne *Benchmark*-Teilnehmer) signifikant höhere Testergebnisse für Jungen vor, in

Abbildung 8.4: Testleistungen in kognitiven Anforderungsbereichen nach Geschlecht im internationalen Vergleich – Mathematik



□ Keine statistisch signifikanten Unterschiede zwischen Mädchen und Jungen ($p > .05$); Linkssseitige Ausprägung = Vorsprung der Mädchen;
 ■ Rechtssseitige Ausprägung = Vorsprung der Jungen
 ■ Statistisch signifikante Unterschiede zwischen Mädchen und Jungen ($p \leq .05$); Linkssseitige Ausprägung = Vorsprung der Mädchen;
 ■ Rechtssseitige Ausprägung = Vorsprung der Jungen
 Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

fünf Staaten weisen Mädchen nominell höhere Testleistungen auf, diese Unterschiede sind jedoch statistisch nicht signifikant. In 24 Teilnehmerstaaten (ohne *Benchmark*-Teilnehmer) können keine signifikanten Unterschiede beobachtet werden. Die größten Unterschiede zugunsten von Jungen liegen in Kanada und in Portugal (jeweils 15 Punkte Differenz) vor. Diese Unterschiede sind signifikant. Im internationalen Durchschnitt beträgt die mittlere Differenz 2 Punkte zugunsten der Jungen. Die mittlere Differenz in der Vergleichsgruppe OECD beträgt 5 Punkte, die der Vergleichsgruppe EU 6 Punkte. Alle Unterschiede sind statistisch signifikant und fallen zugunsten der Jungen aus. In Deutschland beträgt der mittlere Geschlechterunterschied 8 Punkte zugunsten der Jungen. Dieser Unterschied ist signifikant und liegt oberhalb des Differenzwertes in den Vergleichsgruppen. Insgesamt sind die Geschlechterunterschiede deutlich geringer ausgeprägt als in den anderen Inhalts- oder Anforderungsbereichen und erreichen auch seltener das Signifikanzniveau.

Bei Betrachtung der Geschlechterunterschiede im Anforderungsbereich *Problemlösen* wird deutlich, dass in allen Teilnehmerstaaten (mit Ausnahme von Aserbaidschan) eine Geschlechterdifferenz zugunsten der Jungen beobachtet werden kann. In 25 Staaten ist dieser Unterschied statistisch signifikant, in 13 Staaten nicht. Die größten Unterschiede können in Zypern (23 Punkte Differenz) und Spanien (22 Punkte Differenz) beobachtet werden. Der internationale Mittelwert beträgt 7 Leistungspunkte und die mittleren Geschlechterunterschiede in den Vergleichsgruppen OECD und EU betragen 11 Punkte zugunsten der Jungen. Diese Unterschiede sind signifikant. In Deutschland erreichen Jungen im Anforderungsbereich *Problemlösen* durchschnittlich 11 Leistungspunkte mehr als Mädchen. Dieser Unterschied ist statistisch signifikant und entspricht dem Wert in den Vergleichsgruppen OECD und EU.

8.4.4 Trends in geschlechterspezifischen Leistungsunterschieden in Mathematik

Von besonderer Bedeutung ist, wie sich die Geschlechterunterschiede über die Zeit entwickelt haben. Dabei ist zunächst interessant, ob Mädchen und Jungen ihre mathematischen Testleistungen insgesamt verbessern konnten. Eine getrennte Betrachtung der Leistungen von Mädchen und Jungen auf der Gesamtskala Mathematik für die Jahre 2007, 2011, 2015 und 2019 ermöglicht die Abbildung 8.5. In der linken Spalte finden sich die Teilnehmerstaaten sortiert nach dem höchsten Differenzwert im Vergleich der Jahre 2007 und 2019 ($M_{19} - M_{07}$). Links davon befindet sich ein Hinweis auf Besonderheiten der Stichproben in den einzelnen Staaten. Die Ergebnisse für Mädchen und Jungen werden untereinander in zwei getrennten Tabellen ausgewiesen. Für jedes Jahr werden die Mittelwerte (M) der jeweiligen Testleistungen auf der Gesamtskala Mathematik sowie die zugehörigen Standardfehler (SE) berichtet. Rechts daneben werden die Veränderungen des Mittelwertes von TIMSS 2019 im Vergleich zu den vorangegangenen Zyklen TIMSS 2015, TIMSS 2011 und TIMSS 2007 sowie die zugehörigen Standardfehler abgebildet. In der letzten Spalte ist die Differenz des aktuellen Zyklus zu TIMSS 2007 dokumentiert. Dieses Ergebnis wird in Form eines Balkens rechts neben der Tabelle veranschaulicht. Ein grün eingefärbter Balken verweist auf signifikante Leistungsunterschiede zwischen den Testzyklen 2007 und 2019.

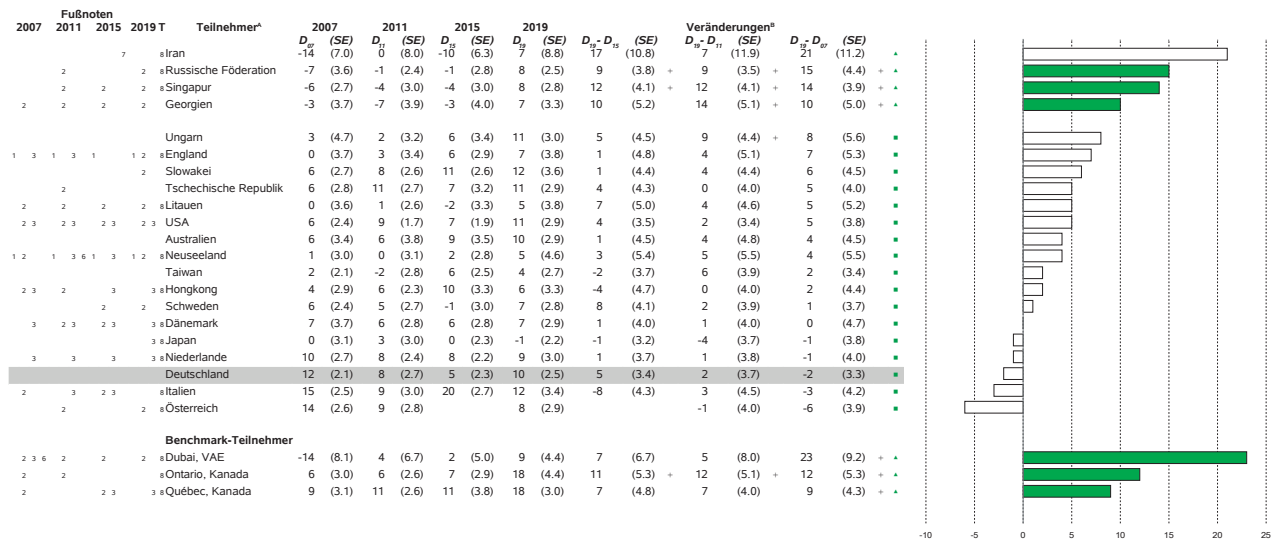
Abbildung 8.5: Vergleich der Leistungsentwicklungen von Mädchen und Jungen in Mathematik in TIMSS 2007, 2011, 2015 und 2019 (internationaler Vergleich)

- Keine statistisch signifikanten Unterschiede zwischen 2007 und 2019 ($p > .05$)
 ■ Statistisch signifikante Unterschiede zwischen 2007 und 2019 ($p \leq .05$)
 ▲ = Differenzwerte im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)
 ▼ = Differenzwerte im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)
 ■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)
 + = Mittelwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)
 - = Mittelwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)
 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 6 = Abweichender Testzeitpunkt
 7 = Teilnahme an TIMSS Less Difficult Mathematics
 8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019
 A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.
 B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Anhand Abbildung 8.5 wird deutlich, dass die Mathematikleistungen von Mädchen in 17 Teilnehmerstaaten (ohne *Benchmark*-Teilnehmer) im Vergleich zu TIMSS 2007 nominell zugenommen haben, in zwölf Staaten ist der Unterschied der Leistungspunkte aus TIMSS 2007 zu TIMSS 2019 signifikant. In vier Staaten erreichen Mädchen geringere Leistungen als in TIMSS 2007, wobei dieser Unterschied jedoch nicht signifikant ist. Den größten Zuwachs erreichen Viertklässlerinnen in der Tschechischen Republik. Dort erreichten sie im Jahr 2007 im Mittel 483 Leistungspunkte. Im Jahr 2019 erreichen sie durchschnittlich 44 Leistungspunkte mehr und kommen auf eine durchschnittliche Leistung von 527 Punkten. Auch in Georgien und Österreich haben sich die Leistungen von Mädchen im aktuellen Zyklus im Vergleich zu TIMSS 2007 deutlich verbessert. In Australien, Deutschland, Hongkong und Neuseeland sind die Leistungen der Mädchen leicht zurückgegangen. Die Unterschiede zwischen den Leistungen in TIMSS 2007 und TIMSS 2019 sind jedoch statistisch nicht signifikant. Konnten Mädchen in Deutschland ihre Testleistung im Jahr 2011 im Vergleich zum Jahr 2007 zunächst verbessern, so zeigt sich in den folgenden Jahren ein abnehmender Trend. Die im aktuellen Zyklus erreichten Testleistungen der Mädchen (516 Punkte) liegen erstmals unter den in TIMSS 2007 erreichten Testleistungen (519 Punkte). Dieser Unterschied ist nicht signifikant. Demnach konnten Mädchen ihre mathematischen Leistungen in den vergangenen zwölf Jahren nicht verbessern; die Leistungen sind stagniert. Zudem deutet sich ein negativer Entwicklungstrend für die mathematischen Testleistungen von Mädchen an.

Für die Jungen kann in vielen Teilnehmerstaaten ein positiver Trend beobachtet werden. In 18 Staaten (ohne *Benchmark*-Teilnehmer) konnten Jungen ihre Testleistungen nominell verbessern, in 14 Staaten ist dieser Trend statistisch signifikant. In drei Staaten liegen die Testleistungen von Jungen in TIMSS 2019 unter denen in TIMSS 2007. Dieser Trend ist statistisch nicht signifikant. Die größten Unterschiede zwischen den Testleistungen in TIMSS 2007 und TIMSS 2019 können im Iran, in der Tschechischen Republik und in Georgien beobachtet werden. Dort liegt die Differenz bei 49 Punkten für die Tschechische Republik und Georgien sowie bei 51 Punkten im Iran. Die Mathematikleistungen der Jungen stagnieren im Vergleichszeitraum für die Staaten Neuseeland, Hongkong und Deutschland mit negativer Tendenz und in Dänemark, Australien und den Niederlanden mit leicht positiver Tendenz. In Deutschland beträgt die Differenz in den Leistungen von TIMSS 2019 im Vergleich zu TIMSS 2007 insgesamt 5 Punkte. Erreichten Jungen in TIMSS 2007 durchschnittlich noch 531 Leistungspunkte, so waren es in TIMSS 2011 durchschnittlich 532 Punkte, in TIMSS 2015 nur noch 524 Punkte und in TIMSS 2019 noch 526 Punkte. Damit haben sich auch die Leistungen der Jungen im Zeitraum der Jahre 2007 bis 2019 zwar leicht verringert, aber nicht signifikant verändert.

Eine Abschätzung des Ausmaßes der Entwicklung von Leistungsunterschieden von Mädchen und Jungen ermöglicht Abbildung 8.6. In der Russischen Föderation, in Singapur und in Georgien gab es in TIMSS 2007 Leistungsvorsprünge zugunsten der Mädchen. In TIMSS 2019 liegen nun deutliche Leistungsvorsprünge zugunsten der Jungen vor. In zehn weiteren Teilnehmerstaaten zeigt sich eine Tendenz hinsichtlich der Zunahme von Geschlechterdisparitäten. In Deutschland sowie in vier weiteren Staaten kann eine Abnahme von Geschlechterunterschieden im Vergleich zu TIMSS 2007 beobachtet werden, wobei die Unterschiede nicht signifikant sind. Für Dänemark, Japan und die Niederlande zeigen sich nahezu keine Veränderungen zwischen TIMSS 2007 und 2019. Der größte beobachtete Geschlechterunterschied auf der Gesamtskala Mathematik konnte für Deutschland

Abbildung 8.6: Vergleich der Leistungsunterschiede in Mathematik von Mädchen und Jungen in TIMSS 2007, 2011, 2015 und 2019 (internationaler Vergleich)

- Keine statistisch signifikanten Unterschiede zwischen 2007 und 2019 ($p > .05$)
 ■ Statistisch signifikante Unterschiede zwischen 2007 und 2019 ($p \leq .05$)
 ▲ = Differenzwerte im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)
 ▼ = Differenzwerte im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)
 ■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)
 + = Mittelwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)
 - = Mittelwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)
 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 6 = Abweichender Testzeitpunkt
 7 = Teilnahme an TIMSS Less Difficult Mathematics
 8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019
 A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.
 B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

in TIMSS 2007 beobachtet werden. Hier lag der mittlere Unterschied in den Testleistungen bei 12 Punkten zugunsten der Jungen. In TIMSS 2011 fiel dieser Unterschied mit 8 Punkten um 4 Punkte kleiner aus. Ein ähnlich niedriger Wert konnte auch in TIMSS 2015 beobachtet werden, wo der Unterschied noch 5 Leistungspunkte betrug. Dieser sich über die drei Erhebungen erstreckende positive Trend kann im aktuellen Zyklus jedoch nicht beobachtet werden. Hier zeigt sich wieder eine Zunahme der Geschlechterunterschiede auf einen Wert von 10 Differenzpunkten. Dieser Wert unterscheidet sich statistisch nicht vom Ausmaß der Geschlechterunterschiede im Jahr 2007. Dieser Befund kann in Bezug auf den Abbau von Geschlechterdisparitäten im Fach Mathematik nicht zufriedenstellen.

8.4.5 Geschlechterspezifische Unterschiede in Mathematiknoten

Im deutschen Schulsystem spielen schulische Noten nach wie vor eine bedeutende Rolle. Zum einen ermöglichen sie eine direkte Einordnung von Leistungsergebnissen, zum anderen kommt ihnen eine bedeutende Rolle bei der Selbstkonzeptgenese und der Einstellung zu einem Fach zu (siehe Abschnitt 8.2 sowie Wendt, Steinmayr & Kasper, 2016). In Tabelle 8.1 sind die Mittelwerte und Verteilungen der Noten im Fach Mathematik für Mädchen und Jungen getrennt ausgewiesen. Die Mädchen erhalten im Halbjahreszeugnis der Klasse 4 im Notendurchschnitt eine 2.6 ($SE = 0.03$), Jungen hingegen eine 2.5 ($SE = 0.03$). Dieser Unterschied ist statistisch signifikant. Insgesamt erreichen

Tabelle 8.1: Mittelwerte und Verteilungen der Noten in Mathematik nach Geschlecht (Anteile der Schülerinnen und Schüler in Prozent)

Mathematiknote (Halbjahreszeugnis 4. Jahrgangsstufe) ^A						
	M (SE)	1 % (SE)	2 % (SE)	3 % (SE)	4 % (SE)	5 % (SE)
Mädchen	2.6* (0.03)	10.1* (0.8)	38.7 (1.4)	31.2 (1.3)	16.8* (1.0)	2.9 (0.4)
Jungen	2.5 (0.03)	13.3 (0.9)	40.7 (1.4)	29.8 (1.4)	13.8 (1.0)	1.9 (0.3)
Gesamt	2.5 (0.02)	11.8 (0.7)	39.7 (1.2)	30.5 (0.9)	15.3 (0.8)	2.4 (0.3)

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

A = Aufgrund zu geringer Fallzahlen kann die Note 6 nicht ausgewertet werden.

* = Unterschied zu Jungen statistisch signifikant ($p \leq .05$)

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

etwa die Hälfte der Viertklässlerinnen und Viertklässler gute oder sogar sehr gute Noten in Mathematik. Von den Mädchen erhalten 48.8 Prozent die Note *sehr gut* oder *gut*. Bei den Jungen liegt der Anteil bei 54 Prozent. Die Note *mangelhaft* erhalten hingegen 2.9 Prozent der Mädchen und 1.9 Prozent der Jungen.

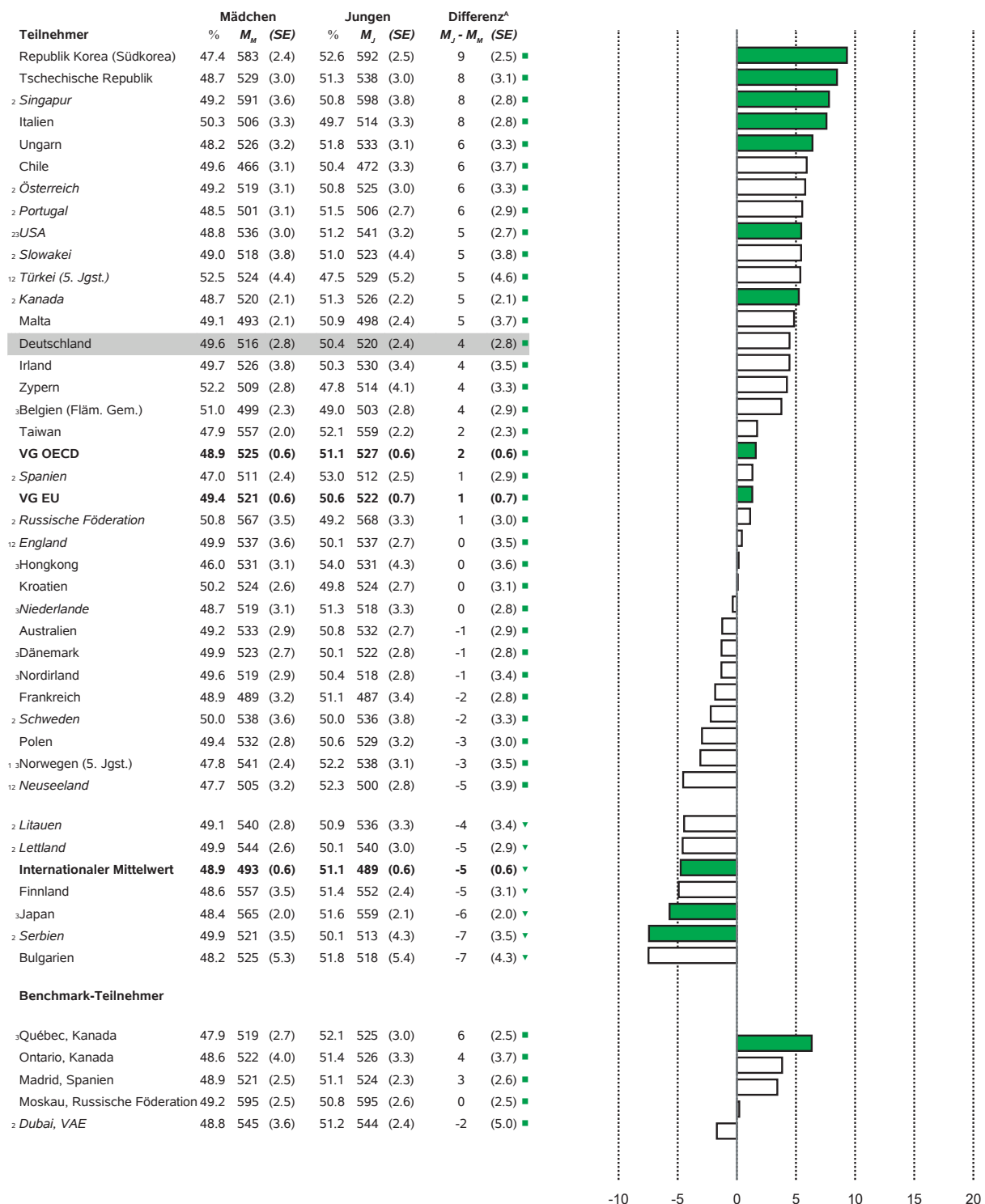
Bezogen auf die Verteilung von Einzelnoten kann beobachtet werden, dass der Anteil an Mädchen, die die Note eins (*sehr gut*) erhalten, signifikant geringer ist als der Anteil an Jungen, die diese Note erhalten. Hinsichtlich der Note vier (*ausreichend*) liegt ein signifikant höherer Anteil an Mädchen vor. Somit zeigen sich Geschlechtereffekte besonders im Extrembereich der sehr guten Noten und im Bereich der Note *ausreichend*, die eine Leistung unterhalb des rechnerischen Mittels ausweist, sowie tendenziell im Bereich *mangelhaft*.

Auf eine Betrachtung der Schulnoten von Mädchen und Jungen differenziert nach Testleistungen wie sie im Berichtsband des vorherigen TIMSS-Zyklus zu finden ist, wird an dieser Stelle verzichtet. Dies begründet sich darin, dass die aktuellen Verteilungen denen in TIMSS 2015 sehr stark ähneln und demnach kein zusätzlicher Erkenntnisgewinn erzielt werden kann (vgl. Wendt, Steinmayr & Kasper, 2016).

8.5 Ergebnisse zu geschlechterspezifischen Leistungsunterschieden in den Naturwissenschaften

Im Anschluss an die Betrachtung der Ergebnisse zu Geschlechterunterschieden in Mathematik werden in diesem Abschnitt die Unterschiede von Mädchen und Jungen im Hinblick auf die naturwissenschaftlichen Leistungen und die Sachunterrichtsnoten dokumentiert.

Abbildung 8.7: Testleistung nach Geschlecht – Gesamtskala Naturwissenschaften



□ Keine statistisch signifikanten Unterschiede zwischen Mädchen und Jungen ($p > .05$)

■ Statistisch signifikante Unterschiede zwischen Mädchen und Jungen ($p \leq .05$)

▲ = Differenzwert im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)

▼ = Differenzwert im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)

■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)

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

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

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

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

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

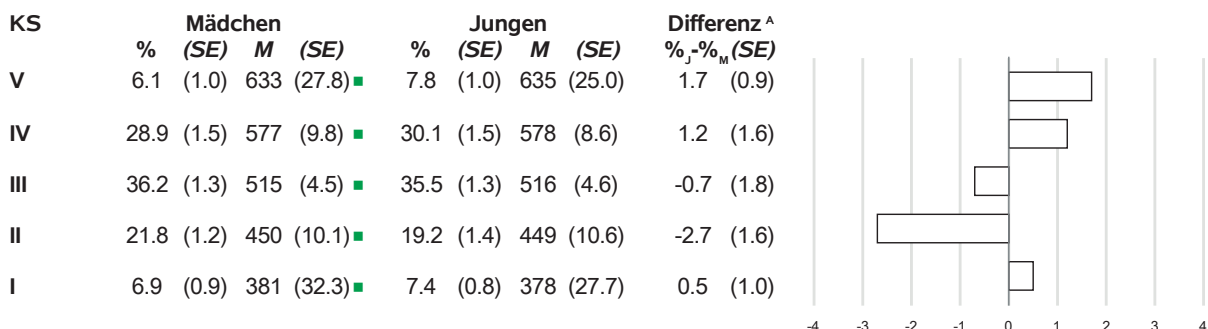
8.5.1 Geschlechterspezifische Leistungsunterschiede auf der Gesamtskala Naturwissenschaften

Die Ergebnisse für die Mädchen und Jungen auf der Gesamtskala Naturwissenschaften sind in Abbildung 8.7 dokumentiert. Diese entspricht in ihrem Aufbau der Abbildung 8.1 in Abschnitt 8.4.1.

Abbildung 8.7 zeigt die durchschnittlichen naturwissenschaftlichen Testleistungen von Mädchen und Jungen für Deutschland und die anderen Teilnehmerstaaten. In sieben Teilnehmerstaaten weisen Jungen signifikant höhere Testleistungen auf als Mädchen (ohne *Benchmark*-Teilnehmer). In zwei Staaten erreichen Mädchen signifikant höhere Testleistungen als Jungen. In der Mehrzahl der Teilnehmerstaaten liegen jedoch keine signifikanten Geschlechterunterschiede vor. Die größten Unterschiede finden sich in der Republik Korea mit einer Differenz von 9 Leistungspunkten sowie in der Tschechischen Republik, Singapur und Italien mit jeweils 8 Punkten Differenz. In Japan erreichen Mädchen im Durchschnitt 6 Leistungspunkte mehr als Jungen, in Serbien sind es 7 Punkte. Diese Unterschiede sind signifikant. Im internationalen Durchschnitt beträgt der Geschlechterunterschied 5 Punkte zugunsten der Mädchen und ist statistisch signifikant. Der Wert für die Vergleichsgruppe OECD beträgt 2 Punkte zugunsten der Jungen und der Wert für die Vergleichsgruppe EU einen Punkt. Beide Werte sind statistisch signifikant. In Deutschland beträgt der mittlere Differenzwert 4 Leistungspunkte zugunsten der Jungen, der allerdings nicht als statistisch signifikant eingestuft werden kann. Demnach können im aktuellen Zyklus keine Geschlechterdisparitäten auf der Gesamtskala Naturwissenschaften beobachtet werden. Weitere Hinweise auf bestehende Geschlechterunterschiede bietet die differenzierte Betrachtung von Unterschieden nach Inhaltsbereichen sowie kognitiven Anforderungsbereichen.

Eine differenzierte Betrachtung von Geschlechterunterschieden nach Kompetenzstufen ermöglicht Abbildung 8.8. Hier zeigt sich eine leichte Tendenz, wonach Jungen auf den höheren Kompetenzstufen IV und V nominell überrepräsentiert sind, wobei Mädchen auf den unteren Kompetenzstufen II und III leicht überrepräsentiert scheinen. Die prozentualen Anteile von Mädchen und Jungen auf den Kompetenzstufen unterscheiden sich jedoch nicht signifikant voneinander.

Abbildung 8.8: Mittlere Testleistungen von Mädchen und Jungen und prozentuale Verteilungen auf die Kompetenzstufen – Naturwissenschaften

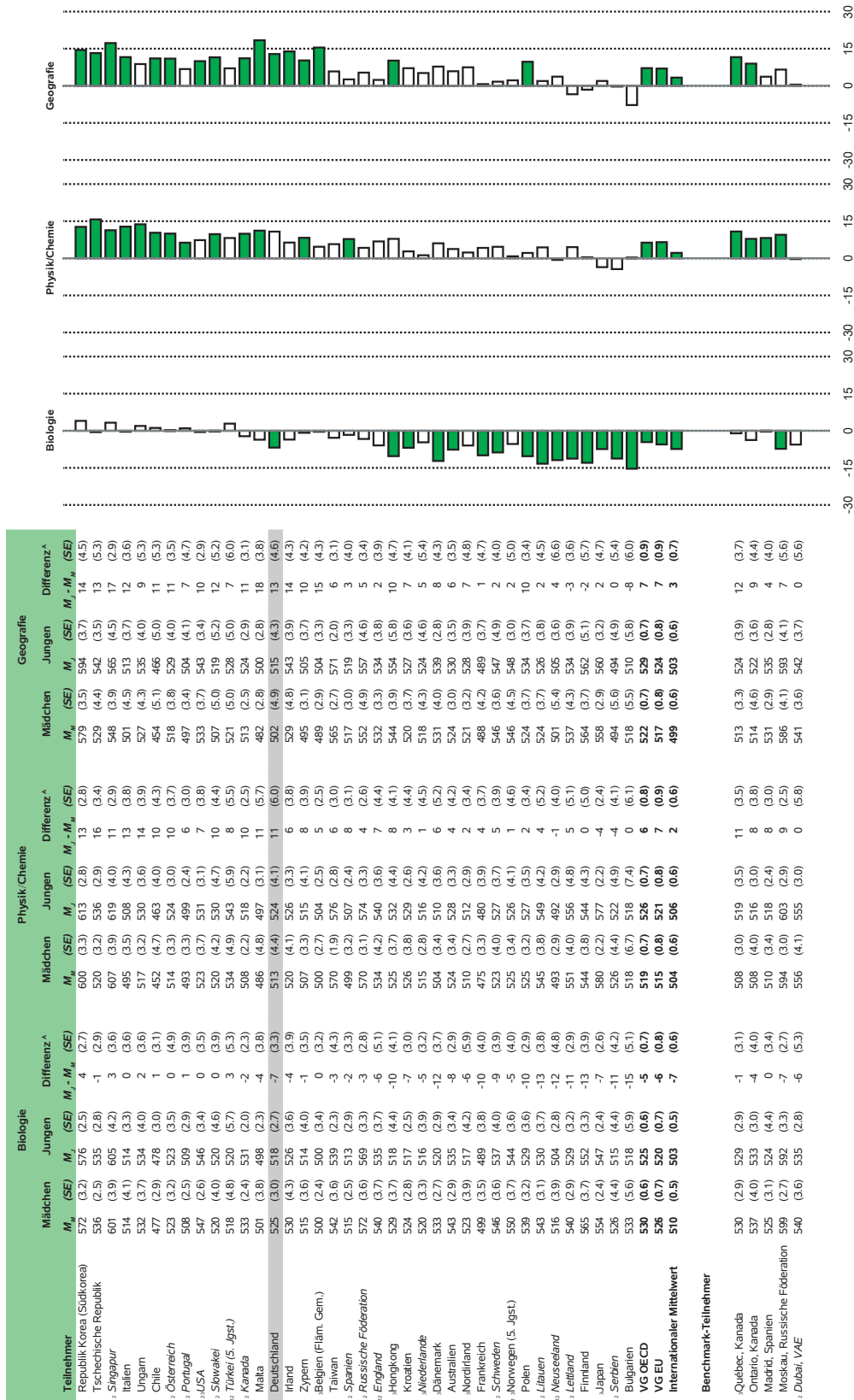


□ Keine statistisch signifikanten Unterschiede zwischen Mädchen und Jungen ($p > .05$); Linksseitige Ausprägung = Vorsprung der Mädchen; Rechtsseitige Ausprägung = Vorsprung der Jungen

■ = Unterschied in den mittleren Leistungswerten zu Jungen nicht statistisch signifikant ($p > .05$)

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

Abbildung 8.9: Testleistungen in Inhaltsbereichen nach Geschlecht im internationalen Vergleich – Naturwissenschaften



□ Keine statistisch signifikanten Unterschiede zwischen Mädchen und Jungen ($p > 0.05$); Linksseltige Ausprägung = Vorsprung der Mädchen;
 ■ Rechtsseltige Ausprägung = Vorsprung der Jungen
 ■ Statistisch signifikante Unterschiede zwischen Mädchen und Jungen ($p \leq 0.05$); Linksseltige Ausprägung = Vorsprung der Mädchen;
 ■ Rechtsseltige Ausprägung = Vorsprung der Jungen
 Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfunggrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

8.5.2 Geschlechterspezifische Leistungsunterschiede nach naturwissenschaftlichen Inhaltsbereichen

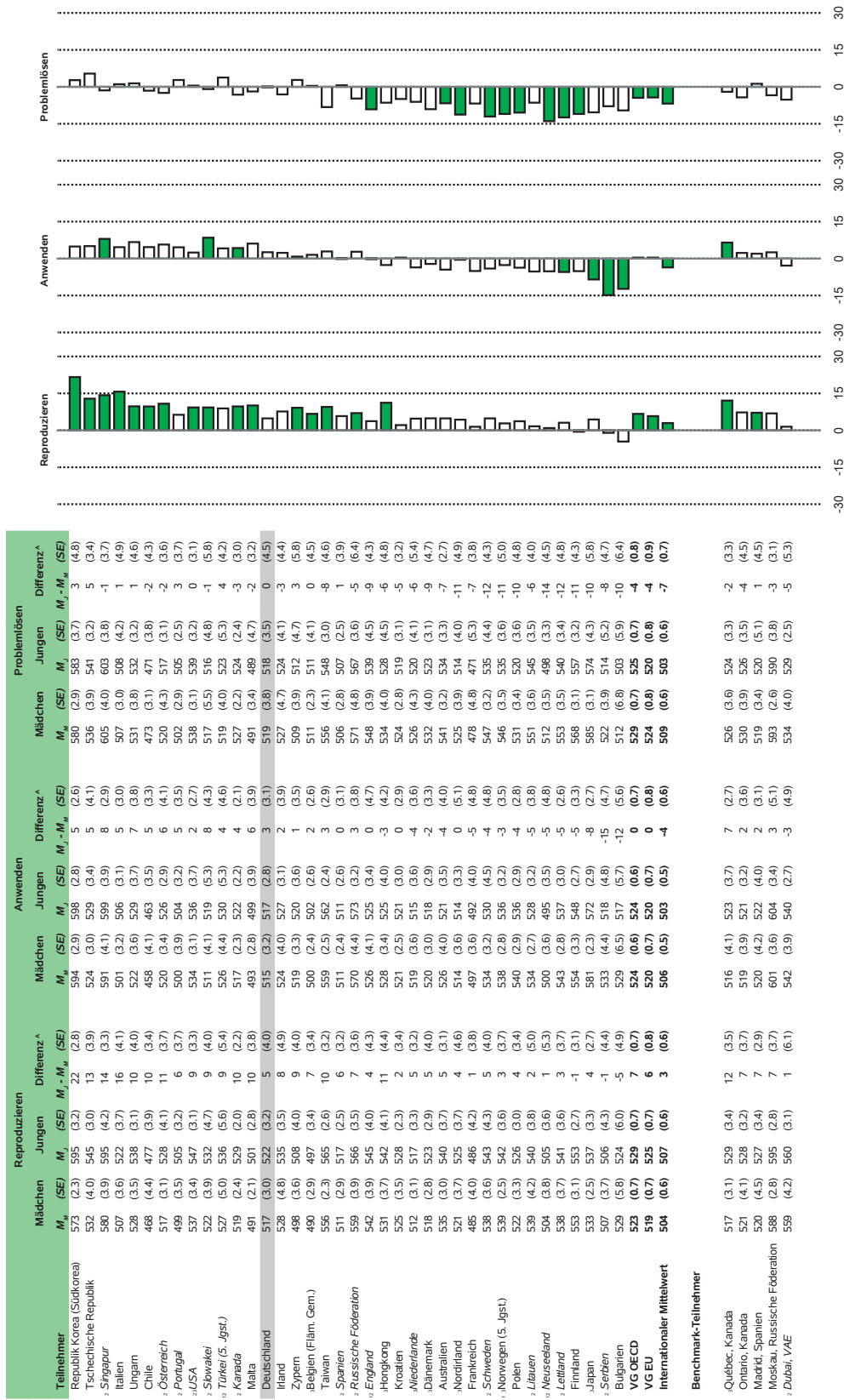
Die Betrachtung der Geschlechterunterschiede auf der Gesamtskala Naturwissenschaften ermöglicht nur begrenzt Aussagen zu Potenzialen und Förderbedarfen in den Naturwissenschaften. Eine differenziertere Betrachtung nach Inhaltsbereichen ermöglicht hingegen Abbildung 8.9. In Abbildung 8.9 werden zunächst die in den drei Inhaltsbereichen *Biologie*, *Physik/Chemie* und *Geografie* erreichten mittleren Leistungspunkte für Mädchen und Jungen, die zugehörigen Standardfehler sowie die mittlere Differenz ($M_J - M_M$) getrennt berichtet. Die Abbildung ist analog zu Abbildung 8.3 in Abschnitt 8.4.2 aufgebaut.

Im Inhaltsbereich *Biologie* können in 15 Teilnehmerstaaten signifikante Geschlechterunterschiede zugunsten der Mädchen beobachtet werden (ohne *Benchmark*-Teilnehmer). In keinem der Teilnehmerstaaten weisen Jungen signifikant höhere Testwerte im Bereich *Biologie* auf als Mädchen. In der Mehrzahl der Staaten liegen keine statistisch bedeutsamen Unterschiede vor. Im internationalen Durchschnitt beträgt die Geschlechterdifferenz 7 Punkte zugunsten der Mädchen, in der Vergleichsgruppe OECD 5 Punkte und in der Vergleichsgruppe EU 6 Punkte. Die Mittelwertunterschiede sind für alle drei Gruppen statistisch signifikant. In Deutschland beträgt der Vorsprung der Mädchen gegenüber den Jungen im Mittel 7 Punkte, ist signifikant und liegt knapp oberhalb der Werte in den Vergleichsgruppen. Insgesamt zeigt sich im Inhaltsbereich *Biologie* in vielen Teilnehmerstaaten sowie in Deutschland ein deutlicher Vorsprung der Mädchen gegenüber den Jungen.

Die Ergebnisse für den Inhaltsbereich *Physik/Chemie* fallen hingegen eher zugunsten der Jungen aus. Hier finden sich in allen Staaten (mit Ausnahme von Serbien und Japan) Mittelwertunterschiede zugunsten der Jungen, in 13 Staaten ist der Unterschied statistisch signifikant. Der größte Differenzwert kann mit 16 Punkten für die Tschechische Republik beobachtet werden. Insgesamt beträgt die internationale Mittelwertdifferenz 2 Punkte. In den Vergleichsgruppen EU und OECD liegen die Werte mit 7 beziehungsweise 6 Punkten darüber. Diese Mittelwertunterschiede sind statistisch signifikant. In Deutschland beträgt der mittlere Unterschied im Inhaltsbereich *Physik/Chemie* 11 Leistungspunkte zugunsten der Jungen. Der Unterschied ist statistisch signifikant und liegt oberhalb der Differenzwerte in den Vergleichsgruppen.

Bei Betrachtung des Inhaltsbereiches *Geografie* können in 16 Staaten (ohne *Benchmark*-Teilnehmer) statistisch signifikante Geschlechterunterschiede beobachtet werden. In drei Teilnehmerstaaten (Bulgarien, Finnland, Lettland) können Tendenzen zugunsten der Mädchen beobachtet werden, diese sind jedoch nicht signifikant. Die größten Unterschiede zugunsten der Jungen sind in Malta (18 Punkte Differenz) und Singapur (17 Punkte Differenz) zu verzeichnen. Im internationalen Mittel fällt die Differenz mit 3 Leistungspunkten signifikant zugunsten der Jungen aus. In den Vergleichsgruppen OECD und EU beträgt die mittlere Differenz 7 Punkte und fällt ebenfalls zugunsten der Jungen aus. In Deutschland zeigt sich eine Geschlechterdifferenz auf der Testskala *Geografie* von 13 Punkten zugunsten der Jungen. Dieser Unterschied ist statistisch signifikant und liegt oberhalb der Differenzwerte in den Vergleichsgruppen.

Abbildung 8.10: Testleistungen in kognitiven Anforderungsbereichen nach Geschlecht im internationalen Vergleich – Naturwissenschaften



□ Keine statistisch signifikanten Unterschiede zwischen Mädchen und Jungen ($p > .05$); Linksseltige Ausprägung = Vorsprung der Mädchen;
 ■ Rechtsseltige Ausprägung = Vorsprung der Jungen
 □ Statistisch signifikante Unterschiede zwischen Mädchen und Jungen ($p \leq .05$); Linksseltige Ausprägung = Vorsprung der Mädchen;
 ■ Rechtsseltige Ausprägung = Vorsprung der Jungen
 Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfunggrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schullebene erreichen nicht die internationalen Vorgaben.
 A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

8.5.3 Geschlechterspezifische Unterschiede nach kognitiven Anforderungsbereichen in den Naturwissenschaften

Analog zur Rahmenkonzeption Mathematik wurde auch im Bereich der Naturwissenschaften eine Differenzierung auf der Grundlage der kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen* vorgenommen. Inwiefern sich Mädchen und Jungen in ihren Testleistungen in den genannten Anforderungsbereichen unterscheiden, zeigt Abbildung 8.10.

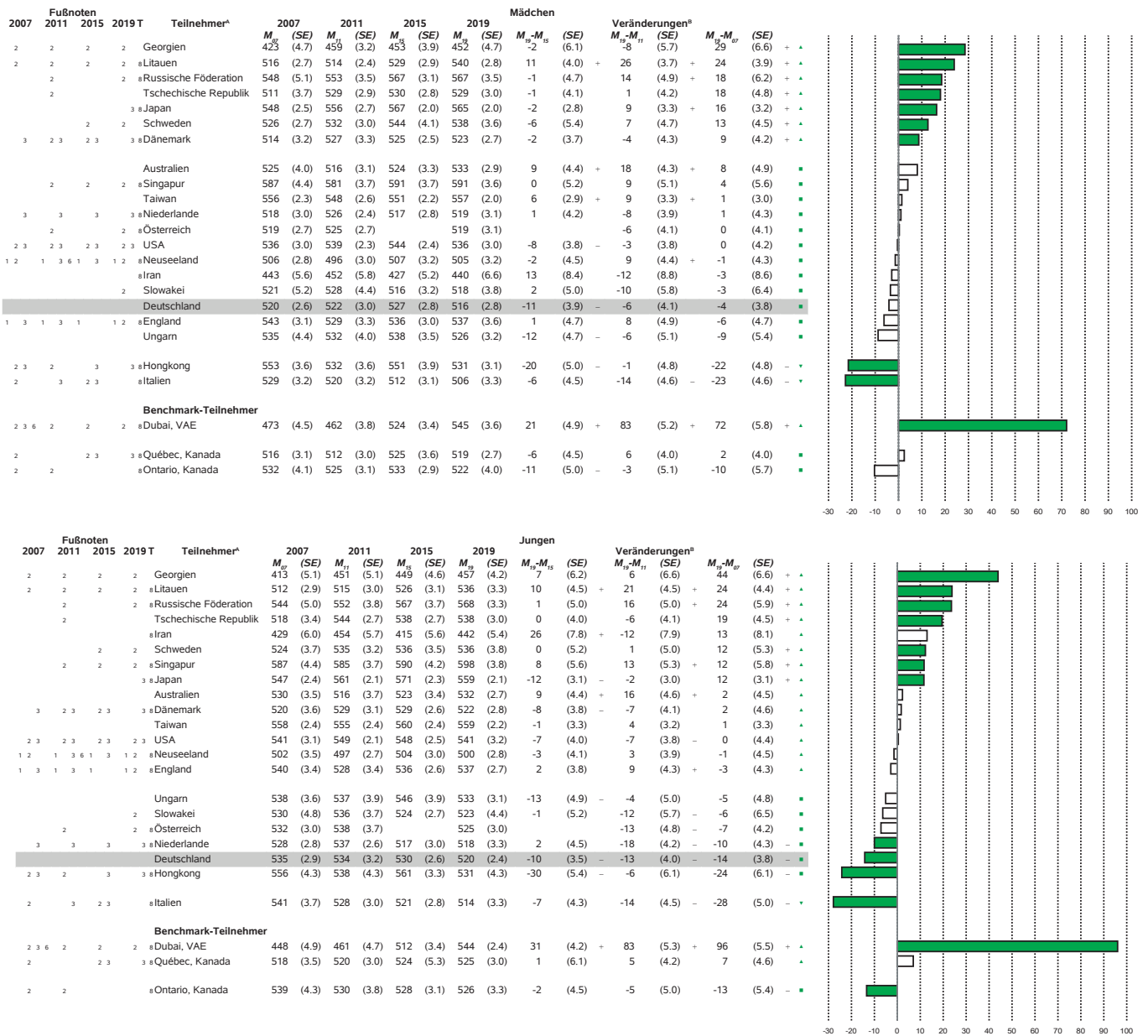
Im Anforderungsbereich *Reproduzieren* erreichen Jungen in 35 Staaten höhere Testwerte als Mädchen, in 16 Staaten ist dieser Unterschied signifikant (ohne *Benchmark*-Teilnehmer). In drei Staaten liegt eine Tendenz zugunsten der Mädchen vor (Serbien, Bulgarien, Finnland), wobei diese Unterschiede gering sind und nicht gegen den Zufall abgesichert werden können. Der internationale Mittelwert für Geschlechterunterschiede in diesem Bereich liegt bei 3 Leistungspunkten. Die Mittelwerte der Vergleichsgruppen liegen bei 7 Punkten (VG_{OECD}) und 6 Punkten (VG_{EU}) und sind statistisch signifikant. In Deutschland beträgt die Differenz 5 Punkte und ist nicht statistisch signifikant. Die Leistungen von Mädchen und Jungen in Deutschland unterscheiden sich im Anforderungsbereich *Reproduzieren* nicht signifikant voneinander.

Im Anforderungsbereich *Anwenden* zeigt sich ein heterogeneres Bild hinsichtlich der beobachteten Geschlechterunterschiede. In drei Staaten können signifikante Geschlechterunterschiede zugunsten der Jungen beobachtet werden (Singapur, Slowakei, Kanada), in vier weiteren Staaten signifikante Unterschiede zugunsten der Mädchen (Lettland, Japan, Serbien, Bulgarien). Der internationale Durchschnitt beträgt 4 Punkte zugunsten der Mädchen und ist statistisch signifikant. In den Vergleichsgruppen beträgt die mittlere Differenz jeweils 0 Punkte. In Deutschland zeigt sich im Anforderungsbereich *Anwenden* ein nicht signifikanter Geschlechterunterschied von 4 Punkten zugunsten der Jungen. Im Bereich *Anwenden* unterscheiden sich demnach Mädchen und Jungen in Deutschland wie auch in *Reproduzieren* nicht signifikant hinsichtlich ihrer Testleistungen.

Im Anforderungsbereich *Problemlösen* erreichen Mädchen in vielen Teilnehmerstaaten im Durchschnitt bessere Testleistungen als Jungen. In neun Teilnehmerstaaten erreichen sie signifikant höhere Testleistungen als Jungen. Jungen hingegen können in keinem der Staaten einen statistisch bedeutsamen Vorsprung gegenüber den Mädchen erzielen. In der Mehrzahl der Staaten sind keine signifikanten Geschlechterunterschiede zu beobachten. Im internationalen Mittel zeigt sich eine signifikante Geschlechterdifferenz von 7 Skalenwertpunkten zugunsten der Mädchen. In den Vergleichsgruppen beträgt der Vorsprung der Mädchen jeweils 4 Punkte. Alle drei Differenzen sind als signifikant einzustufen. In Deutschland kann also für keinen Anforderungsbereich ein Unterschied zwischen Mädchen und Jungen beobachtet werden.

8.5.4 Trends in geschlechterspezifischen Leistungsunterschieden in den Naturwissenschaften

Inwieweit sich die Leistungen von Mädchen und Jungen in den Naturwissenschaften in den Jahren 2007 bis 2019 entwickelt haben, zeigt Abbildung 8.11. Die Betrachtung von Trends ermöglicht eine Einschätzung der Entwicklung

Abbildung 8.11: Vergleich der Leistungsentwicklungen von Mädchen und Jungen in Naturwissenschaften in TIMSS 2007, 2011, 2015 und 2019 (internationaler Vergleich)□ Keine statistisch signifikanten Unterschiede zwischen 2007 und 2019 ($p > .05$)■ Statistisch signifikante Unterschiede zwischen 2007 und 2019 ($p \leq .05$)▲ = Differenzwerte im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)▼ = Differenzwerte im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)+ = Mittelwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)- = Mittelwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

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

6 = Abweichender Testzeitpunkt

8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

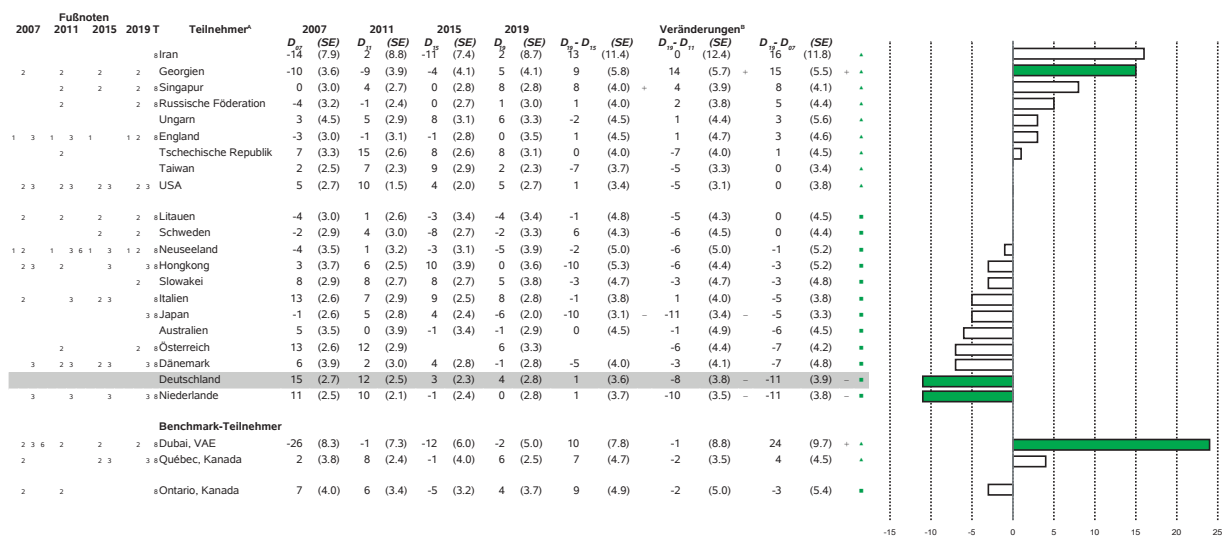
B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

von Geschlechterdisparitäten in den Bildungssystemen. Auch diese Abbildung wurde analog zur Abbildung 8.5 in Abschnitt 8.4.4 erstellt.

Bei Betrachtung der Leistungsentwicklung der Mädchen in den Naturwissenschaften kann in sieben Teilnehmerstaaten (ohne *Benchmark*-Teilnehmer) ein signifikanter, positiver Trend beobachtet werden. Im aktuellen Zyklus konnten Mädchen in Georgien und Litauen ihre Testleistungen in den Naturwissenschaften im Vergleich zu TIMSS 2007 besonders stark steigern. In Georgien erreichten Mädchen in TIMSS 2019 insgesamt 29 Leistungspunkte mehr als noch im Jahr 2007. In Litauen sind es 24 Leistungspunkte. In Hongkong und Italien zeigt sich ein gegenläufiger Trend. Hier erzielten Mädchen in TIMSS 2019 signifikant geringere Testleistungen als in TIMSS 2007. Der Unterschied in Hongkong beträgt 22 Punkte, der in Italien 23 Punkte. In einer Reihe von Teilnehmerstaaten zeichnet sich allerdings keine signifikante Veränderung der Testleistungen im Vergleich von TIMSS 2007 zu TIMSS 2019 ab. In Deutschland kann die Tendenz beobachtet werden, dass die Leistungen der Mädchen in TIMSS 2019 gegenüber vorheriger Testzyklen nominell niedriger ausfallen. Betrug die mittlere Leistung von Mädchen in TIMSS 2007 noch 520 Punkte, so konnten diese ihre Leistungen im Durchschnitt in TIMSS 2011 auf 522 Punkte sowie in TIMSS 2015 auf 527 Punkte steigern. Nach diesem Aufwärtstrend ist nun erstmals eine Abnahme der naturwissenschaftlichen Leistungen zu verzeichnen. Im aktuellen Zyklus beträgt die mittlere Leistung von Mädchen 516 Punkte und nimmt damit im Vergleich zu TIMSS 2015 um 11 Leistungspunkte ab. Damit sinkt der aktuelle Wert erstmals unter das Ausgangsniveau im Jahr 2007. Der Unterschied zwischen TIMSS 2007 und TIMSS 2019 ist jedoch nicht signifikant.

Bei den Jungen zeigt sich in sieben Teilnehmerstaaten ein positiver Trend für die Entwicklung naturwissenschaftlicher Leistungen im Vergleich von TIMSS 2007 zu TIMSS 2019. Besonders deutlich ist dieser in Georgien, wo Jungen in TIMSS 2019 eine im Vergleich zu TIMSS 2007 um 44 Punkte bessere Testleistung erzielen. In vier Staaten kann ein gegenläufiger Trend beobachtet werden. In den Niederlanden, Deutschland, Hongkong und Italien haben die durchschnittlichen Testleistungen von Jungen im Zeitraum von zwölf Jahren signifikant abgenommen. Am stärksten ist dieser negative Trend in Italien ausgeprägt. Dort beträgt die Differenz 28 Leistungspunkte. In Deutschland ist der Trend mit 14 Punkten ebenfalls deutlich ausgeprägt. Hier erreichten Jungen in TIMSS 2007 durchschnittliche Testleistungen in Höhe von 535 Punkten. In den folgenden Testzyklen haben diese Leistungen sukzessive abgenommen. So erreichten Jungen in TIMSS 2011 immerhin noch 534 Punkte, in TIMSS 2011 sank dieser Wert um 4 Punkte auf 530 und im aktuellen Zyklus von TIMSS 2019 um weitere 10 Punkte auf 520 Leistungspunkte. Dieser Wert liegt signifikant unter dem Ausgangswert aus dem Jahr 2007.

Eine Abschätzung zur Entwicklung von Geschlechterunterschieden in den Naturwissenschaften ermöglicht Abbildung 8.12. In drei Staaten (ohne *Benchmark*-Teilnehmer) können signifikante Unterschiede in den Geschlechterdisparitäten zwischen den Zyklen 2007 und 2019 beobachtet werden. Eine deutliche Veränderung hinsichtlich der Ausprägung von Geschlechterunterschieden kann in Georgien beobachtet werden. Hier lag die mittlere Differenz der Leistungsunterschiede zwischen Mädchen und Jungen in TIMSS 2007 bei 10 Punkten zugunsten der Mädchen. In TIMSS 2019 liegt die aktuelle Geschlechterdifferenz bei 5 Leistungspunkten Unterschied zugunsten der Jungen. Damit beträgt der Unterschied in der Leistungsdifferenz von Mädchen und Jungen in Georgien von TIMSS 2007 bis TIMSS 2019 insgesamt 15 Leistungspunkte.

Abbildung 8.12: Vergleich der Leistungsunterschiede in den Naturwissenschaften von Mädchen und Jungen in TIMSS 2007, 2011, 2015 und 2019 (internationaler Vergleich)

- Keine statistisch signifikanten Unterschiede zwischen 2007 und 2019 ($p > .05$)
 ■ Statistisch signifikante Unterschiede zwischen 2007 und 2019 ($p \leq .05$)
 ▲ = Differenzwerte im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)
 ▼ = Differenzwerte im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)
 ■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)
 + = Mittelwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)
 - = Mittelwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)
 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
 6 = Abweichender Testzeitpunkt
 8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019
 A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.
 B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

In Deutschland und den Niederlanden ist die Differenz der Testleistungen von Mädchen und Jungen im aktuellen Zyklus signifikant geringer als in TIMSS 2007. Beide Werte fallen jedoch zugunsten der Jungen aus. So wurde in Deutschland in TIMSS 2007 ein Mittelwertunterschied in den naturwissenschaftlichen Testleistungen von Mädchen und Jungen von 15 Leistungspunkten dokumentiert. Im aktuellen Zyklus beläuft sich diese Differenz auf 4 Punkte. Auch wenn dieser Wert im Vergleich zu TIMSS 2015 stagniert (2015 lag der Unterschied bei 3 Punkten), ist nach wie vor ein positiver Trend für Deutschland zu verzeichnen. Bei genauerer Betrachtung der Leistungsentwicklung von Mädchen und Jungen (siehe Abbildung 8.12) zeigt sich jedoch, dass die Verringerung von Geschlechterunterschieden in den Naturwissenschaften auf der Grundlage absinkender Leistungen der Jungen basiert und nicht vorrangig auf einer Verbesserung der Mädchen.

8.5.5 Geschlechterspezifische Unterschiede in Sachunterrichtsnoten

Analog zu den Darstellungen der schulischen Noten im Fach Mathematik (Tabelle 8.1) weist Tabelle 8.2 die Mittelwerte und Verteilungen der Fachnoten im Sachunterricht aus. Hier erreichen Mädchen im Mittel im Halbjahreszeugnis der Klasse 4 eine 2.2 ($SE = 0.02$), Jungen eine 2.3 ($SE = 0.03$). Dieser Mittelwertunterschied ist statistisch signifikant. Anders als in Mathematik können jedoch keine statistisch signifikanten Unterschiede im Hinblick auf die prozentuale Verteilung von Mädchen und Jungen auf die Noten beobachtet werden.

Tabelle 8.2: Mittelwerte und Verteilungen der Noten im Sachunterricht nach Geschlecht (Anteile der Schülerinnen und Schüler in Prozent)

Sachunterrichtsnote (Halbjahreszeugnis 4. Jahrgangsstufe) ^A												
			1		2		3		4		5	
	M	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)
Mädchen	2.2*	(0.02)	17.7	(1.2)	46.4	(1.6)	27.0	(1.3)	7.8	(0.8)	0.7	(0.2)
Jungen	2.3	(0.03)	15.9	(1.2)	42.8	(1.6)	30.7	(1.3)	8.8	(0.7)	1.5	(0.4)
Gesamt	2.3	(0.02)	16.9	(0.9)	44.6	(1.2)	28.9	(1.0)	8.3	(0.6)	1.1	(0.2)

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

A = Aufgrund zu geringer Fallzahlen kann die Note 6 nicht ausgewertet werden.

* = Unterschied zu Jungen statistisch signifikant ($p \leq .05$)

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

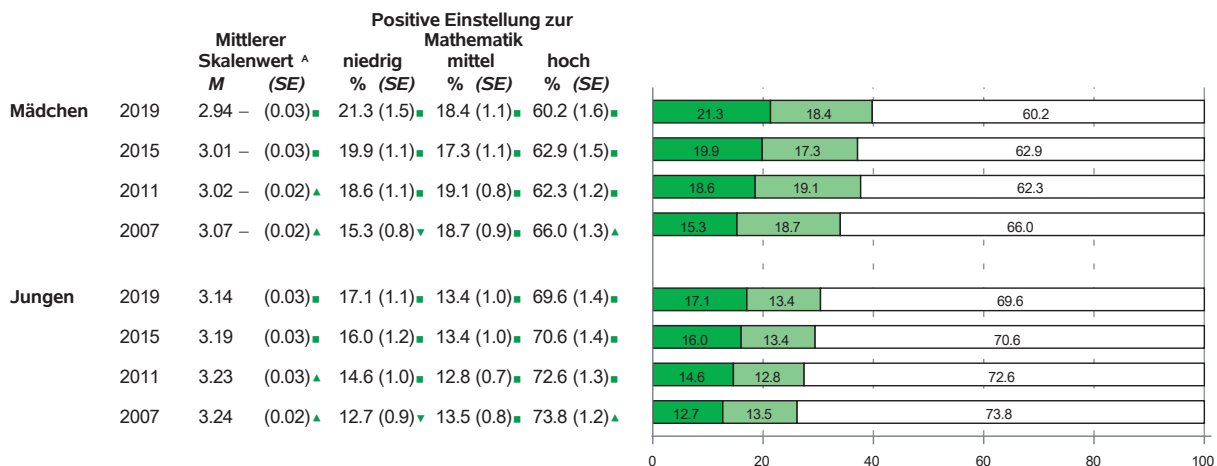
Insgesamt erhalten 64.1 Prozent der Mädchen eine Note im guten oder sehr guten Bereich. Bei den Jungen liegt der Anteil bei 58.7 Prozent. Die Note *mangelhaft* erhalten 0.7 Prozent der Mädchen und 1.5 Prozent der Jungen. Aufgrund geringer Fallzahlen werden die prozentualen Häufigkeiten für die Note *ungenügend* nicht berichtet.

8.6 Ergebnisse zu geschlechterspezifischen Unterschieden in Einstellungen und Selbstkonzepten

Das fachbezogene Fähigkeitsselbstkonzept und die Werte beziehungsweise Einstellungen gegenüber einem Fach stellen wichtige Voraussetzungen für das erfolgreiche Lernen von Schülerinnen und Schülern dar. Im folgenden Abschnitt werden zunächst die Ergebnisse zum Selbstkonzept und zur positiven Einstellung der Grundschülerinnen und Grundschüler in Deutschland gegenüber Mathematik und anschließend gegenüber dem naturwissenschaftlichen Sachunterricht dargestellt. Dabei wird jeweils neben den Mittelwerten für die Mädchen und Jungen aus den Jahren 2007, 2011, 2015 und 2019 auch die Verteilung auf die Gruppen mit niedriger, mittlerer und hoher positiver Einstellung beziehungsweise niedrigem, mittlerem und hohem Fähigkeitsselbstkonzept dargestellt. Eine Dokumentation der eingesetzten Instrumente und der Skalenbildung findet sich in den Kapiteln 3 und 4 in diesem Band.

8.6.1 Geschlechterspezifische Unterschiede in Einstellungen und Selbstkonzepten in Mathematik

In Abbildung 8.13 sind die mittleren Skalenwerte der positiven Einstellung in Mathematik von Mädchen und Jungen für die Jahre 2007, 2011, 2015 und 2019 dargestellt. Je höher der jeweilige Mittelwert ist, desto positiver ist die Einstellung gegenüber Mathematik. Des Weiteren ist jeweils die prozentuale Verteilung der Mädchen und Jungen auf die drei verschiedenen Gruppen *niedrige positive Einstellung zu Mathematik*, *mittlere positive Einstellung zu Mathematik* und *hohe positive Einstellung zu Mathematik* abgebildet.

Abbildung 8.13: Prozentuale Verteilung auf der Skala *positive Einstellung zur Mathematik* in TIMSS 2007, 2011, 2015 und 2019 – Mädchen und Jungen im Vergleich

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

- % der Schülerinnen und Schüler mit niedriger positiver Einstellung zur Mathematik
- % der Schülerinnen und Schüler mit mittlerer positiver Einstellung zur Mathematik
- % der Schülerinnen und Schüler mit hoher positiver Einstellung zur Mathematik
- ▲ = Wert im Vergleich zu 2019 statistisch signifikant höher ($p \leq .05$)
- ▼ = Wert im Vergleich zu 2019 statistisch signifikant niedriger ($p \leq .05$)
- = Unterschied zu 2019 nicht statistisch signifikant ($p > .05$)
- + = Mittlerer Skalenwert im Vergleich zu Jungen statistisch signifikant höher ($p \leq .05$)
- = Mittlerer Skalenwert im Vergleich zu Jungen statistisch signifikant niedriger ($p \leq .05$)
- A = Die Skala umfasst drei Fragen (z.B. Ich mag Mathematik.) mit vierstufigem Antwortformat (1 = Stimme überhaupt nicht zu, ..., 4 = Stimme völlig zu).

IEA: Trends in International Mathematics and Science Study

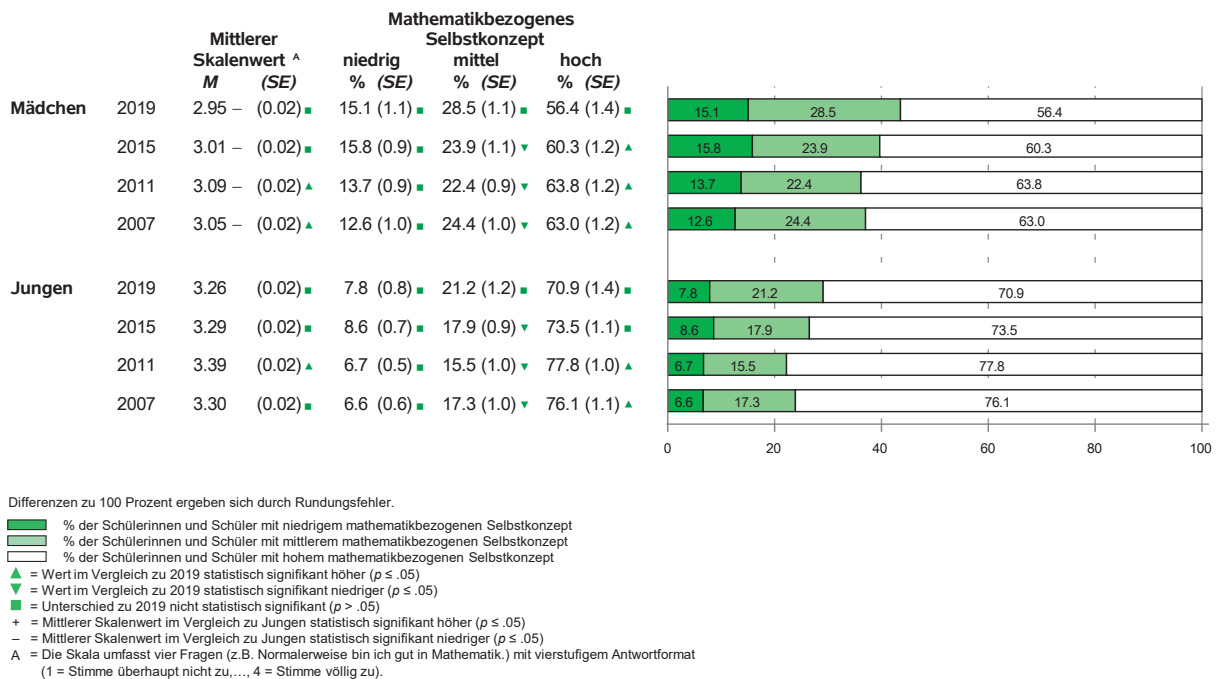
© TIMSS 2019

Es zeigt sich, dass Mädchen und Jungen im Durchschnitt eine positive Einstellung gegenüber Mathematik haben: Über alle Zyklen hinweg berichten mindestens drei Viertel der Schülerinnen und Schüler von einer mittleren bis hohen positiven Einstellung. Verglichen mit den Zyklen 2007 und 2011 zeigen sich jedoch sowohl für Mädchen als auch für Jungen in 2019 im Mittel signifikant niedrigere Werte. Für die Jungen ergab sich im Vergleich zu 2007 eine Differenz von 0.10 Skalenwertpunkten, der Mittelwert der Mädchen sank in diesem Zeitraum um 0.13 auf 2.94 Punkte. Die Differenz in der mittleren positiven Einstellung von Mädchen und Jungen ist wie in den Zyklen zuvor signifikant. Mit einer Effektstärke von $d = .22$ entspricht dieser Unterschied einem kleinen Effekt.

Bei der Betrachtung der prozentualen Verteilung der Schülerinnen und Schüler auf die drei Gruppen lässt sich ein ähnliches Bild für den Vergleich der Zyklen 2007 und 2019 erkennen. Mit einer Differenz von knapp 6 Prozent werden in 2019 signifikant weniger Mädchen und mit etwa 4 Prozent ebenfalls signifikant weniger Jungen der Gruppe *hohe positive Einstellung zur Mathematik* zugeordnet. Gleichzeitig stieg die Zahl der Mädchen und Jungen, die angeben der Gruppe mit niedriger positiver Einstellung zur Mathematik anzugehören, im Vergleich zu 2007 signifikant an. Dabei ist es bei den Mädchen etwas mehr als jede Fünfte, die sich dieser Gruppe zuordnet, bei den Jungen etwas mehr als jeder Sechste.

Abbildung 8.14 können die Informationen zum mathematikbezogenen Selbstkonzept von Viertklässlerinnen und Viertklässlern entnommen werden. Neben den Skalenmittelwerten zum mathematikbezogenen Selbstkonzept von Mädchen und Jungen aus den Jahren 2007, 2011, 2015 und 2019 ist auch die prozentuale Zugehörigkeit der Kinder zu den Gruppen mit niedrigem, mittlerem oder hohem mathematikbezogenen Selbstkonzept dargestellt.

Insgesamt verfügen knapp 85 Prozent der Mädchen und etwa 92 Prozent der Jungen über ein mittleres bis hohes mathematikbezogenes Selbstkonzept. Bei der Betrachtung der Mittelwerte zeigt sich, dass sowohl bei den Mädchen als auch

Abbildung 8.14: Prozentuale Verteilung auf der Skala *mathematikbezogenes Selbstkonzept* in TIMSS 2007, 2011, 2015 und 2019 – Mädchen und Jungen im Vergleich

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

bei den Jungen zwischen 2007 und 2011 ein leichter Anstieg der Werte vorliegt. Diese Entwicklung ist in den nachfolgenden Zyklen jedoch nicht mehr zu beobachten, da seit 2011 das fachbezogene Selbstkonzept der Kinder sinkt. Sowohl der Mittelwert der Mädchen als auch der Mittelwert der Jungen ist in 2019 signifikant niedriger als 2011. Beim Vergleich der Mittelwerte der Mädchen und Jungen zeigt sich, dass der Wert der Mädchen mit 2.95 Punkten im Jahr 2019 um 0.31 niedriger liegt als der der Jungen. Damit ist die Differenz zwischen den Mittelwerten der Mädchen und Jungen wie auch in den vorangegangenen Zyklen signifikant. Der Unterschied im mittleren mathematikbezogenen Selbstkonzept von Mädchen und Jungen ist mit $d = .41$ zwar ebenfalls als klein einzustufen, der Unterschied zwischen den Geschlechterdifferenzen in der positiven Einstellung und dem Selbstkonzept ist jedoch beträchtlich.

Beim Vergleich der Gruppenzugehörigkeit der Mädchen und Jungen zeichnet sich ein ähnliches Bild ab. Während im Jahr 2019 mit 56.4 Prozent im Vergleich zu 2007, 2011 und 2015 signifikant weniger Mädchen zur Gruppe der Kinder mit hohem mathematikbezogenen Selbstkonzept gehören, ist die Anzahl der Schülerinnen, die angeben, ein mittleres mathematikbezogenes Selbstkonzept zu haben, signifikant höher als in den vorangegangenen Zyklen. Bei den Jungen ist eine ähnliche Tendenz erkennbar.

Insgesamt lässt sich ein leichtes Absinken des mathematikbezogenen Selbstkonzeptes und der positiven Einstellung zur Mathematik bei den Schülerinnen und Schülern in Deutschland erkennen. Darüber hinaus gibt es nach wie vor für die positive Einstellung zur Mathematik und im Besonderen für das Selbstkonzept zuungunsten der Mädchen einen signifikanten Unterschied zwischen den Schülerinnen und Schülern.

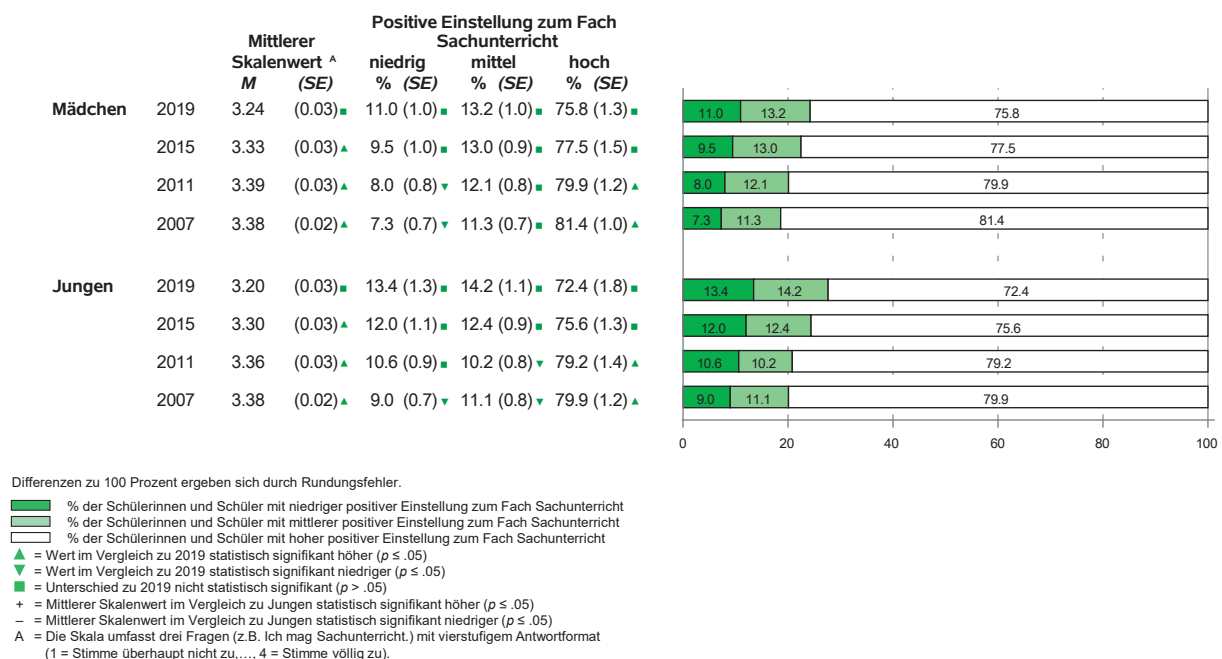
8.6.2 Geschlechterspezifische Unterschiede in Einstellungen und Selbstkonzepten im Sachunterricht

Da die Naturwissenschaften in Deutschland in der Grundschule nicht als eigenständiges Fach, sondern im Rahmen des Sachunterrichts unterrichtet werden, werden in TIMSS die Einstellung und das Selbstkonzept bezogen auf den Sachunterricht erhoben. Dies bedeutet, dass nicht davon ausgegangen werden kann, dass die Schülerinnen und Schüler die naturwissenschaftlichen Inhalte klar von den anderen Inhalten des Sachunterrichts abgrenzen können. Stattdessen müssen die Ergebnisse globaler als positive Einstellung und Selbstkonzept zum Sachunterricht verstanden werden (siehe Kapitel 4 in diesem Band).

In Abbildung 8.15 ist der Mittelwert der positiven Einstellung von Mädchen und Jungen gegenüber dem Sachunterricht für die Jahre 2007, 2011, 2015 und 2019 dargestellt. Es zeigt sich, dass die Schülerinnen und Schüler eine überwiegend positive Einstellung zum Sachunterricht haben. Dabei ergibt sich für die Mädchen seit 2011 ein leicht höherer Mittelwert als für die Jungen. Dieser Unterschied zwischen Mädchen und Jungen ist jedoch in keinem der Zyklen signifikant. Insgesamt ist für die Mädchen seit 2011 und für die Jungen seit 2007 ein Absinken der mittleren positiven Einstellung zum Sachunterricht zu verzeichnen: Während der Mittelwert 2007 für die Schülerinnen und Schüler noch bei 3.38 Skalenwertpunkten lag, ist er 2019 mit einer Differenz von 0.14 beziehungsweise 0.18 – auch im Vergleich zu 2011 und 2015 – signifikant niedriger. Die größte Veränderung in den Mittelwerten kann für den Zeitraum zwischen 2015 und 2019 verzeichnet werden. Hier liegt die Differenz für Mädchen bei 0.09 Punkten und für Jungen bei 0.10 Punkten.

Bei der Betrachtung der ebenfalls in Abbildung 8.15 dargestellten Verteilung der Schülerinnen und Schüler auf die Gruppen mit niedriger, mittlerer und hoher positiver Einstellung, zeigt sich, dass im Jahr 2019 drei von vier Mädchen

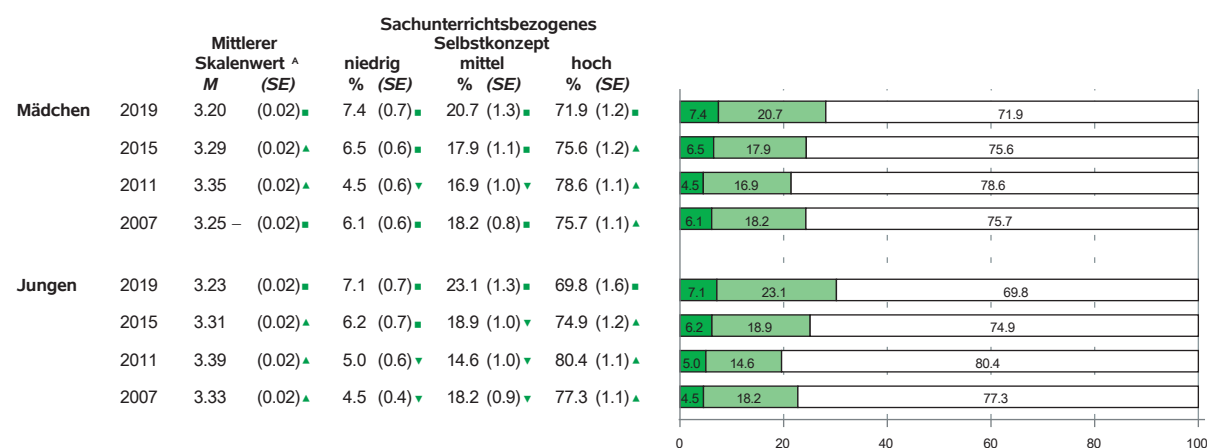
Abbildung 8.15: Prozentuale Verteilung auf der Skala *positive Einstellung zum Fach Sachunterricht* in TIMSS 2007, 2011, 2015 und 2019 – Mädchen und Jungen im Vergleich



der Gruppe der Kinder mit hoher positiver Einstellung zum Sachunterricht zugeordnet werden können. 2007 waren es noch vier von fünf Mädchen. Gleichzeitig steigt die Zahl der Schülerinnen, die der Gruppe mit niedriger positiver Einstellung zum Sachunterricht angehören, seit 2007 an. Für die Jungen zeigt sich eine ähnliche Entwicklung. Die Veränderung in den Gruppen ist dabei – mit einer Ausnahme – beim Vergleich der Werte von 2007 und 2019 signifikant. Beim Vergleich der Mädchen und Jungen lässt sich zudem erkennen, dass sich – anders als in Mathematik – durchgängig prozentual mehr Mädchen als Jungen in der Gruppe *hohe positive Einstellung* befinden. Äquivalent dazu können in allen Zyklen mehr Jungen als Mädchen in die Gruppe *niedrige positive Einstellung* eingeordnet werden.

In Abbildung 8.16 werden zum einen die Skalenmittelwerte zum sachunterrichtsbezogenen Selbstkonzept von Mädchen und Jungen von 2007 bis 2019, zum anderen die prozentuale Zugehörigkeit der Kinder zu den Gruppen mit niedrigem, mittlerem oder hohem sachunterrichtsbezogenen Selbstkonzept dargestellt. Das sachunterrichtsbezogene Selbstkonzept der Grundschülerinnen und Grundschüler in Deutschland kann insgesamt als hoch eingestuft werden. Mehr als 90 Prozent der Mädchen und Jungen geben an, über ein mittleres bis hohes Selbstkonzept zu verfügen. Während zwischen den Jahren 2007 und 2011 die Zahl der Mädchen und Jungen, die sich der Gruppe mit hohem sachunterrichtsbezogenen Selbstkonzept zuordnen lassen, angestiegen ist, ist seit 2011 ein Absinken zu verzeichnen. Seit 2015 gehören dabei mehr Mädchen als Jungen der Gruppe der Kinder mit hohem sachunterrichtsbezogenem Selbstkonzept an. Die prozentuale Verteilung der Jungen in 2019 unterscheidet sich dabei zu deren Ungunsten mit einer Ausnahme im Jahr 2015 über alle Gruppen hinweg signifikant von der Verteilung in den drei vorherigen Zyklen. Dies gilt auch für das mittlere sachunterrichtsbezogene Selbstkonzept der Jungen. Der Mittelwert liegt 2019 bei 3.23 Skalenwertpunkten und ist damit niedriger als 2007, 2011

Abbildung 8.16: Prozentuale Verteilung auf der Skala *sachunterrichtsbezogenes Selbstkonzept* in TIMSS 2007, 2011, 2015 und 2019 – Mädchen und Jungen im Vergleich



Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

■ % der Schülerinnen und Schüler mit niedrigem sachunterrichtsbezogenen Selbstkonzept

■ % der Schülerinnen und Schüler mit mittlerem sachunterrichtsbezogenen Selbstkonzept

■ % der Schülerinnen und Schüler mit hohem sachunterrichtsbezogenen Selbstkonzept

▲ = Wert im Vergleich zu 2019 statistisch signifikant höher ($p \leq .05$)

▼ = Wert im Vergleich zu 2019 statistisch signifikant niedriger ($p \leq .05$)

■ = Unterschied zu 2019 nicht statistisch signifikant ($p > .05$)

+ = Mittlerer Skalenwert im Vergleich zu Jungen statistisch signifikant höher ($p \leq .05$)

– = Mittlerer Skalenwert im Vergleich zu Jungen statistisch signifikant niedriger ($p \leq .05$)

A = Die Skala umfasst vier Fragen (z.B. Normalerweise bin ich gut im Sachunterricht.) mit vierstufigem Antwortformat (1 = Stimme überhaupt nicht zu, ..., 4 = Stimme völlig zu).

und 2015. Der höchste Mittelwert ist für 2011 zu verzeichnen. Ein ähnliches Bild zeigt sich für die Mädchen: Das mittlere sachunterrichtsbezogene Selbstkonzept ist in 2019 mit 3.20 Punkten signifikant niedriger als in 2011 und 2015. Ein signifikanter Unterschied zwischen den Mittelwerten der Mädchen und Jungen wie in 2007 konnte in den nachfolgenden Zyklen nicht mehr nachgewiesen werden.

Insgesamt zeigt sich, dass sowohl die positive Einstellung als auch das Selbstkonzept im Sachunterricht für Mädchen und Jungen höher ausgeprägt ist als in Mathematik. Dabei sind signifikante Geschlechterunterschiede in Mathematik zugunsten der Jungen festzustellen. Im Sachunterricht hingegen zeigen sich geringe, aber nicht signifikante Mittelwertdifferenzen zugunsten der Mädchen. Obwohl in den vergangenen Jahren ein leichtes Absinken für beide Domänen zu verzeichnen ist, verfügen die Grundschülerinnen und Grundschüler in Deutschland nach wie vor über hohe positive Einstellungen und hohe Selbstkonzepte in Mathematik und im Sachunterricht.

8.7 Zusammenfassung

Die in diesem Kapitel berichteten Befunde zu Geschlechterdisparitäten in Mathematik und den Naturwissenschaften geben erste Hinweise auf eine generelle Zunahme von Geschlechterunterschieden. Wie auch in vielen anderen Teilnehmerstaaten und Regionen zeigen sich für Deutschland signifikante Unterschiede in den mathematischen Testleistungen von Mädchen und Jungen. Deutschland liegt mit 10 Punkten Differenz knapp oberhalb der Differenzwerte in den Vergleichsgruppen EU und OECD. Die Betrachtung der Verteilung von Mädchen und Jungen auf die Kompetenzstufen I bis V zeigt, dass Mädchen auf den niedrigeren Kompetenzstufen überrepräsentiert, Jungen hingegen auf den höheren Kompetenzstufen überrepräsentiert sind. Diese Beobachtung konnte bereits in TIMSS 2015 gemacht werden, wobei eine Absicherung gegen den Zufall im letzten Zyklus nicht erfolgen konnte. In TIMSS 2019 ist diese ungleiche Verteilung auf die Kompetenzstufen statistisch signifikant und zeigt dringenden Handlungsbedarf im Abbau von Geschlechterdisparitäten im Fach Mathematik auf. Bei Betrachtung der Inhaltsbereiche zeigt sich ein deutlicher Vorsprung der Jungen im Bereich *Messen und Geometrie* (14 Punkte Differenz) und *Arithmetik* (11 Punkte Differenz). Ein ähnliches Bild zeichnet sich für die kognitiven Anforderungsbereiche ab. Mädchen erreichen in allen drei Bereichen signifikant geringere Testleistungen als Jungen. Besonders deutlich ist der Unterschied im Anforderungsbereich *Reproduzieren* (16 Punkte Differenz) und *Problemlösen* (11 Punkte Differenz). Weitere und zusätzliche Fördermaßnahmen können hier ansetzen und die Kompetenzentwicklung von Mädchen und Jungen in diesen Bereichen unterstützen. Die im aktuellen Zyklus erreichten Testleistungen der Mädchen (516 Punkte) liegen erstmals unter den in TIMSS 2007 erreichten Testleistungen (519 Punkte). Mit Blick auf die erreichten Testleistungen in TIMSS 2011 und TIMSS 2015 deutet sich insgesamt eine negative Entwicklung für die mathematischen Testleistungen von Mädchen an. Aber auch für die Jungen zeichnet sich eine Stagnation des Leistungstrends ab. Erreichten Jungen in TIMSS 2007 durchschnittlich noch 531 Leistungspunkte, so waren es in TIMSS 2011 durchschnittlich 532 Punkte, in TIMSS 2015 nur noch 524 Punkte und in TIMSS 2019 noch 526 Punkte. Damit haben sich die Leistungen der Jungen im Zeitraum der Jahre 2007 bis 2019 nicht signifikant verändert. Auch die Betrachtung der Entwicklung von Geschlechterunterschieden

im Trend weist auf eine Stagnation hin. So unterscheiden sich die beobachteten Geschlechterunterschiede in TIMSS 2007 (12 Punkte) nur marginal von der aktuell beobachteten Differenz (10 Punkte).

Im Bereich Naturwissenschaften zeigt sich auf der Basis von deskriptiven Angaben zunächst kein signifikanter Geschlechterunterschied auf der Gesamtskala Naturwissenschaften. Auch die Verteilung auf den Kompetenzstufen lässt keine Annahme zu Unterschieden in den Verteilungen von Mädchen und Jungen auf den Kompetenzstufen zu. Anhand der Inhaltsbereiche wird allerdings deutlich, dass bei differenzierter Betrachtung Geschlechterunterschiede in den Bereichen *Biologie* (zugunsten der Mädchen) und *Geografie* (zugunsten der Jungen) bestehen. Auch bei TIMSS 2015 konnten Unterschiede im Bereich *Geografie* beobachtet werden, lediglich nominelle Unterschiede lagen hingegen im Inhaltsbereich *Biologie* vor. Demnach konnten Mädchen hier ihren Vorsprung (trotz absinkender Leistung) ausbauen. Keine Unterschiede gibt es hingegen bei der Betrachtung der Testleistungen getrennt nach Anforderungsbereichen. Hier zeigte sich in TIMSS 2015 noch ein Vorsprung der Jungen im Bereich *Reproduzieren*. In Bezug auf die Leistungsentwicklung im Trend kann zunächst beobachtet werden, dass sich die naturwissenschaftlichen Leistungen von Mädchen im Vergleich zu TIMSS 2007 nicht signifikant verändert haben. Bei den Jungen hingegen kann ein signifikanter Unterschied zwischen den Testleistungen in TIMSS 2007 und TIMSS 2019 beobachtet werden. Im berichteten Zeitraum haben die Leistungen von Jungen um 14 Punkte abgenommen. Seit TIMSS 2007 zeigt sich somit ein negativer Trend in der Leistungsentwicklung der Jungen, wobei der aktuelle Testwert nun erstmals signifikant unter den Ausgangswert im Jahr 2007 fällt. Der Befund, wonach sich die aktuell beobachteten Geschlechterunterschiede in den Naturwissenschaften im Vergleich zu TIMSS 2007 signifikant verringert haben, muss dementsprechend vor dem Hintergrund der deutlich abfallenden Testleistungen der Jungen interpretiert werden.

Neben der Betrachtung von geschlechterbezogenen Leistungsunterschieden in Mathematik und den Naturwissenschaften ist für die weitere schulische und spätere berufliche Laufbahn von Schülerinnen und Schülern bedeutsam, welchen subjektiven Wert sie diesen Fächern beimessen (siehe Abschnitt 8.6). Auch in TIMSS 2019 wurden Einstellungen und fachbezogene Selbstkonzepte von Viertklässlerinnen und Viertklässlern in Mathematik und im Sachunterricht erhoben. Über alle Zyklen hinweg berichten mindestens drei Viertel der Schülerinnen und Schüler von einer mittleren bis hohen positiven Einstellung zu Mathematik. Zudem berichten mindestens 85 Prozent von einem mittleren bis hohen Selbstkonzept. Im Vergleich von TIMSS 2019 zu TIMSS 2007 beziehungsweise TIMSS 2011 können für Mädchen und Jungen jedoch signifikant niedrigere Mittelwerte für die Einstellung zu Mathematik und das mathematikbezogene Selbstkonzept festgestellt werden. In allen vier Zyklen können zudem signifikante Mittelwertunterschiede in diesen Bereichen zugunsten der Jungen beobachtet werden. Insgesamt kann ein leichtes Absinken des mathematikbezogenen Selbstkonzepts und der positiven Einstellung zu Mathematik in Deutschland beobachtet werden.

In TIMSS 2019 verfügt ein Großteil der Viertklässlerinnen und Viertklässler (85% bis 90%) über eine mittlere bis hohe sachunterrichtsbezogene positive Einstellung beziehungsweise ein mittleres bis hohes Selbstkonzept. Bei Betrachtung der Einstellungen der Mädchen zum Sachunterricht können seit TIMSS 2011 minimale, nicht signifikante Vorsprünge für die Mädchen beobachtet werden. Für das sachunterrichtsbezogene Selbstkonzept zeigt sich ein klei-

ner Vorsprung zugunsten der Jungen, der jedoch nur in TIMSS 2007 statistisch signifikant war. Hinsichtlich der positiven Einstellung zum Sachunterricht befinden sich prozentual mehr Mädchen in der Gruppe mit hoher positiver Einstellung und weniger Mädchen in der Gruppe mit niedriger positiver Einstellung als Jungen. Für das sachunterrichtsbezogene Selbstkonzept befinden sich seit TIMSS 2015 mehr Mädchen als Jungen in der Gruppe mit hohem Selbstkonzept. Auch hier ist seit TIMSS 2007 und TIMSS 2011 ein negativer Trend beziehungsweise eine negative Entwicklung für das sachunterrichtsbezogene Selbstkonzept und die positive Einstellung zum Sachunterricht bei Mädchen und Jungen zu verzeichnen.

Im Vergleich der beiden Domänen Mathematik und Sachunterricht zeigt sich, dass sowohl die positive Einstellung zum Sachunterricht als auch das sachunterrichtsbezogene Selbstkonzept für Mädchen und Jungen höher ausgeprägt ist als die Einstellung zu Mathematik und das mathematikbezogene Selbstkonzept. Dabei sind signifikante Geschlechterunterschiede in Mathematik zugunsten der Jungen festzustellen. Im Sachunterricht hingegen zeigen sich geringe, aber nicht signifikante Mittelwertsunterschiede zugunsten der Mädchen. Obwohl in den vergangenen Jahren ein leichtes Absinken für beide Domänen zu verzeichnen ist, verfügen die Grundschülerinnen und Grundschüler in Deutschland nach wie vor über hohe positive Einstellungen und hohe Selbstkonzepte in Mathematik und im Sachunterricht.

Insgesamt zeigen sich deutliche Befunde und in einigen Bereichen Tendenzen, wonach Geschlechterdisparitäten in Mathematik und in den Naturwissenschaften im aktuellen Zyklus zugenommen haben. Dieser Befund geht einher mit der Beobachtung, dass die Leistungen sowie Einstellungen und Selbstkonzepte seit TIMSS 2007 beziehungsweise 2011 einem negativen Trend folgen. Da es sich lediglich um deskriptive Befunde handelt, die eine Abschätzung der praktischen Relevanz der beobachteten Unterschiede ohne Bezugnahme auf weitere bedeutende Hintergrundmerkmale verbieten, ist eine multikriteriale Betrachtung dringend angeraten. Inwieweit die beobachteten Entwicklungen und Unterschiede im Zusammenhang mit weiteren Kovariaten wie etwa der Herkunft und dem Anteil an Viertklässlerinnen und Viertklässlern mit Förderbedarf stehen, wird in Kapitel 12 in diesem Band eingehend beleuchtet. Vorweggenommen werden soll an dieser Stelle der Befund, dass sich nach Kontrolle weitere Kovariaten nach wie vor ein signifikanter Einfluss des Geschlechts auf die mathematischen und naturwissenschaftlichen Leistungen von Viertklässlerinnen und Viertklässlern zeigt (siehe Kapitel 12 in diesem Band). Aus diesem Grund ist die Betrachtung von Geschlechterunterschieden im Kontext von Schulleistungsstudien nach wie vor unverzichtbar. Letztendlich deuten die in diesem Kapitel berichteten Befunde ausdrücklich auf einen dringenden Handlungsbedarf hinsichtlich des Abbaus von Geschlechterdisparitäten hin, insbesondere da sich die Befunde für Mädchen und Jungen – nicht zuletzt aufgrund personen- und einstellungsbezogener Merkmale – (siehe Kapitel 12 in diesem Band) deutlich negativer darstellen als noch in den Zyklen zuvor. Zudem sei darauf hingewiesen, dass das Geschlecht in TIMSS 2019 lediglich binär (weiblich/männlich) erfasst wurde, was der Vergleichbarkeit und Konsistenz zu den vorherigen Berichtsbänden geschuldet ist. Eine differenzierte Betrachtung insbesondere unter Berücksichtigung des Geschlechterrollenselbstkonzepts und unter Berücksichtigung der Verschränkung weiterer disparitätserzeugender Merkmale wäre wünschenswert, kann aber auf der Grundlage der erhobenen nationalen und internationalen Daten in TIMSS nicht erfolgen.

Literatur

- Arden, R. & Plomin, R. (2006). Sex differences in variance of intelligence across childhood. *Personality and Individual Differences*, 41, 39–48.
<https://doi.org/10.1016/j.paid.2005.11.027>
- Baye, A. & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-scale Assessments in Education*, 4, 541.
<https://doi.org/10.1186/s40536-015-0015-x>
- Bergold, S., Wendt, H., Kasper, D. & Steinmayr, R. (2017). Academic competencies: Their interrelatedness and gender differences at their high end. *Journal of Educational Psychology*, 109(3), 439–449. <https://doi.org/10.1037/edu0000140>
- Brehl, T., Wendt, H. & Bos, W. (2012). Geschlechtsspezifische Unterschiede in mathematischen und naturwissenschaftlichen Kompetenzen. In W. Bos, H. Wendt, O. Köller & C. Selzer (Hrsg.), *TIMSS 2011: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 203–230). Münster: Waxmann.
- Bundesagentur für Arbeit. (August 2019). *Berichte: Blickpunkt Arbeitsmarkt – MINT-Berufe: Statistik der Bundesagentur für Arbeit*. Verfügbar unter: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Statistiken/Themen-im-Fokus/Berufe/Generische-Publikationen/Broschuere-MINT.pdf?__blob=publicationFile
- Ceci, S. J., Williams, W. M. & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, 135, 218–261. <https://doi.org/10.1037/a0014412>
- Eccles, J. S. (1994). Understanding women's educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *Psychology of Women Quarterly*, 18(4), 585–609. <https://doi.org/10.1111/j.1471-6402.1994.tb01049.x>
- Eccles, J. S. & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, 21, 215–225. <https://doi.org/10.1177/0146167295213003>
- Eccles, J. S. & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Fiske, S. T. & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, 23, 1–74.
[https://doi.org/10.1016/S0065-2601\(08\)60317-2](https://doi.org/10.1016/S0065-2601(08)60317-2)
- Gray, H., Lyth, A., McKenna, C., Stothard, S., Tymms, P. & Copping, L. (2019). Sex differences in variability across nations in reading, mathematics and science: A meta-analytic extension of Baye and Monseur (2016). *Large-scale Assessments in Education*, 7, 219. <https://doi.org/10.1186/s40536-019-0070-9>
- Halpern, D. F. (2014). It's complicated-in fact, it's complex: Explaining the gender gap in academic achievement in science and mathematics. *Psychological Science in the Public Interest*, 15, 72–74. <https://doi.org/10.1177/1529100614548844>
- Hannover, B. & Kessels, U. (2002). Monoedukativer Anfangsunterricht in Physik in der Gesamtschule. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 34, 201–215. <https://doi.org/10.1026//0049-8637.34.4.201>
- Hannover, B. & Kessels, U. (2011). Sind Jungen die neuen Bildungsverlierer? Empirische Evidenz für Geschlechterdisparitäten zuungunsten von Jungen und Erklärungsansätze. *Zeitschrift für Pädagogische Psychologie*, 25, 89–103.
<https://doi.org/10.1024/1010-0652/a000039>
- Hedges, L. V. & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41–45.
<http://dx.doi.org/10.1126/science.7604277>
- Helbig, M. (2010). Sind Lehrerinnen für den geringeren Schulerfolg von Jungen verantwortlich? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62, 93–111.
<https://doi.org/10.1007/s11577-010-0095-0>
- Heyder, A., Kessels, U. & Steinmayr, R. (2017). Explaining academic-track boys' underachievement in language grades: Not a lack of aptitude but students' motivational beliefs and parents' perceptions? *The British Journal of Educational Psychology*, 87(2), 205–223. <https://doi.org/10.1111/bjep.12145>

- Holder, K. & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards perspective. *Social Psychology of Education, 20*, 471–490. <https://doi.org/10.1007/s11218-017-9384-z>
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581–592. <https://doi.org/10.1037/0003-066X.60.6.581>
- Kessels, U. & Heyder, A. (2018). Geschlechtsunterschiede. In D. H. Rost, J. R. Sparfeldt & S. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (5., überarbeitete und erweiterte Aufl., S. 209–217). Weinheim: Beltz.
- Kessels, U., Heyder, A., Latsch, M. & Hannover, B. (2014). How gender differences in academic engagement relate to students' gender identity. *Educational Research, 56*, 220–229. <https://doi.org/10.1080/00131881.2014.898916>
- Lauermann, F., Meißner, A. & Steinmayr, R. (2020). Relative importance of intelligence and ability self-concept in predicting test performance and school grades in the math and language arts domains. *Journal of Educational Psychology, 112*, 364–383. <https://doi.org/10.1037/edu0000377>
- Lazarides, R. & Lauermann, F. (2019). Gendered paths into STEM-related and language-related careers: Girls' and boys' motivational beliefs and career plans in math and language arts. *Frontiers in Psychology, 10*, 1243. <https://doi.org/10.3389/fpsyg.2019.01243>
- Lindberg, S. M., Hyde, J. S., Petersen, J. L. & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*, 1123–1135. <https://doi.org/10.1037/a0021276>
- Machin, S. & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science, 322*, 1331–1332. <http://dx.doi.org/10.1126/science.1162573>
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*, 129–149. <https://doi.org/10.3102/00028312023001129>
- Marsh, H. W., Martin, A. J. & Cheng, J. H. S. (2008). A multilevel perspective on gender in classroom motivation and climate: Potential benefits of male teachers for boys? *Journal of Educational Psychology, 100*, 78–95. <https://doi.org/10.1037/0022-0663.100.1.78>
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O. & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*, 397–416.
- McElvany, N., Kessels, U., Schwabe, F. & Kasper, D. (2017). Geschlecht und Lesekompetenz. In A. Hussmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 177–194). Münster: Waxmann.
- Nowell, A. & Hedges, L. V. (1998). Trends in gender differences in academic achievement from 1960 to 1994: An analysis of differences in mean, variance, and extreme scores. *Sex Roles, 39*, 21–43. <http://dx.doi.org/10.1023/A:1018873615316>
- O'Dea, R. E., Lagisz, M., Jennions, M. D. & Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications, 9*, 3777. <https://doi.org/10.1038/s41467-018-06292-0>
- OECD. (Hrsg.). (2016). *PISA 2015 Ergebnisse (Bd. I): Exzellenz und Chancengerechtigkeit in der Bildung*. Bielefeld: Bertelsmann. <https://doi.org/10.3278/6004573w>
- Parker, P., Nagy, G., Trautwein, U. & Lüdtke, O. (2014). Predicting career aspirations and university majors from academic ability and self-concept. In I. Schoon & J. S. Eccles (Hrsg.), *Gender differences in aspirations and attainment: A life course perspective* (S. 224–246). Cambridge: University Press. <https://doi.org/10.1017/CBO9781139128933.015>
- Reilly, D., Neumann, D. L. & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of national assessment of educational progress assessments. *Journal of Educational Psychology, 107*, 645–662. <http://dx.doi.org/10.1037/edu0000012>
- Reinhold, F., Reiss, K., Diedrich, J., Hofer, S. & Heinze, A. (2019). Mathematische Kompetenz in PISA 2018 – aktueller Stand und Entwicklung. In K. Reiss, M. Weis, E. Klieme & O. Köller (Hrsg.), *PISA 2018 – Grundbildung im internationalen Vergleich* (S. 187–201). Münster: Waxmann.

- Reiss, K., Weis, M., Klieme, E. & Köller, O. (Hrsg.). (2019). *PISA 2018 – Grundbildung im internationalen Vergleich*. Münster: Waxmann.
https://doi.org/10.31244/9783830_991007
- Retelsdorf, J., Köller, O. & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction*, 21(4), 550–559.
<https://doi.org/10.1016/j.learninstruc.2010.11.001>
- Schiepe-Tiska, A., Rönnebecke, S. & Neumann, K. (2019). Naturwissenschaftliche Kompetenz in PISA 2018 – aktueller Stand, Veränderungen und Implikationen für die naturwissenschaftliche Bildung in Deutschland. In K. Reiss, M. Weis, E. Klieme & O. Köller (Hrsg.), *PISA 2018 – Grundbildung im internationalen Vergleich* (S. 211–240). Münster: Waxmann.
- Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (2012). *Die Skalen zur Erfassung des schulischen Selbstkonzepts (SESSKO)* (2., überarbeitete und neuronormierte Aufl.). Göttingen: Hogrefe.
- Spinath, B., Eckert, C. & Steinmayr, R. (2014). Gender differences in school success: What are the roles of students' intelligence, personality and motivation? *Educational Research*, 56, 230–243. <https://doi.org/10.1080/00131881.2014.898917>
- Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S. & Haag, N. (Hrsg.). (2017). *IQB-Bildungstrend 2016: Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Münster: Waxmann.
- Statistisches Bundesamt. (2019). *Bildung und Kultur – Allgemeinbildende Schulen*. Wiesbaden: Statistisches Bundesamt.
- Steffens, M. C., Jelenec, P. & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology*, 102, 947–963.
<https://doi.org/10.1037/a0019920>
- Steinmayr, R., Michels, J. & Weidinger, A. F. (2017). *FA(IR)BULOUS – FAIRe BeUrteilung des LeistungsPotenzials von Schülerinnen Und Schülern*. Dortmund: Technische Universität Dortmund.
- Steinmayr, R. & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality*, 22, 185–209. <https://doi.org/10.1002/per.676>
- Steinmayr, R. & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, 19(1), 80–90.
<https://doi.org/10.1016/j.lindif.2008.05.004>
- Steinmayr, R. & Spinath, B. (2010). Konstruktion und erste Validierung einer Skala zur Erfassung subjektiver schulischer Werte (SESSW): Construction and first validation of a scale assessing subjective educational task values. *Diagnostica*, 56, 195–211.
<https://doi.org/10.1026/0012-1924/a000023>
- Steinmayr, R., Weidinger, A. F., Heyder, A. & Bergold, S. (2019). Warum schätzen Mädchen ihre mathematischen Kompetenzen geringer ein als Jungen? – Ein Erklärungsversuch unter Berücksichtigung von Noten, Leistungstests, Lehrer- und Elterneinschätzungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 51, 71–83.
<https://doi.org/10.1026/0049-8637/a000213>
- Steinmayr, R., Weidinger, A. F., Schwinger, M. & Spinath, B. (2019). The importance of students' motivation for their academic achievement – Replicating and extending previous findings. *Frontiers in Psychology*, 10, 1730.
<https://doi.org/10.3389/fpsyg.2019.01730>
- Voyer, D. & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140, 1174–1204. <https://doi.org/10.1037/a0036620>
- Wang, M.-T., Degol, J. & Ye, F. (2015). Math achievement is important, but task values are critical, too: Examining the intellectual and motivational factors leading to gender disparities in STEM careers. *Frontiers in Psychology*, 6, 36.
<https://doi.org/10.3389/fpsyg.2015.00036>
- Weidinger, A. F., Spinath, B. & Steinmayr, R. (2020). The value of valuing math: Longitudinal links between students' intrinsic, attainment and utility values and grades in math. *Motivation Science*. <https://doi.org/10.1037/mot0000179>

- Weidinger, A. F., Steinmayr, R. & Spinath, B. (2017). Math grades and intrinsic motivation in elementary school: A longitudinal investigation of their association. *The British Journal of Educational Psychology*, 87(2), 187–204.
<https://doi.org/10.1111/bjep.12143>
- Weidinger, A. F., Steinmayr, R. & Spinath, B. (2019). Ability self-concept formation in elementary school: No dimensional comparison effects across time. *Developmental Psychology*, 55, 1005–1018. <https://doi.org/10.1037/dev0000695>
- Wendt, H., Bos, W., Selter, C., Köller, O., Schwippert, K. & Kasper, D. (Hrsg.). (2016). *TIMSS 2015: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Wendt, H., Steinmayr, R. & Kasper, D. (2016). Geschlechterunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 257–298). Münster: Waxmann.
- Wigfield, A., Eccles, J. S., Fredricks, J., Simpkins, S., Roeser, R. & Schiefele, U. (2015). Development of achievement motivation and engagement. In R. Lerner, M. Lamb & C. Garcia Coll (Hrsg.), *Handbook of child psychology and developmental science* (Bd. 3, 7. Aufl., S. 657–700). New York, NY: Wiley.
<https://doi.org/10.1002/9781118963418.childpsy316>
- Wigfield, A., Muenks, K. R. & Rosenzweig, E. Q. (2016). Achievement motivation. In H. L. Miller (Hrsg.), *The SAGE encyclopedia of theory in psychology* (S. 1–4). Thousand Oaks: SAGE.

Kapitel 9

Soziale Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen von Viertklässlerinnen und Viertklässlern

Tobias C. Stubbe, Maria Krieg, Christin Beese und Donieta Jusufi

9.1 Einleitung

National für die Bundesrepublik Deutschland ebenso wie international für eine Vielzahl von Staaten ist seit Jahren durch Ergebnisse der empirischen Bildungsforschung belegt, dass eine enge Kopplung zwischen dem sozioökonomischen Status von Familien und dem Bildungserfolg ihrer Kinder besteht. Abhängig von den sozialen, kulturellen und ökonomischen Verhältnissen, in denen Kinder und Jugendliche aufwachsen, ergeben sich für die Schülerinnen und Schüler unterschiedliche Chancen, erfolgreich am Bildungssystem teilzuhaben. Der Anteil der armutsgefährdeten Bevölkerung in Deutschland, der seit 2005 in Westdeutschland (insbesondere in Nordrhein-Westfalen, Berlin und Bremen) angestiegen ist, verweist auf eine größer werdende soziale Ungleichheit (Statistisches Bundesamt, 2016). Welche Konsequenzen sich daraus im Hinblick auf Teilhabechancen, den Bildungsverlauf sowie den Kompetenzerwerb ergeben, ist breit dokumentiert (zuletzt u. a. Bertelsmann Stiftung, 2016; Laubstein, Holz & Seddig, 2016; Mahler & Kölm, 2019; Müller & Ehmke, 2016; Tophoven, Lietzmann, Reiter & Wenzig, 2017; Weis et al., 2019).

Die Frage nach Chancengleichheit im Bildungssystem steht schon lange im Fokus der Aufmerksamkeit (vgl. u. a. Blossfeld & Shavit, 1993). Breites öffentliches ebenso wie wissenschaftliches Interesse erfährt die Thematik insbesondere auch durch die Fülle an empirisch fundiertem Wissen aus wiederholt durchgeführten Schulleistungsstudien, wie der *Internationalen Grundschul-Lese-Untersuchung* (IGLU), der *Trends in International Mathematics and Science Study* (TIMSS), der *International Computer and Information Literacy Study* (ICILS) und dem *Programme for International Student Assessment* (PISA). Ein bildungspolitisches Ziel ist, den Bildungserfolg von Schülerinnen und Schülern von ihrer sozialen Herkunft zu entkoppeln und für mehr Chancengleichheit zu sorgen.

Hinsichtlich der Ungleichverteilung von Bildungschancen lassen sich nach Boudon (1974) zwei Arten von sogenannten Herkunftseffekten unterscheiden. Primäre Herkunftseffekte beschreiben den schulischen Leistungsvorsprung von Kindern aus bildungsnahen Familien gegenüber Kindern aus bildungsfernen Familien. Sekundäre Herkunftseffekte beziehen sich auf Bildungsentscheidungen, die – auch unter Kontrolle der primären Herkunftseffekte – durch die soziale Herkunft der Schülerinnen und Schüler beeinflusst werden. Mögliche Ursachen primärer Herkunftseffekte sind ein Mangel an ökonomischen Ressourcen, beispielsweise für den Kauf von Lernmaterialien, aber auch ein anregungsarmes soziales Umfeld, in dem kulturelle Aktivitäten, wie Theaterbesuche oder gemeinsames Musizieren, selten stattfinden. Sekundäre Herkunftseffekte resultieren zum Beispiel aus Schulformentscheidungen beim Übergang von der Grundschule auf eine weiterführende Schule (siehe Kapitel 11 in diesem Band).

Neben dem in den letzten Jahren zunehmenden Wissen über jene Bereiche, die Chancenungleichheiten im deutschen Bildungssystem potenziell verstärken, lassen sich Maßnahmen und Programme benennen, die Möglichkeiten zur Förderung, Intervention und Innovation im Bildungssystem hinsichtlich der Schaffung gerechterer Teilhabe eröffnen und eröffnet haben. Im Allgemeinen sind dabei vor allem Veränderungen der letzten Jahre im Bereich frühkindlicher Bildung (z.B. Anders & Roßbach, 2013; Hasselhorn & Kuger, 2014) sowie der Ausbau von Ganztagsangeboten im Primar- und Sekundarschulbereich (z.B. Strietholt, Maniti, Berkemeyer & Bos, 2015; vbw, 2013; Wendt, Goy, Walzebug & Valtin, 2016) hervorzuheben. Die Förderung bildungsbenachteiligter Kinder sowohl mit als auch ohne Migrationshintergrund wird insbesondere auch durch eine Vielzahl von Sprachförderprogrammen und -angeboten forciert (z.B. Stanat, Weirich & Radmann, 2012; Valtin & Tarelli, 2014). Entsprechende Projekte und Programme können praxis- und/oder forschungsorientiert sein, sie finden überregional oder auch regional statt (Gogolin, 2020; McMonagle & Hansen, 2020; Titz et al., 2020). Im naturwissenschaftlichen Bereich gewinnen sowohl in der Primar- als auch in der Sekundarstufe I naturwissenschaftliche Profilklassen (z.B. Forscher-, Entdecker- oder MINT-Klassen) zunehmend an Bedeutung, da diese allen teilnehmenden Schülerinnen und Schülern die Möglichkeit bieten, ihre Interessen unabhängig von den familiären Ressourcen, die ihnen zur Verfügung stehen, zu verfolgen (vgl. u.a. Nonte, Haas & Stubbe, 2019; Seidel et al., 2016).

Das vorliegende Kapitel stellt zunächst zentrale Befunde zum Zusammenhang von sozialer Herkunft und Bildungserfolg in Deutschland vor (siehe Abschnitt 9.2) und beschreibt Indikatoren, die im Rahmen von TIMSS zur Erfassung der sozialen Herkunft verwendet werden (siehe Abschnitt 9.3). Gegliedert nach Indikatoren wird anschließend ihr Zusammenhang mit mathematischen und naturwissenschaftlichen Kompetenzen in internationaler (siehe Abschnitt 9.4) sowie, unter Berücksichtigung ausgewählter Indikatoren, in nationaler Perspektive dargestellt (siehe Abschnitt 9.5). Abschließend werden die zentralen Befunde dieses Kapitels zusammenfassend diskutiert (siehe Abschnitt 9.6).

9.2 Soziale Herkunft und Bildungserfolg

Die in den vergangenen Jahrzehnten stetig zu beobachtende Bildungsexpansion in Deutschland hat gesamtgesellschaftlich betrachtet eine kontinuierliche Höherqualifizierung und Akademisierung der Bevölkerung erbracht (Geißler,

2014). Die mit der Bildungsexpansion verknüpften Hoffnungen auf eine Verringerung der Ungleichheiten herkunftsspezifischer Bildungschancen blieben jedoch aus (Blossfeld, Blossfeld & Blossfeld, 2019; Lucas, 2001). Dass soziale Ungleichheiten weiterhin festzustellen sind, belegen empirisch fundiert vor allem die groß angelegten internationalen Schulleistungsstudien wie IGLU, TIMSS, ICILS und PISA, an denen die Bundesrepublik Deutschland seit Ende der 1990er-Jahre regelmäßig teilnimmt (zuletzt Hußmann, Stubbe & Kasper, 2017; Senkbeil, Drossel, Eickelmann & Vennemann, 2019; Stubbe, Schwippert & Wendt, 2016; Weis et al., 2019). Die Ergebnisse zeigen für Deutschland im internationalen Vergleich große soziale Disparitäten. In der ersten PISA-Erhebung 2000 schnitt Deutschland in diesem Aspekt im Vergleich zu allen Teilnehmerstaaten sogar am ungünstigsten ab (Baumert & Schümer, 2001). Befunde wie diese stützen die Aussagen zahlreicher Autorinnen und Autoren, die Bildungsexpansion in Deutschland habe insgesamt nicht zu einem Abbau an sozialen Disparitäten beigetragen (z.B. Becker & Lauterbach, 2004; Blossfeld, 1993; Blossfeld et al., 2019; Schimpl-Neimanns, 2000; Solga, 2005; Vester, 2004, 2005). Das Problem der ungenügenden sozialen Gerechtigkeit beziehungsweise gesellschaftlichen Teilhabechancen ist in Deutschland damit nach wie vor akut (z.B. vbw, 2017).

Während sich bei PISA im Sekundarschulbereich im Zeitraum von 2000 bis 2018 der Zusammenhang zwischen sozialer Herkunft und den erreichten Kompetenzen leicht verringerte (Weis et al., 2019), zeigt sich für den Grundschulbereich in Deutschland nach wie vor eine eher überdurchschnittliche Kopplung zwischen sozialer Herkunft und dem Bildungserfolg der Schülerinnen und Schüler am Ende der vierten Jahrgangsstufe, die im Zeitverlauf relativ stabil ausfällt. So wurde im Rahmen von TIMSS 2015 berichtet, dass hinsichtlich des Ausmaßes sozialer Disparitäten keine signifikanten Veränderungen seit TIMSS 2007 festzustellen waren (Stubbe et al., 2016). Ebenso zeigten die Ergebnisse von IGLU 2001 bis IGLU 2011 keine signifikanten Veränderungen. In IGLU 2016 fielen die sozialen Disparitäten hingegen signifikant höher aus (Hußmann et al., 2017).

Der regelmäßig vom *Institut zur Qualitätsentwicklung im Bildungswesen* (IQB) durchgeführte IQB-Bildungstrend (bis 2012: Ländervergleich) ermöglicht es, den Blick auf soziale Disparitäten im Vergleich der Länder der Bundesrepublik Deutschland zu richten. Gemessen am sozialen Gradienten¹ zeigten die Befunde des IQB-Ländervergleichs 2012 für das Fach Mathematik in Deutschland einen Wert von 40 (Kuhl, Siegle & Lenski, 2013). Im Vergleich zu 2012 hat sich im Jahr 2018 der soziale Gradient im Fach Mathematik auf Bundesebene praktisch nicht verändert (39 Punkte). Auch die Daten von 2018 verweisen folglich auf statistisch signifikante Zusammenhänge zwischen der sozialen Herkunft und den Kompetenzen der Schülerinnen und Schüler in allen Ländern der Bundesrepublik Deutschland (Mahler & Kölm, 2019). Analog sind die sozialen Gradienten für den Kompetenzbereich Fachwissen in den Fächern Biologie, Chemie sowie Physik in allen Ländern statistisch signifikant und die erreichten Kompetenzen von Schülerinnen und Schülern daher substanziell mit dem sozioökonomischen Status der Familien verknüpft. Zwischen den Jahren 2012

1 Der soziale Gradient ist der Regressionskoeffizient (b_1) einer Regression mit der jeweiligen Kompetenz als abhängiger Variable und dem z-standardisierten HISEI (siehe Abschnitt 9.3) als unabhängiger Variable. Abweichend davon wird im Rahmen von TIMSS und IGLU der unstandardisierte HISEI zur Berechnung des sozialen Gradienten verwendet.

und 2018 sind die sozialen Gradienten in den drei Fächern Biologie, Chemie und Physik sowohl bundesweit als auch in den einzelnen Ländern weitestgehend stabil geblieben, sodass sich keine signifikante Veränderung der sozialen Disparitäten feststellen lässt (Mahler & Kölm, 2019).

9.3 Indikatoren der sozialen Herkunft

Es gibt unterschiedliche Ansätze und Instrumente, das Konstrukt soziale Herkunft empirisch zu erfassen. Zentral bei der Analyse von sozialen Disparitäten ist die Frage der Validität, also der Gültigkeit dessen, wie der sozioökonomische Status von Familien (oder auch andere Aspekte von sozialer Herkunft) in Deutschland und im internationalen Vergleich gemessen wird. Erhebungsinstrumente, die sich in der empirischen Bildungsforschung zur Erfassung der sozialen Herkunft von Schülerinnen und Schülern etabliert haben, basieren auf der theoretischen Grundlage der soziologischen Arbeiten von Pierre Bourdieu (1983). Bourdieu unterscheidet ökonomisches, kulturelles und soziales Kapital und beschreibt damit verbundene Ressourcen, auf die Familien zurückgreifen können, um in bestimmten Kontexten, wie dem schulischen, erfolgreich handeln zu können.

Als ökonomisches Kapital beschreibt Bourdieu die monetären und materiellen Ressourcen einer Familie. Das kulturelle Kapital beinhaltet verschiedene Aspekte: (1) erworbenes Wissen und Fähigkeiten einer Person (inkorporiertes Kulturkapital), (2) kulturelle Besitztümer, wie Musikinstrumente oder Bücher, die nur von Personen genutzt werden können, die über das notwendige inkorporierte Kulturkapital verfügen (objektiviertes Kulturkapital) und (3) formale schulische und akademische Titel (institutionalisiertes Kulturkapital). Mit sozialem Kapital werden nach Bourdieu die Beziehungsnetzwerke der Menschen beschrieben – also die Menge aller Personen, zu denen soziale Beziehungen bestehen. Dabei bestimmen einerseits die Größe dieses sozialen Netzwerkes und andererseits die Kapitalausstattung der Personen innerhalb des Netzwerkes über das Ausmaß des sozialen Kapitals. Coleman (1988), auf den zur Beschreibung des sozialen Kapitals ergänzend zu Bourdieu Bezug genommen werden kann, grenzt das soziale Kapital außerhalb der Familie vom sozialen Kapital innerhalb der Familie ab und betont dessen Wichtigkeit für den Erwerb von schulischen Kompetenzen. Generell gilt: Jede Kapitalform kann in eine andere transformiert werden. Unter Umständen kann dies mit Transformationskosten (z. B. zu investierende Zeit) verbunden sein. Dies ist beispielsweise dann der Fall, wenn Familien ihre Kinder im Verlauf ihrer Bildungsbiografien zur Erreichung von (Hoch-) Schulabschlüssen (institutionalisiertes Kulturkapital) unterstützen, indem ökonomisches (z. B. Investition in Nachhilfeunterricht), kulturelles (z. B. Unterstützung durch Familienangehörige bei Schularbeiten) und soziales (z. B. Unterstützung durch befreundete Lehrkräfte) Kapital genutzt werden.

Nicht von Bourdieu explizit aufgegriffen, aber eng an die Kapitalformen und Transformationspotenziale von Kapitalformen gebunden, ist der Aspekt der Armut beziehungsweise *Armutsgefährdung*. Diese wird über das Einkommen von Haushalten zusammen mit Informationen zur Haushaltsgröße berechnet. Als armutsgefährdet gelten nach Definition der EU (Bardone & Guio, 2005) Haushalte, die zum jeweiligen Zeitpunkt über weniger als 60 Prozent des medianen Nettoäquivalenzeinkommens in einem Staat verfügen. Dazu muss zunächst die Äquivalenzgröße der einzelnen Haushalte erhoben werden. Dies geschieht, indem bei zusammenlebenden Gemeinschaften ein Erwachsener ein Gewicht von

1.0 erhält, jedes weitere Haushaltsmitglied, das mindestens 14 Jahre alt ist, ein Gewicht von 0.5 und jedes Kind, das maximal 13 Jahre alt ist, ein Gewicht von 0.3. Ein Haushalt mit zwei Erziehungsberechtigten und zwei Kindern im Alter von acht und zehn Jahren hat also beispielsweise eine Äquivalenzgröße von 2.1. Das mediane Nettoäquivalenzeinkommen betrug im Jahr 2018 (für 2019 lagen bei Berichtslegung noch keine Daten vor) in Deutschland 22 713 Euro. Daraus ergibt sich eine Armutsgefährdungsgrenze (60% des Medians) von 13 628 Euro pro Jahr beziehungsweise 1 136 Euro pro Monat. Die Beispielfamilie mit einer Äquivalenzgröße von 2.1 würde also als armutsgefährdet gelten, wenn ihr verfügbares Einkommen monatlich unterhalb von 2 385 Euro liegt. Die Armutsgefährdungsquote lag in Deutschland 2018 bei 16.0 Prozent (DESTATIS, 2020). Dem fünften Armuts- und Reichtumsbericht des *Bundesministeriums für Arbeit und Soziales* (BMAS) zufolge, liegt die Armutsgefährdungsquote von Kindern in Deutschland derzeit bei 64 Prozent, wenn kein erwachsenes Haushaltsmitglied erwerbstätig ist. Arbeiten beide Erziehungsberechtigten, eine/r davon in Vollzeit, verringert sich die Armutsgefährdungsquote auf 5 Prozent (BMAS, 2017, S. 253).

Empirisch besteht folglich ein enger Zusammenhang zwischen Armut und anderen Indikatoren der sozialen Herkunft. Zugleich bietet der Aspekt der *Armutsgefährdung* zusätzliche Informationen zur sozialen Lage von Familien. So ist zum Beispiel bekannt, dass Familien in Deutschland zwar von Armut bedroht sein können, obwohl sie über ein (relativ) hohes Kultur- und Sozialkapital verfügen (z. B. Bosch & Kalina, 2007; Butterwegge, 2000).

Zur Erfassung der sozialen Herkunft von Schülerinnen und Schülern haben sich in der empirischen Bildungsforschung in den letzten Jahren verschiedene Indikatoren bewährt. Nicht immer lassen sich diese eindeutig einer Kapitalform zuordnen, wie etwa dem ökonomischen oder dem kulturellen Kapital. Ein Beispiel dafür sind die höchsten Bildungsabschlüsse der Erziehungsberechtigten. Sie gelten als ein Indikator für das institutionalisierte und inkorporierte Kulturkapital, sagen aber zugleich etwas über die mögliche finanzielle Lage einer Familie (ökonomisches Kapital) aus, weil höhere Bildungsabschlüsse tendenziell mit einem höheren Einkommen verbunden sind. Auch ist festzustellen, dass das soziale Kapital – wenn überhaupt – zumeist nur rudimentär erfasst wird; so etwa hinsichtlich der Interaktionen zwischen Kind und Erziehungsberechtigten innerhalb der Familie. Dass dies aus theoretischer Perspektive nicht ausreichend ist, wurde bereits vielfach kritisiert (z. B. Bos, Stubbe & Buddeberg, 2010; Walzebug, 2015). Wie schon im Rahmen von TIMSS 2015 und IGLU 2016 wurden bei TIMSS 2019 daher weitere Instrumente zur Erfassung des sozialen Kapitals im Rahmen der nationalen Ergänzung der Elternfragebogenerhebung eingesetzt (Lorenz & Stubbe, 2017).

Geradezu als ‚Klassiker‘ im Rahmen der empirischen Bildungsforschung ist die *Anzahl der im Haushalt vorhandenen Bücher* zu bezeichnen. Die Variable erfasst primär das objektivierte Kulturkapital und liefert damit auch einen Hinweis auf das in einer Familie zur Verfügung stehende ökonomische Kapital. Indirekt misst die Variable zudem das inkorporierte Kulturkapital, da ein Zusammenhang zwischen dem Bildungsniveau einer Familie und der Anzahl der Bücher, die diese besitzt (Bildungsnähe der Familie), besteht. Die Frage nach den Büchern im Haushalt weist für Schüler- und Elternbefragungen gute Messeigenschaften auf und ist einfach und ökonomisch in der Durchführung (siehe aktuell Schwippert, 2019). Auch im Zeitalter zunehmender Digitalisierung hat sich die Anzahl der Bücher im Haushalt als besonders aussagekräftiger Indikator bewährt.

Angaben zu weiteren Besitztümern im Haushalt liefern zudem Informationen zum ökonomischen beziehungsweise kulturellen Kapital von Familien (z.B. Anzahl an Autos beziehungsweise Vorhandensein von Musikinstrumenten oder Tageszeitungen).

Ein weiterer direkter Indikator zur Erfassung des kulturellen Kapitals in einer Familie ist die Frage nach dem höchsten Bildungsniveau der Erziehungsberechtigten. In TIMSS wird dieses Merkmal mithilfe der internationalen Bildungsskala *International Standard Classification of Education* (ISCED) erhoben, die von der UNESCO entwickelt wurde (Schroedter, Lechert & Lüttinger, 2006; UNESCO, 2003).

Neben dem Bildungsniveau der Erziehungsberechtigten ist ihre berufliche Tätigkeit von Bedeutung. Der Berufsstatus von Personen liefert Informationen über das kulturelle und ökonomische Kapital, da einerseits gewisse Bildungsniveaus notwendig für die Ausübung bestimmter Berufe sind und andererseits die Einkommensverhältnisse durch bestimmte Berufe meist klar definiert sind. Der Beruf wird offen erfragt und mithilfe der Nominalskala *International Standard Classification of Occupations* (ISCO) kodiert. Dabei wird jedem Beruf ein vierstelliger Code zugewiesen, der allerdings noch keine Aussagen über den Status eines Berufes erlaubt (Hoffmann, 2003). Diese Codes lassen sich dann in die verschiedenen Indizes des Berufsstatus überführen. Alle zwanzig Jahre werden die ISCO-Codes aktualisiert (zuletzt 2008). In dem TIMSS-Erhebungszyklus 2007 wurde noch der ISCO-88 genutzt, seit TIMSS 2011 der ISCO-08.

Ausgehend von den ISCO-Codes werden zwei Indizes des Berufsstatus der Erziehungsberechtigten gebildet. Sie basieren auf verschiedenen theoretischen Annahmen und weisen unterschiedliche statistische Eigenschaften auf:

- Der *International Socio-Economic Index of Occupational Status* (ISEI) ist ein international vergleichbarer sozioökonomischer Index, der das Einkommen und das Bildungsniveau berücksichtigt und angibt, wie gut ein Beruf geeignet ist, um die Ausbildung einer Person in Einkommen umzuwandeln. Mittels des ISEI werden Berufe hinsichtlich ihres sozioökonomischen Status auf einer Skala von 10 bis 90 Punkten bewertet. Niedrige Werte kennzeichnen Berufe mit einem geringen sozioökonomischen Status (z.B. Hilfsarbeiter in der Land- und Forstwirtschaft); hohe Werte charakterisieren einen hohen sozioökonomischen Status der Berufe (z.B. Richter) (Ganzeboom, de Graaf & Treiman, 1992; Ganzeboom & Treiman, 1996, 2010). In der empirischen Bildungsforschung wird häufig der höchste ISEI im Haushalt (HISEI) genutzt, um den sozioökonomischen Status der Familie abzubilden.
- Erikson, Goldthorpe und Portocarero (1979) teilen Berufe in distinkte Gruppen ein (EGP-Klassen), die sich nach Art der Tätigkeit, der Weisungsbefugnis, Art der Beschäftigung (selbstständig oder angestellt) sowie der erforderlichen Qualifikation unterscheiden. In Hinblick auf die unterschiedlichen Dimensionen zeichnen sich die Gruppen durch eine hohe Homogenität innerhalb der EGP-Klasse und eine hohe Heterogenität zwischen den EGP-Klassen aus. Bei den EGP-Klassen handelt es sich um eine Nominalskala. Häufig werden die EGP-Klassen weiter zur *service class* (EGP I und II), *intermediate class* (EGP III und IV) und *working class* (EGP V, VI und VII) zusammengefasst.

Schließlich liefert das Einkommen von Haushalten direkte Angaben über das ökonomische Kapital. Eine weitere Variable, die aus dem Haushaltseinkommen

und der Haushaltsgröße abgeleitet werden kann, ist die *Armutsgefährdung* von Familien (siehe Beginn dieses Abschnitts).

Die vorgestellten Indikatoren wurden im Rahmen von TIMSS 2019 in Deutschland erhoben. International kamen deutlich weniger Variablen für den sozioökonomischen Status von Familien zum Einsatz. Die nachfolgenden Befunde gliedern sich daher in internationale und nationale Analysen: In Abschnitt 9.4 werden zunächst Befunde im internationalen Vergleich berichtet und anschließend im Abschnitt 9.5 vertiefend soziale Disparitäten im deutschen Bildungssystem analysiert.

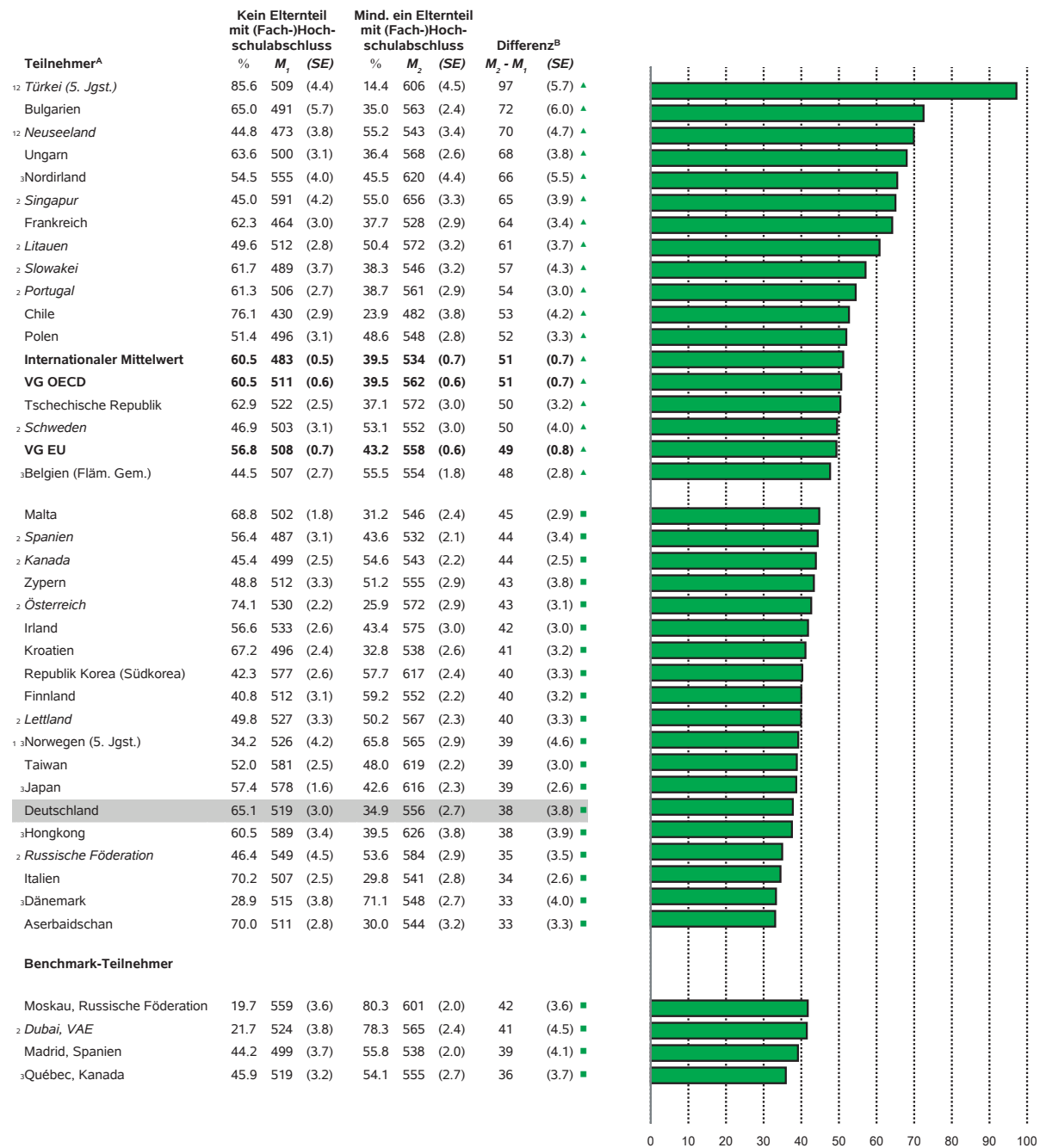
9.4 Soziale Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen im internationalen Vergleich

Im Folgenden werden zunächst Befunde zu sozialen Disparitäten in der vierten Jahrgangsstufe im internationalen Vergleich präsentiert. Als Indikatoren werden dabei zum einen die Anzahl der Bücher im Haushalt und zum anderen das Bildungsniveau der Erziehungsberechtigten (ISCED) genutzt (siehe Abschnitt 9.3). Dazu werden in den Abbildungen 9.1 und 9.2 Familien, in denen mindestens ein Elternteil einen (Fach-)Hochschulabschluss (d.h. tertiären Bildungsabschluss) erreicht hat (mindestens ISCED-Level 6), mit Familien verglichen, in denen kein Elternteil einen tertiären Bildungsabschluss besitzt (maximal ISCED-Level 5). Für die einzelnen Teilnehmerstaaten, für den Durchschnitt aller EU- beziehungsweise OECD-Staaten sowie für den internationalen Mittelwert werden zum einen die Anteile der beiden Gruppen in Prozent und zum anderen die erreichte Mathematikkompetenz (Abbildung 9.1) beziehungsweise die erreichte Naturwissenschaftskompetenz (Abbildung 9.2) dargestellt. Die Differenz in den Kompetenzen zwischen den beiden Gruppen wird zudem grafisch veranschaulicht.

In Deutschland gehört rund ein Drittel (34.9%) der Familien in die Gruppe mit tertiärem Bildungsabschluss. Dieser Wert liegt im Durchschnitt aller OECD-Staaten beziehungsweise im internationalen Mittelwert etwas höher (jeweils 39.5%). Noch deutlicher ist der Unterschied zu Deutschland bei der Betrachtung des Durchschnitts aller EU-Staaten. Dort gilt für 43.2 Prozent der Familien, dass mindestens ein Elternteil einen tertiären Abschluss hat. Abgesehen von den *Benchmark*-Teilnehmern finden sich die höchsten Werte in Dänemark (71.1%), Norwegen (65.8%), Finnland (59.2%) und in der Republik Korea (57.7%). Geringere Anteile als in Deutschland zeigen sich in den EU-Staaten Österreich (25.9%), Italien (29.8%), Malta (31.2%) und Kroatien (32.8%) sowie in der Türkei (14.4%), in Chile (23.9%) und in Aserbaidshan (30.0%).

Im Bereich Mathematik (Abbildung 9.1) erreichen Schülerinnen und Schüler in Deutschland einen Kompetenzwert von 556 Punkten, wenn mindestens ein Elternteil einen tertiären Bildungsabschluss besitzt. Ist dies nicht der Fall, liegt der Wert bei 519 Punkten, was einer Differenz von 38 Punkten entspricht. Diese Differenz ist, wie in allen anderen Teilnehmerstaaten, signifikant. Im internationalen Vergleich fällt diese Differenz der mathematischen Kompetenzwerte jedoch relativ gering aus. Zum einen ist dieser Wert in keinem Staat signifikant kleiner als in Deutschland und zum anderen zeigen sich in den Vergleichsgruppen signifikant höhere Werte (EU: 49 Punkte; OECD: 51 Punkte; internationaler

Abbildung 9.1: Leistungsvorsprung in Mathematik von Kindern aus Familien, in denen mindestens ein Elternteil einen tertiären Bildungsabschluss erreicht hat (ISCED-Level 6, 7 und 8), vor Kindern aus Familien, in denen dies nicht der Fall ist



□ Keine statistisch signifikanten Unterschiede ($p > .05$)

■ Statistisch signifikante Unterschiede ($p \leq .05$)

▲ = Differenzwert im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)

▼ = Differenzwert im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)

■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)

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

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

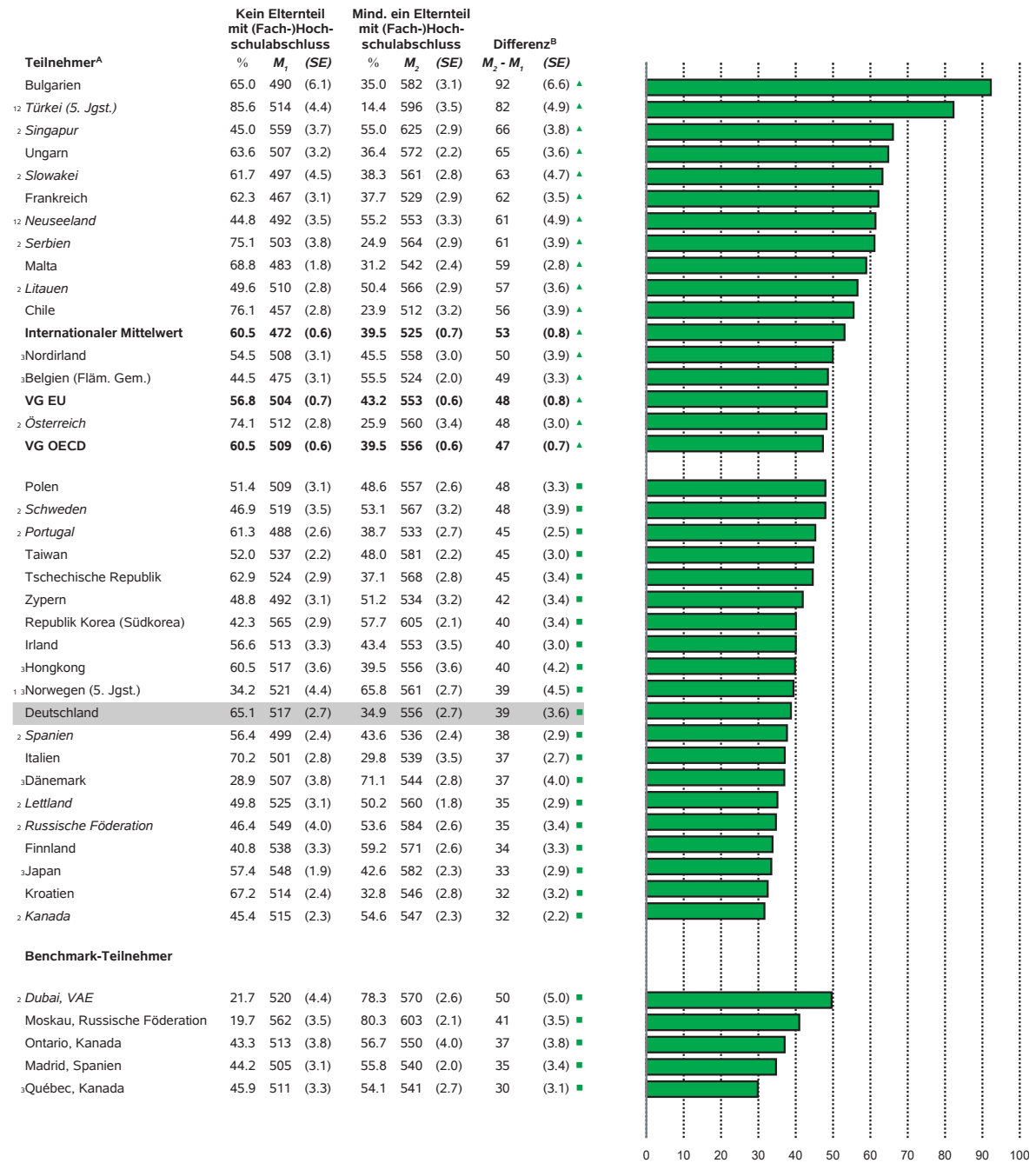
2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.

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

A = Für Australien, England, die Niederlande und die USA liegen keine Ergebnisse vor. Somit sind diese Staaten nicht Teil dieser Darstellung und der berichteten Mittelwerte (International, VG EU, VG OECD).

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

Abbildung 9.2: Leistungsvorsprung in Naturwissenschaften von Kindern aus Familien, in denen mindestens ein Elternteil einen tertiären Bildungsabschluss erreicht hat (ISCED-Level 6, 7 und 8), vor Kindern aus Familien, in denen dies nicht der Fall ist



□ Keine statistisch signifikanten Unterschiede ($p > .05$)

■ Statistisch signifikante Unterschiede ($p \leq .05$)

▲ = Differenzwert im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)

▼ = Differenzwert im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)

■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)

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

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

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

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

A = Für Australien, England, die Niederlande und die USA liegen keine Ergebnisse vor. Somit sind diese Staaten nicht Teil dieser Darstellung und der berichteten Mittelwerte (International, VG EU, VG OECD).

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

Mittelwert: 51 Punkte). Die größten Disparitäten innerhalb der EU finden sich in Bulgarien (72 Punkte), Ungarn (68 Punkte) und Frankreich (64 Punkte).

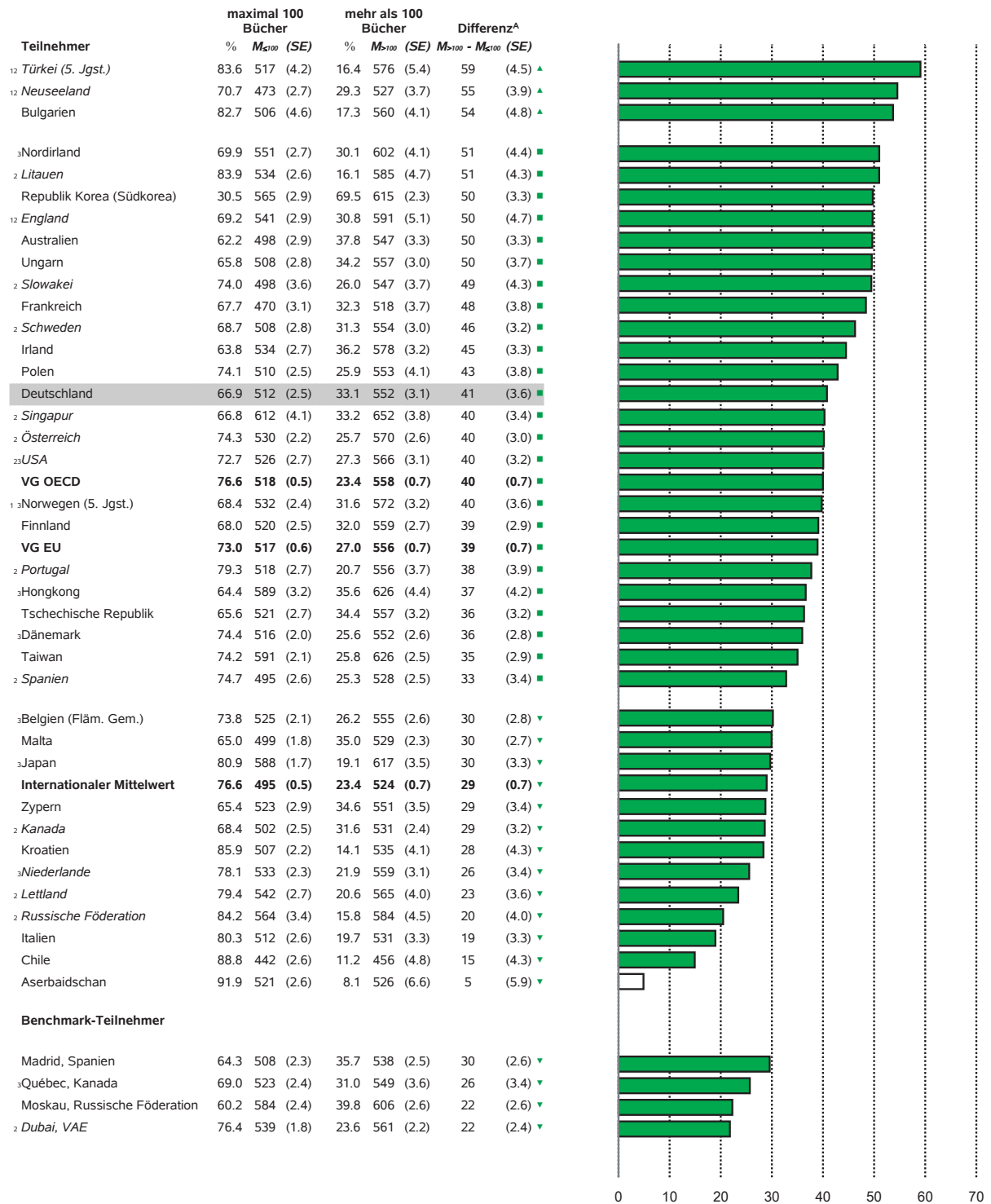
Ein ähnliches Bild zeigt sich für den Kompetenzbereich Naturwissenschaften (Abbildung 9.2). Die Leistungsdifferenz in Deutschland beträgt 39 Punkte (maximal ISCED-Level 5: 517 Punkte; mindestens ISCED-Level 6: 556 Punkte). In allen Teilnehmerstaaten finden sich signifikante Unterschiede zwischen den beiden Gruppen. In keinem Teilnehmerstaat sind die Disparitäten signifikant kleiner als in Deutschland. Im Durchschnitt der OECD (47 Punkte) und der EU (48 Punkte) sowie im Mittel aller Staaten (53 Punkte) sind sie hingegen signifikant größer. Die mit Abstand größten Ungleichheiten innerhalb der EU finden sich in Bulgarien (92 Punkte) gefolgt von Ungarn (65 Punkte), der Slowakei (63 Punkte) und Frankreich (62 Punkte).

Die Abbildungen 9.3 und 9.4 sind ähnlich aufgebaut wie die beiden vorangegangenen. Indikator für den sozialen Status der Familien ist die Anzahl der im Haushalt vorhandenen Bücher, wobei die ursprünglich fünfstufige Variable dichotomisiert wurde. In Deutschland besitzen zwei Drittel der Familien maximal 100 Bücher und entsprechend ein Drittel mehr als 100 Bücher. Diese Werte sind praktisch identisch mit denen aus TIMSS 2015 und TIMSS 2011. Im internationalen Vergleich ist der Anteil von Familien mit mehr als 100 Büchern in Deutschland überdurchschnittlich. Im Mittel aller EU-Staaten gehören 27 Prozent zur Gruppe mit mehr als 100 Büchern im Haushalt; in der OECD sowie im Durchschnitt aller Staaten sind es 23.4 Prozent. Die höchsten Anteile von Familien mit vielen Büchern finden sich (wiederum ohne Berücksichtigung der *Benchmark*-Teilnehmer) in Australien (37.8%), Irland (36.2%) und Hongkong (35.6%), wobei diese Werte nur geringfügig über dem Prozentwert für Deutschland liegen. Ein Ausreißer nach oben ist wie schon 2015 und 2011 die Republik Korea, wo in 69.5 Prozent der Haushalte mehr als 100 Bücher vorhanden sind. Besonders gering ist der Anteil von Familien mit vielen Büchern innerhalb der EU in Kroatien (14.1%), Litauen (16.1%), Bulgarien (17.3%) und Italien (19.7%).

In Deutschland erreichen Viertklässlerinnen und Viertklässler durchschnittlich eine Mathematikkompetenz (Abbildung 9.3) von 552 Punkten, wenn es bei ihnen zu Hause mehr als 100 Bücher gibt. Ist dies nicht der Fall, beträgt der Wert nur 512 Punkte und die Differenz zwischen beiden Gruppen damit 41 Punkte. Abgesehen von Aserbaidschan sind alle Differenzen zwischen den beiden Gruppen innerhalb eines Teilnehmerstaates signifikant. Der Leistungsunterschied in Deutschland weicht nicht signifikant vom OECD- (40 Punkte) und vom EU-Mittelwert (39 Punkte) ab. Die Leistungsdifferenz über alle Staaten hinweg liegt mit 29 Punkten jedoch signifikant unter dem Wert für Deutschland. In drei Staaten finden sich signifikant größere Differenzen als in Deutschland: der Türkei, Neuseeland und Bulgarien. Die Mehrzahl der Staaten unterscheidet sich nicht signifikant von Deutschland und in zwölf Staaten fallen die Disparitäten signifikant geringer aus, darunter in den EU-Staaten Italien, Lettland, Niederlande, Kroatien, Zypern, Malta und in der Flämischen Gemeinschaft in Belgien.

Für die Naturwissenschaftskompetenz (Abbildung 9.4) finden sich wiederum ähnliche Ergebnisse. In Deutschland erreichen Kinder aus Haushalten mit vielen Büchern im Durchschnitt eine Kompetenz von 555 Punkten; sind maximal 100 Bücher vorhanden, fällt dieser Wert 47 Punkte niedriger aus (508 Punkte). Diese Differenz unterscheidet sich nicht signifikant von den Mittelwerten der EU (41 Punkte) und der OECD (40 Punkte), liegt aber signifikant über dem

Abbildung 9.3: Leistungsvorsprung in Mathematik von Kindern aus Familien mit mehr als 100 Büchern vor Kindern aus Familien mit maximal 100 Büchern



□ Keine statistisch signifikanten Unterschiede ($p > .05$)

■ Statistisch signifikante Unterschiede ($p \leq .05$)

▲ = Differenzwert im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)

▼ = Differenzwert im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)

■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)

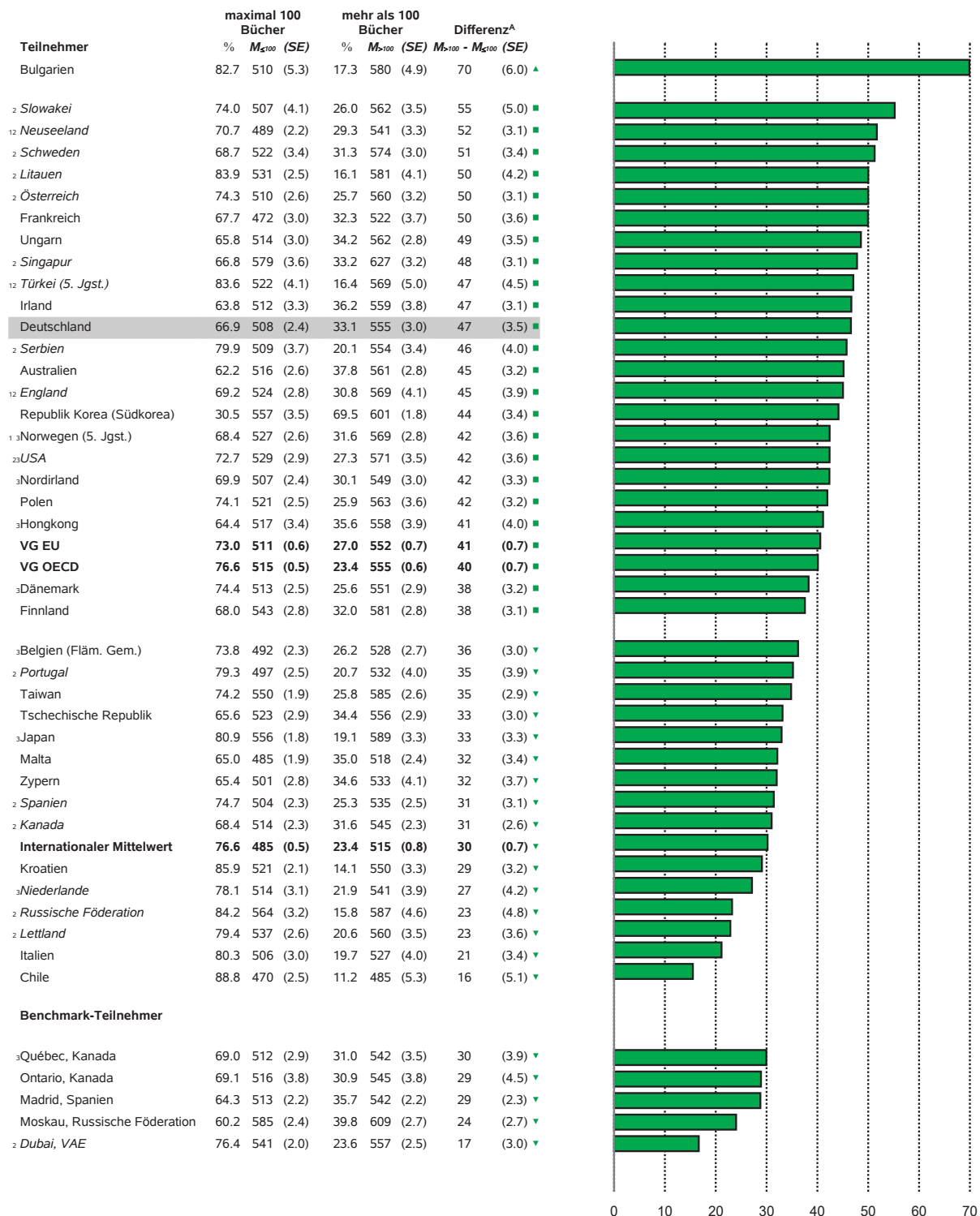
Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

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

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

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

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

Abbildung 9.4: Leistungsvorsprung in Naturwissenschaften von Kindern aus Familien mit mehr als 100 Büchern vor Kindern aus Familien mit maximal 100 Büchern□ Keine statistisch signifikanten Unterschiede ($p > .05$)■ Statistisch signifikante Unterschiede ($p \leq .05$)▲ = Differenzwert im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)▼ = Differenzwert im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)

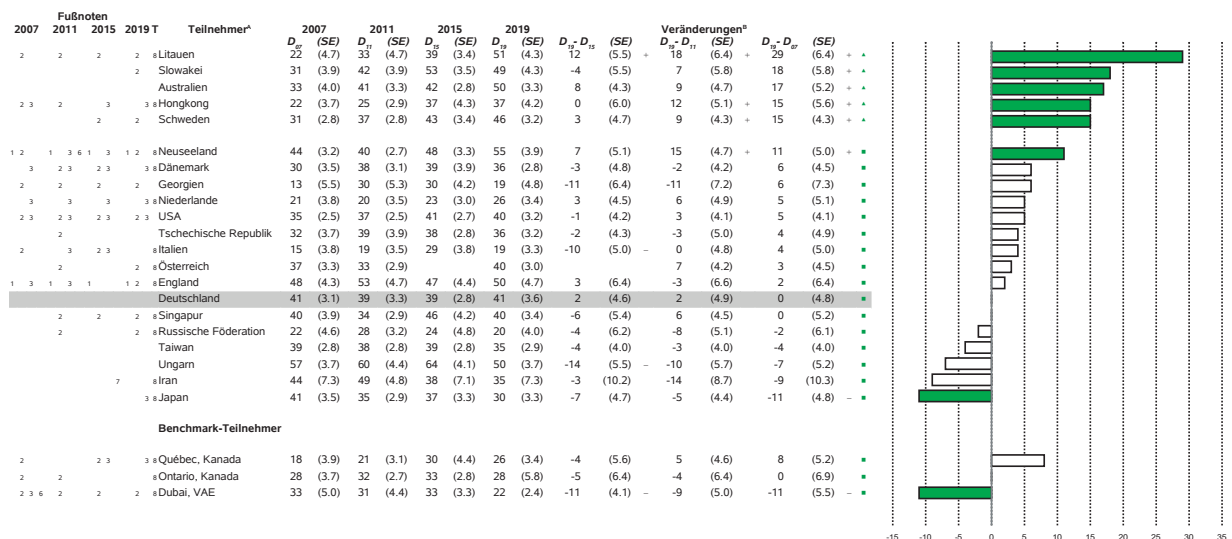
Kursiv gesetzt sind die Teilnehmer, für die von einer eingeschränkten Vergleichbarkeit der Ergebnisse ausgegangen werden muss.

¹ = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.² = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.³ = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.^A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

Durchschnitt aller Teilnehmerstaaten (30 Punkte). Nur in Bulgarien fällt der Leistungsunterschied signifikant größer aus als in Deutschland, während sich in mehreren EU-Staaten signifikant geringere Disparitäten zeigen. Dies betrifft Italien, Lettland, die Niederlande, Kroatien, Spanien, Zypern, Malta, die Tschechische Republik, Portugal und die Flämische Gemeinschaft in Belgien. In allen Teilnehmerstaaten sind die Differenzen zwischen den beiden Gruppen jedoch signifikant.

Die Leistungsunterschiede in Abhängigkeit von den im Haushalt vorhandenen Büchern werden in den Abbildungen 9.5 und 9.6 im Zeitverlauf seit TIMSS 2007 dargestellt. Neben den Mittelwertdifferenzen für die vier Studienzyklen finden sich in den Abbildungen auch die Veränderungen zwischen TIMSS 2019 und den drei vorangegangenen Erhebungen, wobei die Zwölf-Jahres-Differenz (Vergleich mit TIMSS 2007) zudem grafisch dargestellt wird. Die Auswahl der Staaten beschränkt sich auf diejenigen, die seit 2007 an allen TIMS-Studien mit der Primarstufen-Population teilgenommen haben. Hinzu kommt – trotz fehlender Teilnahme im Jahr 2015 – Österreich als weiterer deutschsprachiger Staat.

Abbildung 9.5: Unterschiede im Leistungsvorsprung in Mathematik von Kindern aus Familien mit mehr als 100 Büchern vor Kindern aus Familien mit maximal 100 Büchern im Vergleich von TIMSS 2007, 2011, 2015 und 2019



□ Keine statistisch signifikanten Unterschiede ($p > .05$)

■ Statistisch signifikante Unterschiede ($p \leq .05$)

▲ = Differenzwerte im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)

▼ = Differenzwerte im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)

■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)

+ = Mittelwert in 2019 signifikant höher als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

- = Mittelwert in 2019 signifikant niedriger als 2007 bzw. 2011 bzw. 2015 ($p \leq .05$)

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

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

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

6 = Abweichender Testzeitpunkt

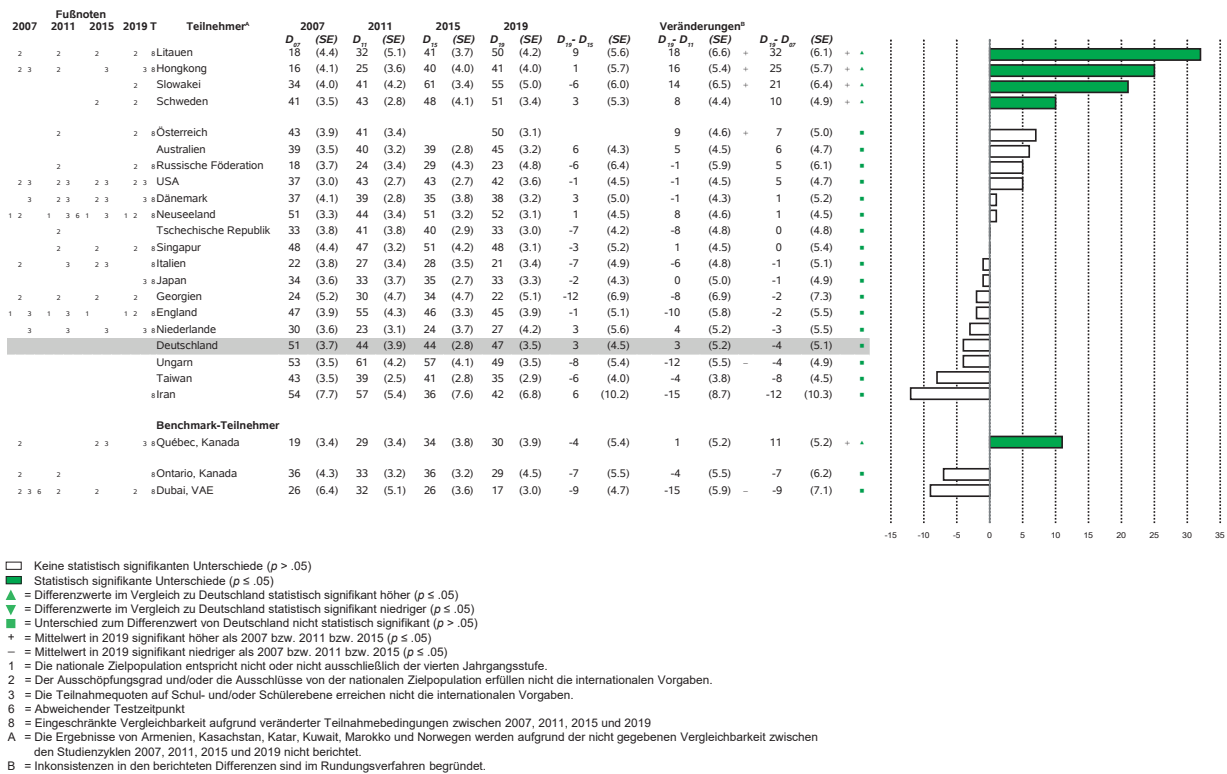
7 = Teilnahme an TIMSS Less Difficult Mathematics

8 = Eingeschränkte Vergleichbarkeit aufgrund veränderter Teilnahmebedingungen zwischen 2007, 2011, 2015 und 2019

A = Die Ergebnisse von Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen werden aufgrund der nicht gegebenen Vergleichbarkeit zwischen den Studienzyklen 2007, 2011, 2015 und 2019 nicht berichtet.

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

Abbildung 9.6: Unterschiede im Leistungsvorsprung in Naturwissenschaften von Kindern aus Familien mit mehr als 100 Büchern vor Kindern aus Familien mit maximal 100 Büchern im Vergleich von TIMSS 2007, 2011, 2015 und 2019



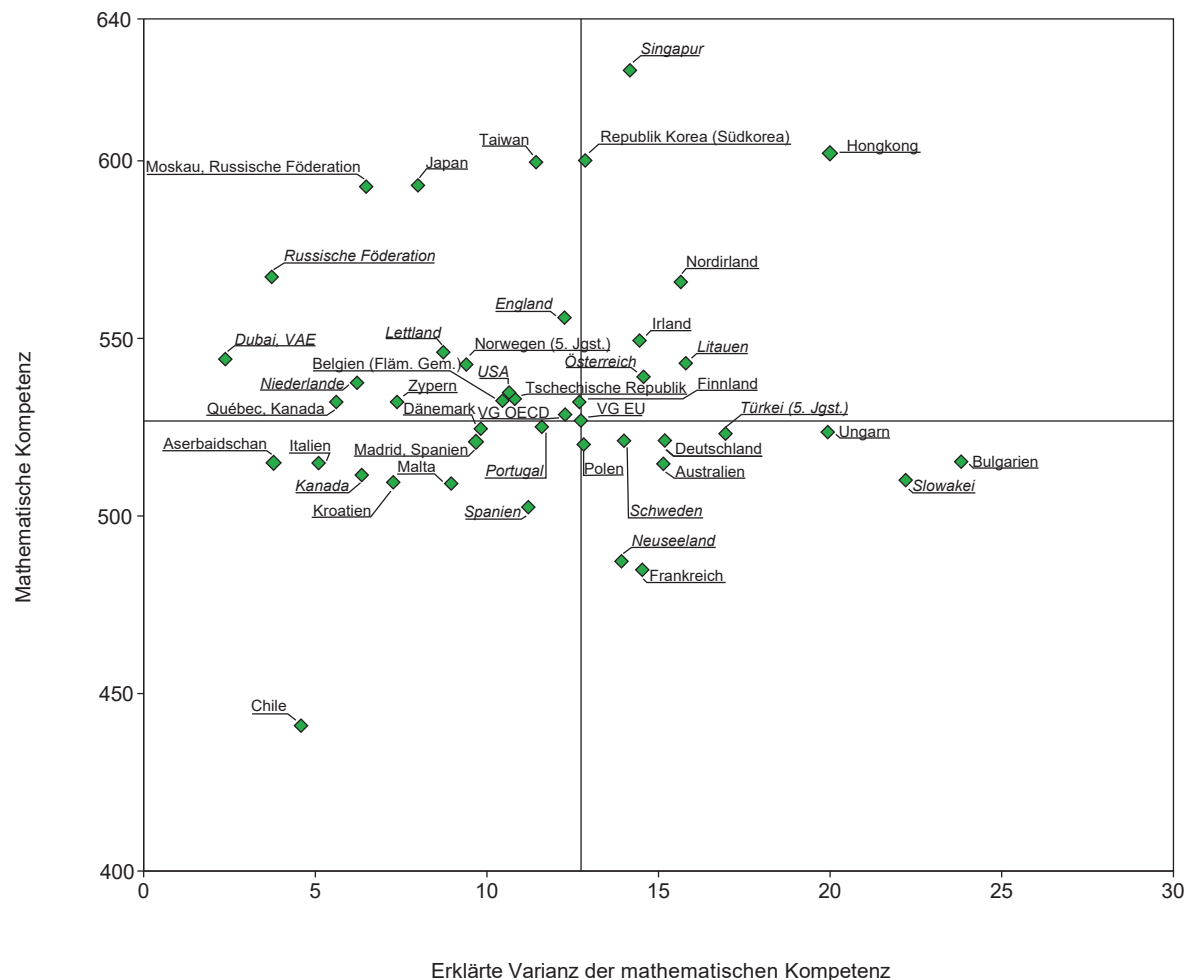
Im Bereich Mathematik (Abbildung 9.5) liegen die Leistungsdifferenzen zwischen Schülerinnen und Schülern aus Familien mit mehr als 100 Büchern und aus Familien mit maximal 100 Büchern in Deutschland im Zeitverlauf zwischen 39 und 41 Punkten, sodass zu keinem Zeitpunkt eine signifikante Veränderung festgestellt werden kann. Lediglich in Japan lässt sich im Trend eine signifikante Verringerung der sozialen Disparitäten beobachten. 2007 war der Wert noch identisch mit der Differenz in Deutschland; 2019 fällt der Unterschied in Japan signifikant geringer aus als in Deutschland. In zwei Staaten – Italien und Ungarn – ergibt sich 2019 ein signifikant niedrigerer Wert als 2015; allerdings haben sich die Differenzen in diesen Staaten zwischen 2007 und 2015 zunächst vergrößert, sodass es im Zwölf-Jahres-Vergleich keine signifikante Veränderung gibt. In sechs Staaten zeichnet sich jedoch ein deutlicher Trend ab: die sozialen Disparitäten verstärken sich. Besonders groß ist diese Veränderung in Litauen (2007: 22 Punkte; 2019: 51 Punkte) – gefolgt von der Slowakei, Australien, Hongkong, Schweden und Neuseeland. In den übrigen Staaten haben sich in den vergangenen zwölf Jahren keine signifikanten Veränderungen der sozialen Disparitäten hinsichtlich der Mathematikkompetenz ergeben.

Für die Naturwissenschaftskompetenz (Abbildung 9.6) zeigen sich im Trend ähnliche Ergebnisse wie für die Differenzen in der Mathematikkompetenz zwischen Viertklässlerinnen und Viertklässlern aus Familien mit mehr als 100 Büchern und aus Familien mit maximal 100 Büchern. Für Deutschland (zwischen 44 und 51 Punkten Differenz) und die meisten anderen Teilnehmer finden sich keine signifikanten Veränderungen zwischen 2007 und 2019. In keinem Staat

verringern sich die sozialen Disparitäten im Zeitverlauf signifikant, es gibt aber drei Teilnehmer, die einen deutlichen Anstieg der Mittelwertdifferenzen zu verzeichnen haben: Litauen (2007: 18 Punkte; 2019: 50 Punkte), Hongkong (2007: 16 Punkte; 2019: 41 Punkte) und die Slowakei (2007: 34 Punkte; 2019: 55 Punkte). Auch in Schweden ist die Leistungsdifferenz 2019 signifikant größer als 2007, allerdings fällt die Veränderung mit 10 Punkten nicht ganz so deutlich aus.

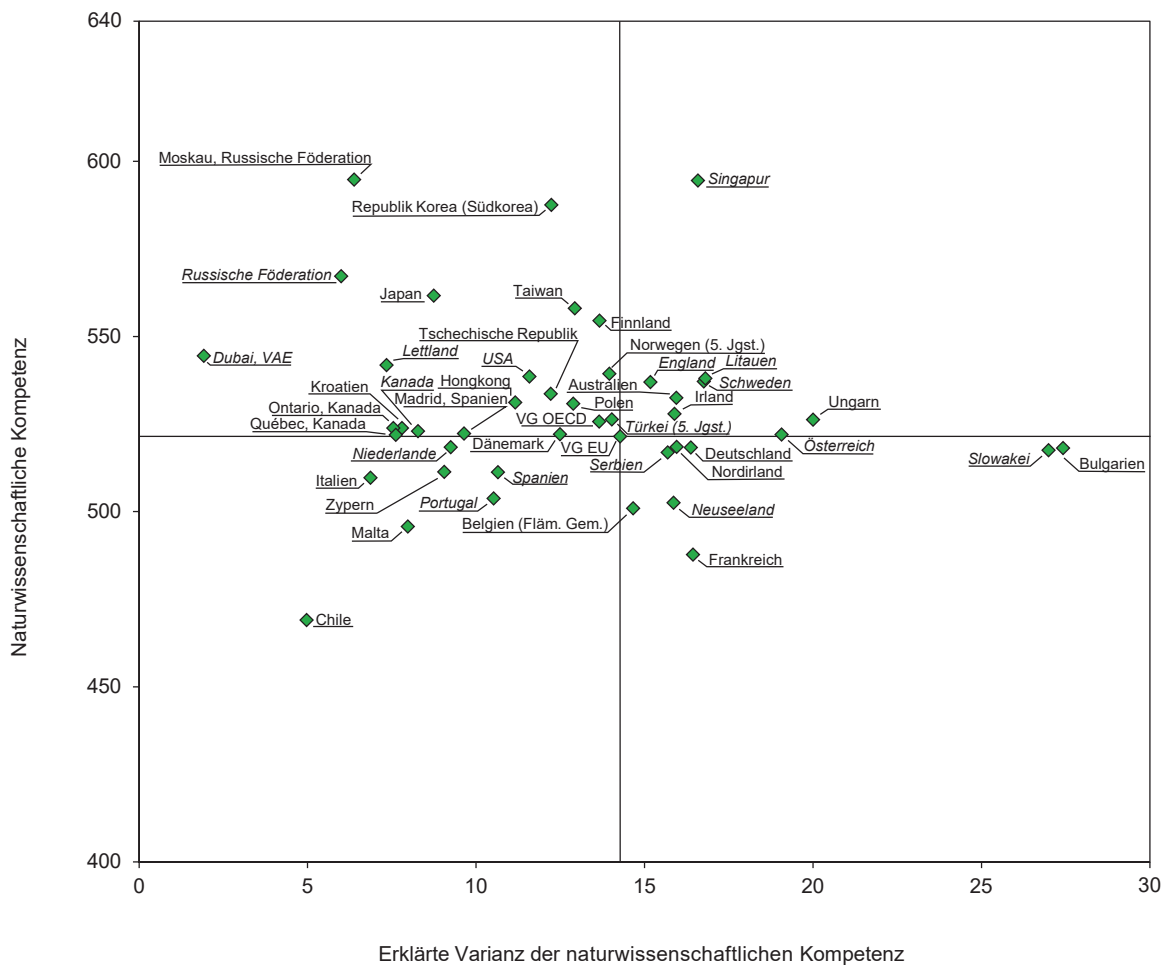
Mithilfe von Regressionsanalysen kann für alle Teilnehmerstaaten der Varianzanteil der Mathematik- beziehungsweise Naturwissenschaftskompetenz bestimmt werden, der durch die fünfstufige Variable *Bücher im Haushalt* erklärt wird. Die Abbildungen 9.7 und 9.8 zeigen für TIMSS 2019 den Zusammenhang zwischen dieser Varianzaufklärung und der Ausprägung der mittleren Mathematik- beziehungsweise Naturwissenschaftskompetenz in den Teilnehmerstaaten.

Abbildung 9.7: Zusammenhang zwischen der erklärten Varianz der Mathematikkompetenz auf Individualebene durch die Anzahl der Bücher im Haushalt und den mittleren Mathematikkompetenzen in den Teilnehmerstaaten



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

Abbildung 9.8: Zusammenhang zwischen der erklärten Varianz der Naturwissenschaftskompetenz auf Individual-ebene durch die Anzahl der Bücher im Haushalt und den mittleren Naturwissenschaftskompetenzen in den Teilnehmerstaaten



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

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Die mittlere erklärte Varianz der Mathematikkompetenz (VG_{EU}) beträgt 12.7 Prozent und die mittlere Mathematikkompetenz 527 Punkte. Für die Naturwissenschaftskompetenz werden im Durchschnitt (VG_{EU}) 14.3 Prozent der Varianz erklärt und der Mittelwert liegt bei 522 Punkten.

In den oberen linken Quadranten der beiden Abbildungen finden sich diejenigen Staaten, die es schaffen, an ihren Grundschulen ein hohes Leistungsniveau zu erreichen und bei denen gleichzeitig die durch die *Anzahl der Bücher im Haushalt* erklärbaren Unterschiede in den erreichten Kompetenzen relativ gering ausfallen. In Abhängigkeit von der jeweiligen Situation in den Staaten kann dies als ein Hinweis auf geringe soziale Disparitäten in den Leistungen der Schülerinnen und Schüler interpretiert werden. In beiden Kompetenzbereichen gehören die Russische Föderation, Japan, Taiwan, Lettland, die USA und die Tschechische Republik in diese Gruppe. Einen ebenfalls unterdurchschnittlichen Zusammenhang zwischen der Anzahl der Bücher im Haushalt und den individuellen Kompetenzen, aber gleichzeitig unterdurchschnittlichen Leistungen, zeigen in beiden Domänen beispielsweise Chile, Malta, Italien und Spanien (untere linke Quadranten).

In den jeweils rechten Quadranten befinden sich die Teilnehmer mit einem überdurchschnittlichen Anteil an erklärter Varianz, was auf eine enge Koppelung zwischen sozialer Herkunft und Leistung hinweisen kann. Besonders hohe Werte erreichen diesbezüglich Bulgarien und die Slowakei, die in Mathematik zudem unterdurchschnittliche Leistungen aufweisen. Die Kombination aus unterdurchschnittlichen Kompetenzen und überdurchschnittlichen Disparitäten findet sich für beide Domänen zudem in Frankreich, Neuseeland und Deutschland. Auch in Singapur ist die Koppelung zwischen sozialem Status und Kompetenzen der Schülerinnen und Schüler überdurchschnittlich. Allerdings liegen dort auch die Durchschnittsleistungen deutlich über dem EU-Mittelwert.

9.5 Soziale Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen in Deutschland

Wie in Abschnitt 9.3 beschrieben, sind die unterschiedlichen Variablen zum Berufsstatus von Personen besonders geeignete Indikatoren für den sozialen Status von Familien. Daher wurden die Erziehungsberechtigten in Deutschland in offener Form nach ihren Berufen gefragt, um anschließend EGP-Klassen und den ISEI berechnen zu können. Auch das Merkmal *Armutsgefährdung* (siehe Abschnitt 9.3) wurde nur in Deutschland erhoben. Daher werden in diesem Abschnitt Befunde zu diesen drei Hintergrundmerkmalen der Viertklässlerinnen und Viertklässler auf nationaler Ebene berichtet.

Der Zusammenhang zwischen dem höchsten ISEI der Erziehungsberechtigten und den erreichten Leistungen lässt sich am besten mithilfe von Regressionsanalysen darstellen. In Tabelle 9.1 ist der soziale Gradient (Steigung der Regressionsgeraden b_1) für die abhängigen Variablen *Mathematik*- und *Naturwissenschaftskompetenz* angegeben. Für beide Domänen kann festgestellt werden, dass der soziale Gradient – und damit der Zusammenhang zwischen sozialer Herkunft und Leistung – 2011 und 2015 signifikant kleiner ist als 2007. Die Werte für TIMSS 2019 unterscheiden sich allerdings nicht signifikant von denjenigen der vorangegangenen Zyklen. Nominell liegen sie unter den Werten von 2007 und über denen aus 2011 und 2015. Ein systematischer Trend lässt sich für Deutschland folglich nicht feststellen.

Tabelle 9.1: Soziale Gradienten für die mathematischen und naturwissenschaftlichen Kompetenzen in Deutschland bei TIMSS 2007, 2011, 2015 und 2019

	Mathematik		Naturwissenschaften	
	sozialer Gradient	(SE)	sozialer Gradient	(SE)
2019	1.27	(0.096)	1.39	(0.087)
2015	1.07	(0.088)	1.22	(0.080)
2011	1.06	(0.055)	1.25	(0.069)
2007	1.41	(0.088)	1.63	(0.101)

Tabelle 9.2 zeigt die mittlere Mathematik- und Naturwissenschaftskompetenz der Schülerinnen und Schüler in Abhängigkeit von der EGP-Klassenzugehörigkeit der Erziehungsberechtigten (Angabe des Vaters und wenn diese fehlt Angabe der Mutter). Kinder, deren Erziehungsberechtigte der oberen Dienstklasse angehören, erreichen in Mathematik 40 Punkte² und in Naturwissenschaften 43 Punkte mehr als ihre Mitschülerinnen und Mitschüler, deren Erziehungsberechtigte der Gruppe der (Fach-)Arbeiter angehören. Dies entspricht in Mathematik einem Vorsprung von knapp einem Lernjahr und in Naturwissenschaften rund anderthalb Lernjahren (Wendt, Kasper, Bos, Vennemann & Goy, 2017). Die Differenz zu Kindern von un- und angelernten Arbeitern beträgt sogar 59 Punkte (Mathematik) beziehungsweise 61 Punkte (Naturwissenschaften).

Tabelle 9.2: Mittlere mathematische und naturwissenschaftliche Kompetenzen pro EGP-Klasse in Deutschland

Sozioökonomische Stellung der Familie (EGP-Klasse)	Mathematik				Naturwissenschaften			
	M	(SE)	SD	(SE)	M	(SE)	SD	(SE)
Obere Dienstklasse (I)	562	(3.6)	63	(2.4)	563	(3.4)	68	(3.3)
Untere Dienstklasse (II)	546	(3.7)	60	(2.9)	549	(3.4)	67	(3.7)
Routinedienstleistungen (III)	518	(6.0)	70	(4.4)	521	(6.1)	75	(4.8)
Selbstständige (IV)	529	(5.0)	58	(3.6)	531	(5.6)	63	(3.8)
(Fach-)Arbeiter (V, VI)	523	(4.4)	67	(3.2)	520	(4.1)	74	(3.9)
Un- und angelernte Arbeiter (VII)	504	(3.7)	66	(2.5)	502	(4.5)	76	(3.5)
Gesamt	534	(2.4)	68	(1.3)	534	(2.4)	74	(2.1)

IEA: Trends in International Mathematics and Science Study

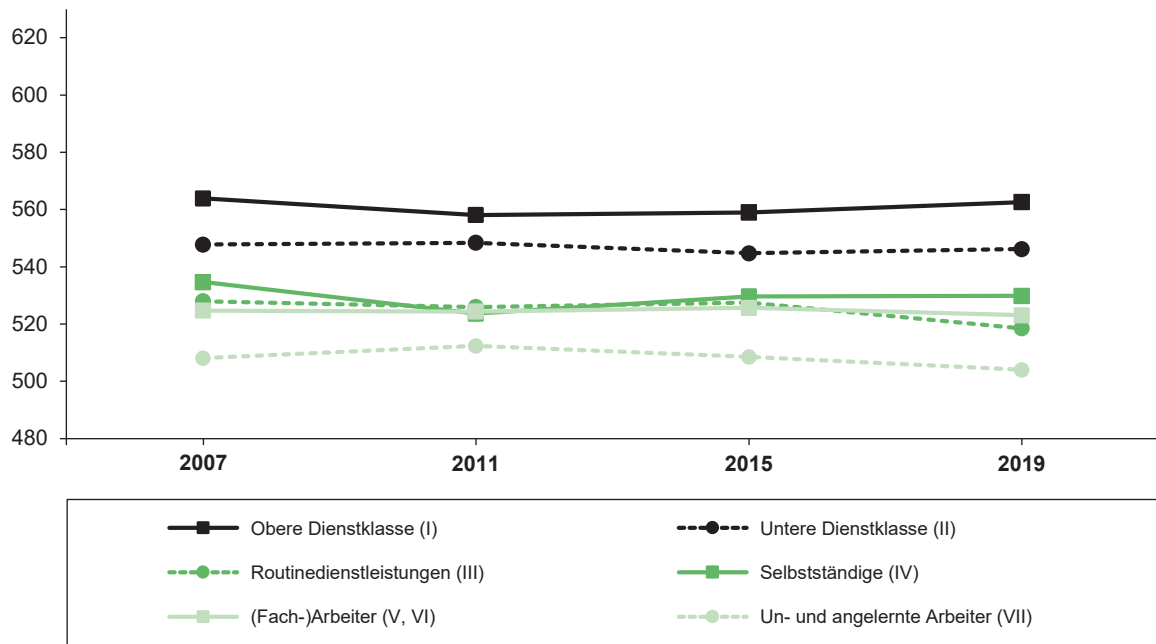
© TIMSS 2019

Der Trend in den Kompetenzen pro EGP-Klasse wird in den Abbildungen 9.9 (Mathematik) und 9.10 (Naturwissenschaften) visualisiert. In beiden Kompetenzbereichen zeigt sich, dass sich zwischen den Gruppen Routinedienstleistungen, Selbstständige und (Fach-)Arbeiter nur geringe Leistungsunterschiede feststellen lassen (insbesondere 2011 und 2015), während sich die anderen drei Gruppen deutlich unterscheiden. Bedeutsame Veränderungen im Zeitverlauf sind nicht zu verzeichnen.

Wie Abbildung 9.11 zu entnehmen ist, beträgt die Armutsgefährdungsquote für TIMSS 2019 insgesamt 21.5 Prozent und liegt damit geringfügig aber nicht signifikant unter dem Wert der ebenfalls repräsentativen Stichprobe von Viertklässlerinnen und Viertklässlern in Deutschland aus IGLU 2016 (24.5%). Theoriekonform unterscheiden sich die Armutsgefährdungsquoten deutlich zwischen verschiedenen gesellschaftlichen Gruppen. Während die Quote in Familien ohne Migrationshintergrund nur 14.3 Prozent beträgt, sind 44.5 Prozent armutsgefährdet, wenn beide Elternteile im Ausland geboren wurden. Weist mindestens ein Elternteil eine abgeschlossene Berufsausbildung oder einen (Fach-) Hochschulabschluss auf (mindestens ISCED-Level 4), so liegt die Quote nur knapp über 10 Prozent. Verfügen die Erziehungsberechtigten über einen Bildungsabschluss auf ISCED-Level 2, beträgt die Quote hingegen fast 50 Prozent. Familien der *service class* gelten nur in 4.6 Prozent (Obere Dienstklasse)

2 Inkonsistenzen zwischen den im Text und in den Abbildungen berichteten Werten sind im Rundungsverfahren begründet.

Abbildung 9.9: Mittlere mathematische Kompetenzen pro EGP-Klasse in Deutschland bei TIMSS 2007, 2011, 2015 und 2019

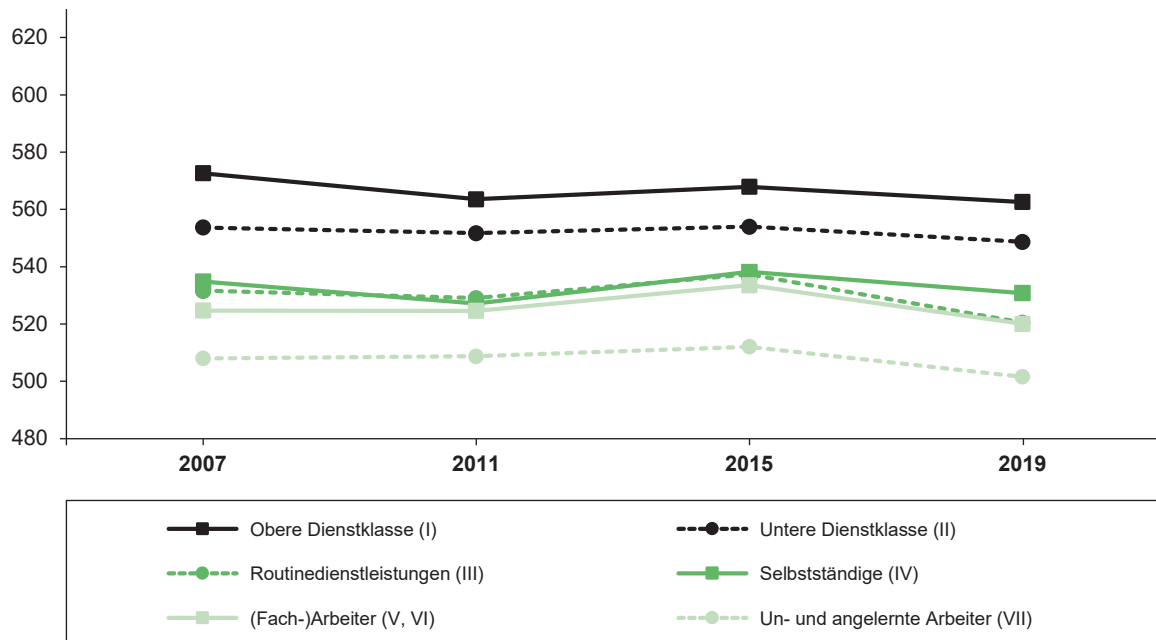


Abweichungen zu den im Rahmen von TIMSS 2015 berichteten Werten (Stubbe, Schwippert & Wendt, 2016) ergeben sich daraus, dass dort für 2015 mit imputierten Daten gearbeitet wurde.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

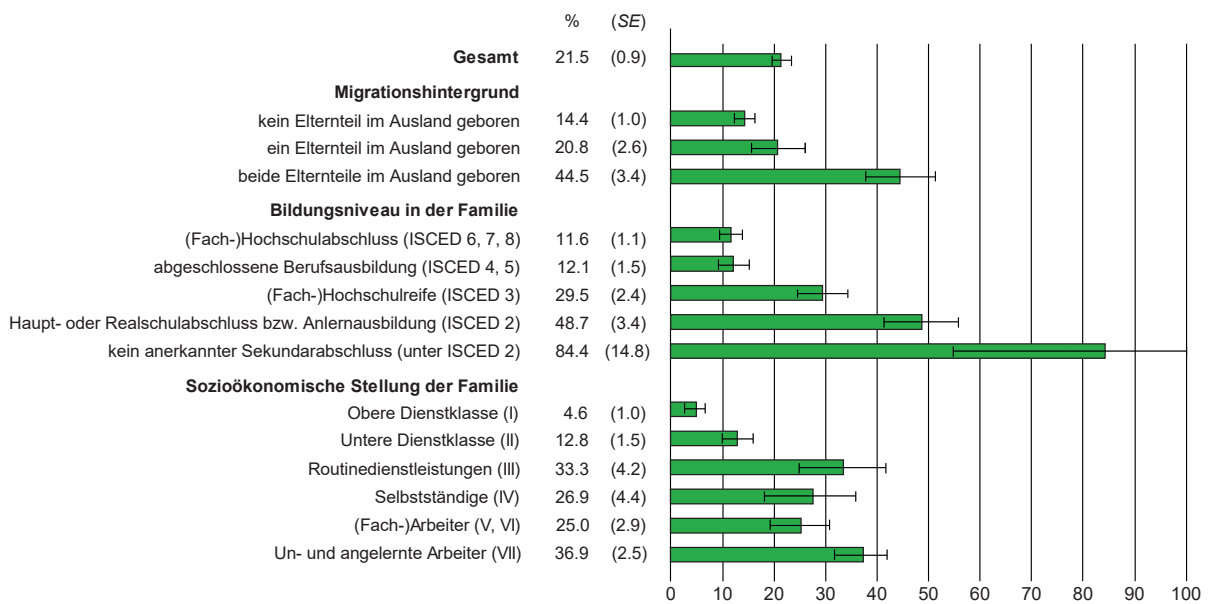
Abbildung 9.10: Mittlere naturwissenschaftliche Kompetenzen pro EGP-Klasse in Deutschland bei TIMSS 2007, 2011, 2015 und 2019



Abweichungen zu den im Rahmen von TIMSS 2015 berichteten Werten (Stubbe, Schwippert & Wendt, 2016) ergeben sich daraus, dass dort für 2015 mit imputierten Daten gearbeitet wurde.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Abbildung 9.11: Armutsgefährdungsquoten nach Migrationshintergrund, höchstem Bildungsabschluss und EGP-Klasse in Deutschland (in Prozent)

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Tabelle 9.3: Mittlere mathematische und naturwissenschaftliche Kompetenzen nach *Armutsgefährdung* und EGP-Klasse in Deutschland

Sozioökonomische Stellung der Familie (EGP-Klasse)	Mathematik							
	nicht armutsgefährdet				armutsgefährdet			
	M_n	(SE)	SD	(SE)	M_a	(SE)	SD	(SE)
Obere Dienstklasse (I)	568	(3.9)	61	(2.8)	504	(15.3)	68	(9.5)
Untere Dienstklasse (II)	552	(4.0)	59	(3.1)	528	(9.4)	54	(6.5)
Routinedienstleistungen (III)	542	(7.6)	60	(4.7)	497	(12.3)	73	(9.9)
Selbstständige (IV)	537	(6.2)	57	(4.9)	516	(11.4)	62	(7.2)
(Fach-)Arbeiter (V, VI)	532	(5.5)	63	(3.1)	510	(8.8)	66	(5.4)
Un- und angelernte Arbeiter (VII)	517	(5.8)	64	(3.1)	493	(6.4)	62	(4.7)
Gesamt	546	(2.8)	64	(1.5)	507	(3.8)	68	(2.8)
Sozioökonomische Stellung der Familie (EGP-Klasse)	Naturwissenschaften							
	nicht armutsgefährdet				armutsgefährdet			
	M_n	(SE)	SD	(SE)	M_a	(SE)	SD	(SE)
Obere Dienstklasse (I)	568	(3.5)	64	(3.7)	510	(21.5)	89	(16.2)
Untere Dienstklasse (II)	557	(3.8)	62	(3.4)	517	(10.7)	69	(7.3)
Routinedienstleistungen (III)	553	(6.3)	56	(4.3)	490	(11.4)	74	(8.0)
Selbstständige (IV)	541	(7.1)	62	(5.5)	511	(13.6)	64	(7.7)
(Fach-)Arbeiter (V, VI)	531	(5.1)	70	(4.3)	507	(9.7)	67	(6.3)
Un- und angelernte Arbeiter (VII)	519	(6.2)	69	(3.1)	482	(8.4)	77	(5.3)
Gesamt	549	(2.6)	67	(1.9)	497	(4.6)	76	(2.8)

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

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

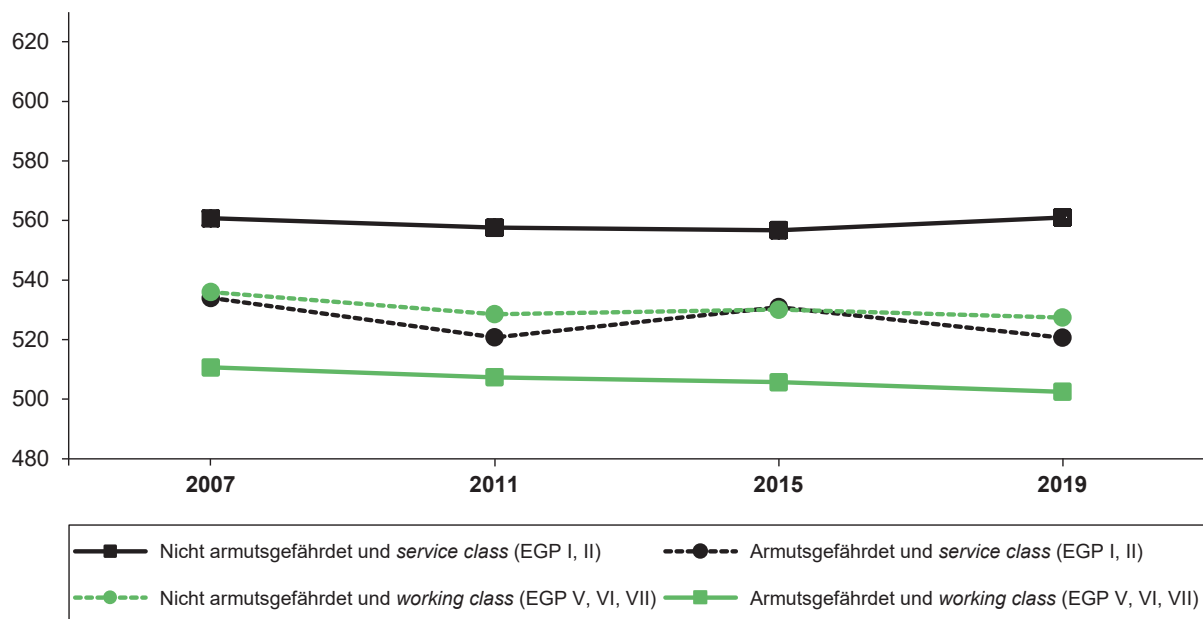
beziehungsweise 12.8 Prozent (Untere Dienstklasse) als armutsgefährdet, während dies in den übrigen EGP-Klassen für 25.0 bis 36.9 Prozent der Familien zutrifft.

Wie Tabelle 9.3 entnommen werden kann, liegt die Mathematikkompetenz von Kindern aus nicht armutsgefährdeten Elternhäusern 39 Punkte (knapp ein Lernjahr) über der Mathematikkompetenz von armutsgefährdeten Schülerinnen und Schülern. Für den Bereich Naturwissenschaften fällt diese Differenz mit 51 Punkten (knapp zwei Lernjahre) größer aus. Im Trend seit TIMSS 2007 zeigen sich keine signifikanten Veränderungen (Bos et al., 2010).

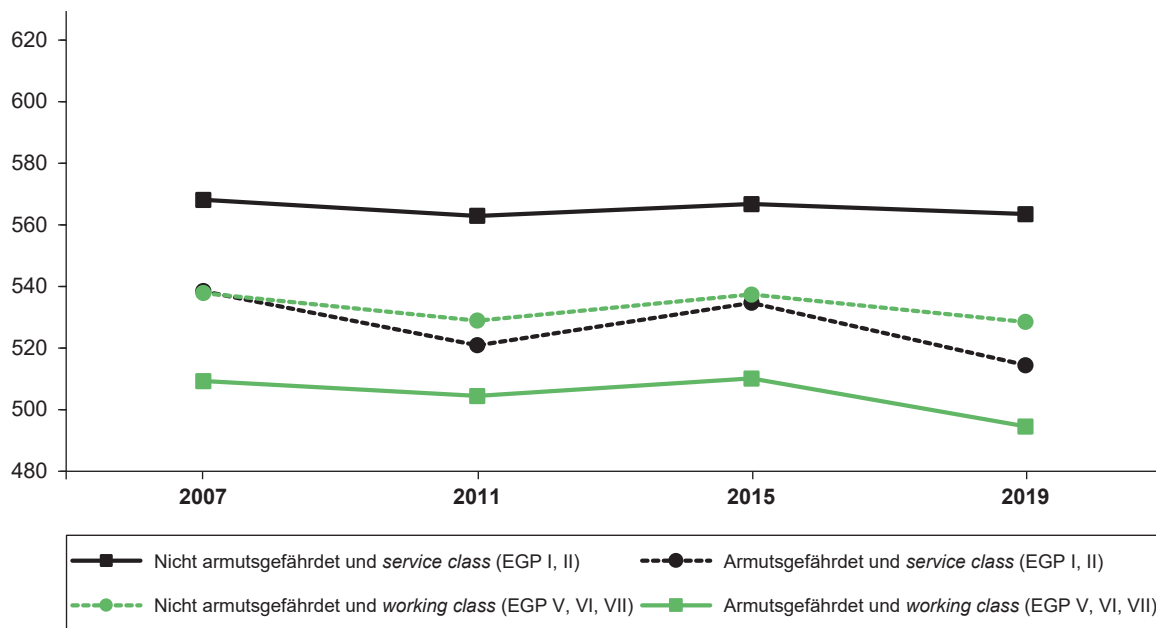
Bedingt durch die geringe Größe der Teilstichproben – insbesondere für armutsgefährdete Schülerinnen und Schüler aus bildungsnahen Elternhäusern – erlauben die vorliegenden Daten kaum generalisierbare Aussagen zu Unterschieden zwischen den einzelnen EGP-Klassen. Tendenziell finden sich aber theoriekonform in den oberen sozialen Lagen größere Leistungsdifferenzen zwischen Schülerinnen und Schülern aus armutsgefährdeten und nicht armutsgefährdeten Familien als in den unteren sozialen Lagen.

Die Abbildungen 9.12 (Mathematikkompetenz) und 9.13 (Naturwissenschaftskompetenz) zeigen für vier gesellschaftliche Gruppen den Trend seit 2007. Die mit Abstand höchsten Kompetenzwerte werden von Viertklässlerinnen und Viertklässlern erreicht, deren Erziehungsberechtigte der *service class* angehören und nicht armutsgefährdet sind (zwischen 557 und 561 Punkten in Mathematik und zwischen 563 und 568 Punkten in Naturwissenschaften). Schülerinnen und Schüler, deren Erziehungsberechtigte der *working class* angehören und armutsgefährdet sind, schneiden gegenüber der vorherigen Gruppe in der Mathematikkompetenz in den Jahren 2007 bis 2015 rund 50 Punkte schlechter ab (2019 sogar 59 Punkte). Im Bereich Naturwissenschaften betragen diese Differenzen in den Jahren 2007 bis 2015 circa 58 Punkte (2019 sogar 69 Punkte). Zwischen den beiden anderen Gruppen – ‚armutsgefährdet und *service class*‘ so-

Abbildung 9.12: Mittlere mathematische Kompetenzen nach *Armutsgefährdung* und EGP-Klasse in Deutschland bei TIMSS 2007, 2011, 2015 und 2019



Abweichungen zu den im Rahmen von TIMSS 2015 berichteten Werten (Stubbe, Schwippert & Wendt, 2016) ergeben sich daraus, dass dort für 2015 mit imputierten Daten gearbeitet wurde.

Abbildung 9.13: Mittlere naturwissenschaftliche Kompetenzen nach *Armutsgefährdung* und EGP-Klasse in Deutschland bei TIMSS 2007, 2011, 2015 und 2019

Abweichungen zu den im Rahmen von TIMSS 2015 berichteten Werten (Stubbe, Schwippert & Wendt, 2016) ergeben sich daraus, dass dort für 2015 mit imputierten Daten gearbeitet wurde.

wie ‚nicht armutsgefährdet und *working class*‘ – zeigen sich in den Jahren 2007 und 2015 in beiden Kompetenzbereichen praktisch keine Leistungsunterschiede. In den Jahren 2011 (beide Domänen) und 2019 (Mathematikkompetenz) liegen die Werte für die Gruppe ‚armutsgefährdet und *service class*‘ nominell rund 8 Punkte unter denjenigen für die Gruppe ‚nicht armutsgefährdet und *working class*‘ (nicht signifikant). Für die Naturwissenschaftskompetenz bei TIMSS 2019 zeigt sich mit einer Differenz von 14 Punkten sogar ein signifikanter Unterschied zugunsten der Gruppe ‚nicht armutsgefährdet und *working class*‘.

9.6 Zusammenfassung

Wie bereits in den vorangegangenen Erhebungen TIMSS 2007, 2011 und 2015 weisen auch die in diesem Kapitel präsentierten Ergebnisse auf eine nach wie vor enge Kopplung zwischen dem sozioökonomischen Status von Familien und dem Bildungserfolg ihrer Kinder hin. Zwischen den Teilnehmerstaaten bestehen allerdings erhebliche Unterschiede im Ausmaß der sozialen Disparitäten.

Dass der Kompetenzerwerb in den Domänen Mathematik und Naturwissenschaften maßgeblich mit dem Bildungsstatus der Erziehungsberechtigten in Zusammenhang steht, zeigt sich an den betrachteten Leistungsunterschieden zwischen Schülerinnen und Schülern, bei denen mindestens eine erziehungsberechtigte Person über einen tertiären Bildungsabschluss verfügt, gegenüber Kindern von Erziehungsberechtigten ohne einen (Fach-)Hochschulabschluss: Sowohl für Deutschland als auch für alle übrigen Teilnehmerstaaten sind die Kompetenzunterschiede zwischen diesen Gruppen in Mathematik und den Naturwissenschaften signifikant. Allerdings sind in keinem Teilnehmerstaat die Disparitäten in den mathematischen und naturwissenschaftlichen Kompetenzen

signifikant kleiner als in Deutschland; jedoch im EU- und OECD-Durchschnitt ebenso wie im internationalen Mittelwert signifikant größer.

Hinsichtlich der Leistungsunterschiede der Viertklässlerinnen und Viertklässler in Abhängigkeit von den im Haushalt vorhandenen Büchern zeigt sich ein ähnliches Bild: Kinder aus Haushalten mit vielen Büchern erreichen sowohl national als auch international signifikant höhere mathematische und naturwissenschaftliche Kompetenzwerte als Kinder aus Haushalten mit wenigen Büchern. Im internationalen Trend lässt sich (mit Ausnahme von Japan) zwischen 2007 und 2019 keine signifikante Verringerung der sozialen Disparitäten beobachten – weder hinsichtlich der Mathematikkompetenz noch der Naturwissenschaftskompetenz. In einigen Staaten verstärken sich die sozialen Disparitäten hingegen zum Teil drastisch (z. B. Litauen, Slowakei, Hongkong, Schweden).

Ergänzend zu den international eingesetzten Variablen zur Erfassung der sozialen Lage wurden in Deutschland zusätzliche Indikatoren, wie der Berufsstatus der Erziehungsberechtigten oder das Merkmal *Armutsgefährdung* erhoben, die vertiefende Analysen auf nationaler Ebene ermöglichen. Analog zu den internationalen Ergebnissen zeigen sich auch anhand dieser Indikatoren Leistungsunterschiede im deutschen Bildungssystem in Abhängigkeit der sozialen Herkunft. Viertklässlerinnen und Viertklässler aus den oberen sozialen Lagen haben in Mathematik einen Vorsprung von circa einem Lernjahr und in den Naturwissenschaften sogar von anderthalb Lernjahren gegenüber Kindern aus Familien mit einem niedrigen sozialen Status. Rund ein Fünftel der in TIMSS 2019 untersuchten Schülerinnen und Schüler fallen mit ihren Familien unter die Armutsgefährdungsgrenze. Besonders hoch ist die *Armutsgefährdung* für Kinder aus Familien mit Migrationshintergrund und Familien mit einem niedrigen Bildungsniveau. Kinder aus nicht armutsgefährdeten Elternhäusern haben gegenüber armutsgefährdeten Schülerinnen und Schülern in Mathematik einen Lernvorsprung von knapp einem Jahr. Im Bereich Naturwissenschaften beträgt diese Differenz sogar knapp zwei Lernjahre. Die Unterschiede zwischen den Gruppen sind seit 2007 unverändert hoch.

Nach wie vor besteht folglich die Notwendigkeit, soziale Disparitäten im Hinblick auf den Bildungsverlauf sowie den Kompetenzerwerb zu berücksichtigen. Dies bekräftigen die in diesem Kapitel präsentierten Ergebnisse. Für zukünftige Untersuchungen wird zu beobachten sein, wie sich die bereits wiederholt dokumentierten sozialen Disparitäten insbesondere unter den aktuellen Bedingungen (z. B. durch temporäre Schulschließungen oder *Homeschooling*) entwickeln werden und welche Konsequenzen sich daraus im Hinblick auf Teilhabechancen zwischen sozial privilegierten und benachteiligten Kindern ergeben (Fickermann & Edelstein, 2020).

Literatur

- Anders, Y. & Roßbach, H. G. (2013). Frühkindliche Bildungsforschung in Deutschland. In M. Stamm & D. Edelmann (Hrsg.), *Handbuch frühkindliche Bildungsforschung* (S. 183–195). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-19066-2_13
- Bardone, L. & Guio, A.-C. (2005). In-work poverty. New commonly agreed indicators at EU level. *Statistics in focus*, 5, 1–11.
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiss (Hrsg.), *PISA 2000*.

- Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–407). Opladen: Leske + Budrich.
https://doi.org/10.1007/978-3-322-83412-6_10
- Becker, R. & Lauterbach, W. (2004). Dauerhafte Bildungsungleichheiten – Ursachen, Mechanismen, Prozesse und Wirkungen. In R. Becker & W. Lauterbach (Hrsg.), *Bildung als Privileg?* (S. 9–40). Wiesbaden: VS Verlag für Sozialwissenschaften.
https://doi.org/10.1007/978-3-322-93532-8_1
- Bertelsmann Stiftung. (2016). *Steigende Kinderarmut beeinträchtigt Chancen fürs ganze Leben*. Pressemitteilung vom 12.09.2016. Verfügbar unter: <http://www.bertelsmannstiftung.de/de/presse/pressemitteilungen/pressemitteilung/pid/steigende-kinderarmut-beeintraechtigt-chancen-fuers-ganze-leben/>
- Blossfeld, H.-P. (1993). Changes in educational opportunities in the Federal Republic of Germany. A longitudinal study of cohorts born between 1916 and 1965. In Y. Shavit & H.-P. Blossfeld (Hrsg.), *Persistent inequality. Changing educational attainment in thirteen countries* (S. 51–74). Boulder, CO: Westview Press.
- Blossfeld, H.-P., Blossfeld, G. J. & Blossfeld, P. N. (2019). Soziale Ungleichheiten und Bildungsentscheidungen im Lebensverlauf: Die Perspektive der Bildungssoziologie. *Journal for Educational Research Online*, 11(1), 16–30.
- Blossfeld, H.-P. & Shavit, Y. (1993). Dauerhafte Ungleichheiten: zur Veränderung des Einflusses der sozialen Herkunft auf die Bildungschancen in dreizehn industrialisierten Ländern. *Zeitschrift für Pädagogik*, 39(1), 25–52.
- BMAS (Bundesministerium für Arbeit und Soziales). (2017). *Lebenslagen in Deutschland. Armuts- und Reichtumsberichterstattung der Bundesregierung. Der fünfte Armuts- und Reichtumsbericht der Bundesregierung*. Verfügbar unter: https://www.armuts-und-reichtumsbericht.de/SharedDocs/Downloads/Berichte/5-arb-langfassung.pdf?__blob=publicationFile&v=6
- Bos, W., Stubbe, T. C. & Buddeberg, M. (2010). Gibt es eine armutsbedingte Bildungsbenachteiligung? Die Operationalisierung verschiedener Indikatoren der sozialen Herkunft in der empirischen Bildungsforschung. In D. H. Rost (Hrsg.), *Intelligenz, Hochbegabung, Vorschulerziehung, Bildungsbenachteiligung* (S. 165–208). Münster: Waxmann.
- Bosch, G. & Kalina, T. (2007). Niedriglöhne in Deutschland – Zahlen, Fakten, Ursachen. In G. Bosch & C. Weinkopf (Hrsg.), *Arbeiten für wenig Geld. Niedriglohnbeschäftigung in Deutschland* (S. 20–105). Frankfurt a. M.: Campus.
- Boudon, R. (1974). *Education, opportunity, and social inequality*. New York: Wiley & Sons.
- Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In R. Kreckel (Hrsg.), *Soziale Ungleichheiten* (Soziale Welt, Sonderbd. 2, S. 183–198). Göttingen: Schwartz.
- Butterwegge, C. (Hrsg.). (2000). *Kinderarmut in Deutschland. Ursachen, Erscheinungsformen und Gegenmaßnahmen*. Frankfurt a. M.: Campus.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(1), 95–120. <https://doi.org/10.1086/228943>
- DESTATIS (2020, Februar). *Armutsschwelle und Armutsgefährdung (monetäre Armut) in Deutschland*. Verfügbar unter: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Lebensbedingungen-Armutsgefaehrung/Tabellen/armutsschwelle-gefaehrung-silc.html;jsessionid=D282227F34EB-4C254C3BDE6B3EB9C091.internet8741>
- Erikson, R., Goldthorpe, H. J. & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology*, 30(4), 415–441. <https://doi.org/10.2307/589632>
- Fickermann, D. & Edelstein, B. (Hrsg.). (2020). *„Langsam vermisste ich die Schule...“: Schule während und nach der Corona-Pandemie*. Münster: Waxmann.
<https://doi.org/10.31244/9783830992318>
- Ganzeboom, H. B. G., Graaf, P. M. de & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1–56.
[https://doi.org/10.1016/0049-089X\(92\)90017-B](https://doi.org/10.1016/0049-089X(92)90017-B)
- Ganzeboom, H. B. G. & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 international standard classification of occupations. *Social Science Research*, 25(3), 201–239. <https://doi.org/10.1006/ssre.1996.0010>

- Ganzeboom, H. B. G. & Treiman, D. (2010). *Occupational status measures for the new International Standard Classification of Occupations ISCO-08; With a discussion of the new classification*. Verfügbar unter: <http://www.harryganzeboom.nl/isol/isol2010c2-ganzeboom.pdf>
- Geißler, R. (2014). Bildungsexpansion und Bildungschancen. In Bundeszentrale für politische Bildung (Hrsg.), *Sozialer Wandel in Deutschland* (Informationen zur politischen Bildung, Heft 324). Verfügbar unter: <http://www.bpb.de/izpb/198031/bildungsexpansion-und-bildungschancen?p=all>
- Gogolin, I. (2020). Sprachliche Förderung, sprachliche Bildung und Lernen im Deutschen als Zweitsprache während und nach der Pandemie. In D. Fickermann & B. Edelstein (Hrsg.), *„Langsam vermisste ich die Schule...“ Schule während und nach der Corona-Pandemie* (S. 175–188). Münster: Waxmann. <https://doi.org/10.31244/9783830992318.11>
- Hasselhorn, M. & Kuger, S. (2014). Wirksamkeit schulrelevanter Förderung in Kindertagesstätten. *Zeitschrift für Erziehungswissenschaft*, 17(2), 299–314. <https://doi.org/10.1007/s11618-013-0473-2>
- Hoffmann, E. (2003). International statistical comparisons of occupational and social structures. Problems, possibilities and the role of ISCO-88. In J. H. Hoffmeyer-Zlotnik & C. Wolf (Hrsg.), *Advances in cross-national comparison. A European working book for demographic and socio-economic variables* (S. 137–158). New York: Plenum Press.
- Hußmann, A., Stubbe, T. C. & Kasper, D. (2017). Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195–217). Münster: Waxmann.
- Kuhl, P., Siegle, T. & Lenski, A. E. (2013). Soziale Disparitäten. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 275–296). Münster: Waxmann.
- Laubstein, C., Holz, G. & Seddig, N. (2016). *Armutsfolgen für Kinder und Jugendliche. Erkenntnisse aus empirischen Studien in Deutschland*. Gütersloh: Bertelsmann Stiftung.
- Lorenz, J. & Stubbe, T. C. (2017, 24. August). *Families' social capital and students' achievement in mathematics and science* [Konferenzpräsentation]. European Conference on Educational Research (ECER), Kopenhagen, Dänemark.
- Lucas, S. R. (2001). Effectively maintained inequality. Education transitions, track mobility, and social background effects. *American Journal of Sociology*, 106(6), 1642–1690. <https://doi.org/10.1086/321300>
- Mahler, N. & Kölm, J. (2019). Soziale Disparitäten. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 265–293). Münster: Waxmann.
- McMonagle, S. & Hansen, A. (2020). *Forschungsschwerpunkt Sprachliche Bildung und Mehrsprachigkeit 2013–2020. Projektvorstellungen und Ergebnisse 1. und 2. Förderphase*. Verfügbar unter: <https://www.mehrsprachigkeit.uni-hamburg.de/bilder/broschue-re-forschungsschwerpunkt-sprachliche-bildung-mehrsprachigkeit.pdf>
- Müller, K. & Ehmke, T. (2016). Soziale Herkunft und Kompetenzerwerb. In K. Reiss, C. Sälzer, A. Schiepe-Tiska, E. Klieme & O. Köller (Hrsg.), *PISA 2015. Eine Studie zwischen Kontinuität und Innovation* (S. 285–316). Münster: Waxmann.
- Nonte, S., Haas, M. & Stubbe, T. C. (2019). Unterrichtsqualität in MINT-Klassen – Befunde aus der quasi-experimentellen Studie ProBiNi. In T. Ehmke, P. Kuhl & M. Pietsch (Hrsg.), *Lehrer. Bildung. Gestalten. Beiträge zur empirischen Forschung in der Lehrerbildung* (S. 325–335). Weinheim: Juventa.
- Schimpl-Neimanns, B. (2000). Soziale Herkunft und Bildungsbeteiligung. Empirische Analysen zu herkunftsspezifischen Bildungsungleichheiten zwischen 1950 und 1989. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 52(4), 636–669. <https://doi.org/10.1007/s11577-000-0102-y>
- Schroedter, J. H., Lechert, Y. & Lüttinger, P. (2006). *Die Umsetzung der Bildungsskala ISCED – 1997 für die Volkszählung 1970, die Mikrozensus-Zusatzerhebung 1971*

- und die Mikrozensen 1976–2004. ZUMA-Methodenbericht 2006/08. Verfügbar unter: https://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2006/06_08_Schroedter.pdf
- Schwippert, K. (2019). Was wird aus den Büchern? Sozialer Hintergrund von Lernenden und Bildungsungleichheit aus Sicht der international vergleichenden Erziehungswissenschaft. *Journal for Educational Research Online*, 11(1), 92–117.
- Seidel, T., Reinhold, S., Holzberger, D., Mok, S. Y., Schiepe-Tiska, A. & Reiss, K. (2016). *Wie gelingen MINT-Schulen? Anregungen aus Forschung und Praxis*. Münster: Waxmann.
- Senkbeil, M., Drossel, K., Eickelmann, B. & Vennemann, M. (2019). Soziale Herkunft und computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich. In B. Eickelmann, W. Bos, J. Gerick, F. Goldhammer, H. Schaumburg, K. Schwippert, M. Senkbeil & J. Vahrenhold (Hrsg.), *ICILS 2018 #Deutschland. Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking* (S. 301–333). Münster: Waxmann.
- Solga, H. (2005). Meritokratie – die moderne Legitimation ungleicher Bildungschancen. In P. A. Berger & H. Kahlert (Hrsg.), *Institutionalisierte Ungleichheiten. Wie das Bildungswesen Chancen blockiert* (S. 19–38). Weinheim: Juventa.
- Stanat, P., Weirich, S. & Radmann, S. (2012). Sprach- und Leseförderung. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. Ergebnisse des IQB-Ländervergleichs 2011* (S. 251–276). Münster: Waxmann.
- Statistisches Bundesamt. (2016). *Armutsgefährdung in Westdeutschland im 10-Jahres-Vergleich gestiegen* (Pressemitteilung vom 22. September 2016 – 334/16). Wiesbaden: Statistisches Bundesamt.
- Strietholt, R., Manitiuss, V., Berkemeyer, N. & Bos, W. (2015). Bildung und Bildungsungleichheit an Halb- und Ganztagschulen. *Zeitschrift für Erziehungswissenschaft*, 18(4), 737–761. <https://doi.org/10.1007/s11618-015-0634-6>
- Stubbe, T. C., Schwippert, K. & Wendt, H. (2016). Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 299–316). Münster: Waxmann.
- Titz, C., Weber, S., Wagner, H., Ropeter, A., Geyer, S. & Hasselhorn, M. (Hrsg.). (2020). *Sprach- und Schriftsprachförderung wirksam gestalten: Innovative Konzepte und Forschungsimpulse*. Stuttgart: Kohlhammer.
- Tophoven, S., Lietzmann, T., Reiter, S. & Wenzig, C. (2017). *Armutsmuster in Kindheit und Jugend. Längsschnittbetrachtungen von Kinderarmut*. Verfügbar unter: https://www.bertelsmann-stiftung.de/fileadmin/files/Projekte/Familie_und_Bildung/Studie_WB_Armutsmuster_in_Kindheit_und_Jugend_2017.pdf
- UNESCO. (2003). International Standard Classification of Education, ISCED 1997. In J. H. Hoffmeyer-Zlotnik & C. Wolf (Hrsg.), *Advances in cross-national comparison. A European working book for demographic and socio-economic variables* (S. 195–220). New York: Plenum Press.
- Valtin, R. & Tarelli, I. (2014). *Lesekompetenz nachhaltig stärken. Evidenzbasierte Maßnahmen und Programme*. Berlin: Deutsche Gesellschaft für Lesen und Schreiben.
- vbw (Vereinigung der Bayerischen Wirtschaft e. V.). (Hrsg.). (2013). *Zwischenbilanz Ganztagsgrundschulen: Betreuung oder Rhythmisierung?* Münster: Waxmann.
- vbw (Vereinigung der Bayerischen Wirtschaft e. V.). (Hrsg.). (2017). *Bildung 2030 – veränderte Welt. Fragen an die Bildungspolitik*. Münster: Waxmann.
- Vester, M. (2004). Die Illusion der Bildungsexpansion. Bildungsöffnungen und soziale Segregation in der Bundesrepublik Deutschland. In S. Engler & B. Kraus (Hrsg.), *Das kulturelle Kapital und die Macht der Klassenstrukturen. Sozialstrukturelle Verschiebungen und Wandlungsprozesse des Habitus* (S. 13–54). Weinheim: Juventa.
- Vester, M. (2005). Die selektive Bildungsexpansion. Die ständische Regulierung der Bildungschancen in Deutschland. In P. A. Berger & H. Kahlert (Hrsg.), *Institutionalisierte Ungleichheiten. Wie das Bildungswesen Chancen blockiert* (S. 39–70). Weinheim: Juventa.

- Walzebug, A. (2015). *Sprachlich bedingte soziale Ungleichheit. Theoretische und empirische Betrachtungen am Beispiel mathematischer Testaufgaben und ihrer Bearbeitung*. Münster: Waxmann.
- Weis, M., Müller, K., Mang, J., Heine, J.-H., Mahler, N. & Reiss, K. (2019). Soziale Herkunft, Zuwanderungshintergrund und Lesekompetenz. In K. Reiss, M. Weis, E. Klieme & O. Köller (Hrsg.), *PISA 2018. Grundbildung im internationalen Vergleich* (S. 129–162). Münster: Waxmann.
- Wendt, H., Goy, M., Walzebug, A. & Valtin, R. (2016). Bildungsangebote an Ganz- und Halbtagsgrundschulen in Deutschland. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 225–245). Münster: Waxmann.
- Wendt, H., Kasper, D., Bos, W., Vennemann, M. & Goy, M. (2017). Wie viele Punkte auf der TIMSS-Metrik entsprechen einem Lernjahr? Leistungszuwächse in Mathematik und Naturwissenschaften am Ende der Grundschulzeit. In T. Eckert & B. Gniewosz (Hrsg.), *Bildungsgerechtigkeit* (S. 121–153). Wiesbaden: Springer VS.
https://doi.org/10.1007/978-3-658-15003-7_8

Kapitel 10

Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund

Heike Wendt, Knut Schwippert, Tobias C. Stubbe und Donieta Jusufi

10.1 Einleitung

Die schulbezogene Integration von Schülerinnen und Schülern mit Zuwanderungshintergrund ist ein wichtiges bildungspolitisches Ziel (Autorengruppe Bildungsberichterstattung, 2016). Fachliche Kompetenzen sind eine entscheidende Voraussetzung für die weitere Schullaufbahn, eine gesellschaftliche Teilhabe und ein wichtiger Bestandteil kultureller Integration. Der Grundschule kommt die Aufgabe zu, Kindern grundlegende Kompetenzen zu vermitteln. Es gilt ausdrücklich als bildungspolitisches Ziel, sowohl zuwanderungsbezogene Disparitäten in der Bildungsbeteiligung als auch im Kompetenzerwerb möglichst gering zu halten und zu reduzieren (KMK, 2002).

Migration hat viele Facetten, die bedeutend für die Bildungsintegration sind. Die Integration und Förderung von Kindern und Jugendlichen, die noch nicht lange in Deutschland leben oder deren Eltern nach Deutschland zugewandert sind, stellt eine Bereicherung des Schulalltags dar, ist aber auch mit speziellen Anforderungen an die Gestaltung von Schule und Unterricht verbunden (KMK, 2015; McElvany, Jungermann, Bos & Holtappels, 2017).

Kompetenzstände und -entwicklungen von Kindern und Jugendlichen mit Zuwanderungshintergrund im Bildungssystem fortlaufend zu beobachten, ist eine Aufgabe des Bildungsmonitorings. Im Rahmen der *Trends in International Mathematics and Science Study* (TIMSS) werden daher seit 2007 Disparitäten in mathematischen und naturwissenschaftlichen Kompetenzen von Grundschulkindern in Deutschland mit und ohne Migrationshintergrund im internationalen Vergleich untersucht. Mithilfe der TIMSS-Daten ist es möglich, auf repräsentativer Basis Leistungsdisparitäten in ihren Größenordnungen sowohl im internationalen Vergleich als auch im nationalen Trendvergleich von 2007 bis 2019 einzuordnen. Damit können Veränderungen in Leistungsständen von Viertklässlerinnen und Viertklässlern aus Familien mit und ohne Migrationshintergrund in Deutsch-

land über einen Zeitraum von mehr als zehn Jahren untersucht werden. Analysen von TIMSS 2019 sind für die Betrachtung von zuwanderungsbezogenen Disparitäten in zweierlei Hinsicht besonders interessant: Angesichts der im letzten Jahrzehnt gestiegenen Zahl neu zugewanderter Kinder und Jugendlicher im schulpflichtigen Alter (Autorengruppe Bildungsberichterstattung, 2016) ist zu untersuchen, welche Bedeutung dies für mögliche Veränderungen von Leistungsständen von Kindern mit einem Migrationshintergrund hat. Zudem gilt es zu untersuchen, inwieweit sich die Ergebnisse aus TIMSS 2011 und 2015 (Tarelli, Schwippert & Stubbe, 2012; Wendt, Schwippert & Stubbe, 2016), die auf eine Verringerung von Leistungsunterschieden zwischen Kindern mit und ohne Migrationshintergrund in den Naturwissenschaften hindeuteten, im Trend fortgesetzt haben und auch in Mathematik sichtbar werden.

Das Kapitel stellt zunächst Forschungsbefunde zu zuwanderungsbezogenen Unterschieden in mathematischen und naturwissenschaftlichen Leistungen sowie theoretische Erklärungsansätze vor. Im Zentrum des Beitrags steht die Betrachtung von mathematischen und naturwissenschaftlichen Kompetenzen von Viertklässlerinnen und Viertklässlern mit und ohne Migrationshintergrund im internationalen Vergleich sowie im Trend für Deutschland seit 2007. Weiter werden die Bedeutung der Sprache sowie des sozioökonomischen Hintergrundes als mögliche Erklärungsfaktoren für beobachtbare Leistungsunterschiede untersucht. Im Sinne eines breiteren Kompetenzverständnisses werden darüber hinaus Unterschiede in den jeweiligen fachbezogenen Selbstkonzepten und Einstellungen betrachtet sowie die Wahrnehmung zur sozialen Eingebundenheit untersucht.

10.2 Migration und Schulleistungen

Befunde aus TIMSS beziehungsweise der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) belegen seit 2001, dass Viertklässlerinnen und Viertklässler mit Migrationshintergrund in Deutschland in Mathematik und in den Naturwissenschaften, wie auch im Lesen, in den niedrigen Kompetenzstufen überrepräsentiert, in den höchsten Kompetenzstufen hingegen deutlich unterrepräsentiert sind (Bonsen, Kummer & Bos, 2008; Tarelli et al., 2012; Schurig, Wendt, Kasper & Bos, 2015; Schwippert, Bos, & Lankes, 2003; Schwippert, Wendt & Tarelli, 2012; Wendt et al., 2016; Wendt & Schwippert, 2017). Die Leistungsunterschiede liegen in Mathematik in der Größenordnung dessen, was Kinder in einem Schuljahr durchschnittlich dazulernen, in den Naturwissenschaften sogar noch deutlich darüber (Vennemann, 2018; Wendt, Kasper, Bos, Vennemann & Goy, 2017).

Leistungsunterschiede, die zuungunsten von Schülerinnen und Schülern aus Familien mit Zuwanderungsgeschichte ausfallen, sind durch ein komplexes Zusammenspiel individueller Merkmale und Benachteiligungen sowie Strukturen und Praktiken des Bildungssystems zu erklären. Studien zeigen sehr deutlich, dass ein nicht unerheblicher Teil beobachteter Leistungsunterschiede zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund durch sozioökonomische Benachteiligungen bedingt ist (z. B. Autorengruppe Bildungsberichterstattung, 2016; Tarelli et al., 2012). Zudem spielen Sprachkompetenzen eine bedeutsame Rolle. Da die Unterrichtssprache an den Schulen in der Regel Deutsch ist, sind Kenntnisse der deutschen Sprache für positive Bildungsverläufe notwendig (Gogolin, 1994; Gogolin, Neumann & Roth, 2003). Sprache ist für je-

des Fach ein wichtiges Lernmedium: Lernende, die nicht in der Lage sind, den Erklärungen der Lehrkraft zu folgen oder Fachtexte und Aufgaben nicht verstehen, können Lerngelegenheiten nicht optimal nutzen (Prediger, 2013). Aber auch die Sprache selbst bildet in jedem Fach einen wichtigen Lerngegenstand, ist doch die erfolgreiche Nutzung erlernter Fachwörter und spezifischer Redemittel eine wichtige Voraussetzung für den weiteren Wissenserwerb sowie die Leistungsbewertung (Esser, 2006; KMK, 2004). Und schließlich ist die Sprache auch ein Symbol, mit dem sich unter Umständen Stereotype aktivieren lassen, die sich in der Folge – beispielsweise in Form von diskriminierendem Verhalten der Lehrkräfte – auf den Kompetenzerwerb von Kindern auswirken können (Esser, 2006).

Der enge Zusammenhang zwischen sprachlichen, mathematischen und naturwissenschaftlichen Kompetenzen ist durch zahlreiche empirische Studien gut belegt (z. B. Heinze, Herwartz-Emden & Reiss, 2007). Ergebnisse aus verschiedenen repräsentativen Schulleistungsuntersuchungen zeigen, dass Schülerinnen und Schüler mit Migrationshintergrund, deren Eltern mit ihnen zu Hause Deutsch sprechen, sowohl in Mathematik als auch Naturwissenschaften – auch nach Kontrolle des sozioökonomischen Status und des Bildungshintergrundes der Eltern – bessere Leistungen aufweisen als jene, deren Eltern nicht mit ihren Kindern zu Hause Deutsch sprechen (Tarelli et al., 2012; Wendt et al., 2016). Dabei spielt auch die Art, wie Sprache(n) in der Familie erworben, genutzt und gefördert werden, eine Rolle für den Kompetenzerwerb (Kempert et al., 2016). Sprachlich bedingte Benachteiligungen entstehen, wenn Sprachkenntnisse nicht ausreichend in vor-, inner- und außerschulischen Kontexten und Förderangeboten erworben und weiterentwickelt werden können (Stanat & Christensen, 2006). Benachteiligungen können durch Faktoren wie Klassenzusammensetzungen oder sozialräumliche Segregation verstärkt werden, auch wenn hier begrenzte Gelegenheiten zur Weiterentwicklung von Sprachkompetenzen geboten werden (Esser, 2006; Kristen & Olczyk, 2013; Rjosk et al., 2014; Stanat, Schwippert & Gröhlich, 2010).

Über den Sprachgebrauch hinaus kann auch das soziokulturelle Milieu, in dem ein Kind aufwächst und welches auch durch den Zuwanderungsstatus geprägt sein kann, familiäre Prozessmerkmale (wie kulturelle Aktivitäten oder die kommunikative Praxis in Familien) beeinflussen. Diese sind ihrerseits für die Ausprägungen individueller Merkmale von Kindern (wie Einstellungen und Verhaltensweisen) und den Kompetenzerwerb bedeutsam. Befinden sich in den außerschulischen Lernkontexten und persönlichen Netzwerken der Kinder nur wenige Personen, die über Ressourcen verfügen, die für Bildungsprozesse relevant sind, kann sich dies nachteilig auf die Kompetenzentwicklung auswirken (Kristen & Olczyk, 2013).

Im Kontext von Zuwanderung spielen weitere Aspekte wie Gründe, Zeitpunkt sowie Bedingungen von Zuwanderung eine Rolle. Brüche in Bildungsbiografien, neue Lebensumstände, Trennung von wichtigen Bezugspersonen oder auch traumatische Erlebnisse, die mit besonderen psychischen Belastungen gerade für Kinder und Jugendliche einhergehen können, können es erschweren, sich im Bildungssystem zurechtzufinden (z. B. Kunz, 2008; Stanat, 2008). Aber auch Unsicherheiten im Aufenthaltsstatus können sich negativ auf Einstellungen, Verhaltensweisen und Schulerfolg auswirken (Scherr & Niermann, 2012). Netzwerken und Unterstützungsstrukturen kommt hier eine besondere Bedeutung zu (Kristen & Olczyk, 2013).

Auch Merkmale der Institution Schule beeinflussen Leistungen von Grundschulkindern in unterschiedlicher Weise (Gogolin, 1994; Gomolla & Radtke, 2014; Hasse & Schmidt, 2012; Vennemann, 2018). Hier ist zu fragen, inwieweit Schülerinnen und Schüler mit Zuwanderungsgeschichte Unterstützungsangebote nutzen können, ihre Kompetenzen gewürdigt und anerkannt werden und sie die gleiche Qualität schulischer Sozialbeziehungen und akademischer Entwicklungsmöglichkeiten erfahren wie ihre Mitschülerinnen und Mitschüler ohne Zuwanderungsgeschichte. Diese Aspekte spielen eine besondere Rolle für das schulische Wohlbefinden, das Lern- und Leistungsverhalten von Schülerinnen und Schülern, die Motivation und auch das Vertrauen in die eigenen fachspezifischen Fähigkeiten (Mehringer, 2013). Die empirische Befundlage lässt Zusammenhänge zwischen Migrationshintergrund und zentralen motivationalen Komponenten des schulischen Lernens vermuten, allerdings fallen die Ergebnisse in Abhängigkeit von betrachteten Indikatoren, Fächern und der Stichprobe unterschiedlich aus. Hinsichtlich der generellen Einstellungen zur Schule konnte im repräsentativen IQB-Bildungstrend 2016 für sechs Länder der Bundesrepublik Deutschland und auch im Bundesdurchschnitt ein kleiner signifikanter Effekt zugunsten von Kindern mit Migrationshintergrund festgestellt werden (Rjosk, Haag, Heppt & Stanat, 2017). Dieser Befund korrespondiert zwar mit älteren Studienergebnissen (Lehmann, Peek & Gänßfuß, 1997; Rauer & Schuck, 2003), hingegen können eine Reihe von Studien keine signifikanten Unterschiede in den Schuleinstellungen feststellen (Harazd & Schürer, 2006; Gisdakis, 2007; Kohl, Striegler, Peters & Leyendecker, 2011; Nicklaussen, 2012; van Ophuysen, 2008). In anderen Untersuchungen bringen Grundschulkindern mit Migrationshintergrund ein geringeres psychisches Wohlbefinden zum Ausdruck (Hofmann, Bonitz, Lippert & Gläser-Zikuda, 2018) oder fühlen sich von ihren Lehrkräften in geringerem Ausmaß angenommen (Walsen, 2013).

Fachspezifische Motivation und das Vertrauen in die eigenen fachspezifischen Fähigkeiten sind für Grundschulkindern mit und ohne Migrationshintergrund, insbesondere für den Mathematik- und Sachunterricht in Deutschland, wenig erforscht: Für Mathematik zeigte sich im IQB-Bildungstrend 2016 ein kleiner signifikanter Effekt dahingehend, dass Kinder mit Zuwanderungshintergrund eine höhere Lernfreude zum Ausdruck bringen als Kinder ohne Zuwanderungsgeschichte (Rjosk et al., 2017). Auch lassen Studien vermuten, dass Kinder mit Migrationshintergrund im Fach Mathematik ihre eigenen Fähigkeiten sehr optimistisch und unerwartet hoch einschätzen (Faber, Tiedemann & Billmann-Mahecha, 2011; Roebbers, Mecheril & Schneider, 1998). Zuwanderungsbedingte motivationale Unterschiede sind für den Sachunterricht nicht repräsentativ erforscht. Allerdings konnten Lossen, Tillmann, Holtappels, Rollett und Hannemann (2016) kleine Effekte für niedriger ausgeprägte sachunterrichtsbezogene Selbstkonzepte von Kindern mit Migrationshintergrund beobachten.

Ein etwas ungünstigeres Bild zeigt sich für andere Kompetenzbereiche: Im IQB-Bildungstrend 2016 ergab sich zwar ein kleiner signifikanter Effekt dahingehend, dass Kinder mit Zuwanderungshintergrund auch im Fach Deutsch eine höhere Lernfreude zum Ausdruck bringen als Kinder ohne Zuwanderungsgeschichte (Rjosk et al., 2017). Allerdings zeigt sich auch, dass Grundschulkindern mit Migrationshintergrund ein signifikant geringeres Leseverhalten zum Ausdruck bringen als ihre Mitschülerinnen und Mitschüler ohne Migrationshintergrund (Wendt & Schwippert, 2017). Unterschiede hinsichtlich der Lesemotivation zwischen Kindern mit und ohne Zuwanderungsgeschichte sind gering und lassen sich nicht in allen Studien nachweisen (Kigel, McElvany & Becker, 2015; Miyamoto,

Pfost & Artelt, 2017; Wendt & Schwippert, 2017). Die Befundlage lässt jedoch signifikant geringere Selbsteinschätzungen der eigenen Fähigkeiten im Lesen (Mielke, Goy & Pietsch, 2006; Roebers et al., 1998; Wendt & Schwippert, 2017) sowie der Rechtschreibung (Faber, 2013) vermuten.

10.3 Indikatoren des Migrationshintergrundes

Eine Herausforderung bei der Untersuchung von Leistungsdisparitäten zwischen Kindern aus Familien mit und ohne Migrationshintergrund im internationalen Vergleich stellen unterschiedliche Möglichkeiten bei der Definition und Erfassung des Migrationshintergrundes dar. Familienkonstellationen, Lebensbedingungen und Zuwanderungsgeschichten von Kindern und Jugendlichen mit Migrationshintergrund sind vielfältig und lassen sich in Gänze im Rahmen von Schulleistungsuntersuchungen nicht differenziert erfassen. Die Nutzung von übergeordneten kategorialen Unterscheidungen von Kindern in Gruppen mit und ohne Migrationshintergrund – wie sie auch in TIMSS zum Einsatz kommt – ist dabei nicht unproblematisch (z.B. Scherr & Niermann, 2012). Für die Interpretation ist zu berücksichtigen, dass der Kategorie Migrationshintergrund sehr heterogene Teilpopulationen zugeordnet werden, die sich in einer Vielzahl bildungsrelevanter Merkmale nicht eindeutig voneinander oder von der Gruppe ohne Migrationshintergrund abgrenzen lassen. Ein Interpretationsschluss von mittleren Gruppenmerkmalen auf Individuen und Subgruppen ist unzulässig. Auch bei weiteren Differenzierungen nach Nationalität oder Herkunft handelt es sich um indirekte Indikatoren für potenziell bildungsrelevante Unterschiede (wie z.B. Sprache oder geringe Vertrautheit mit dem Bildungssystem).

Gleichwohl sind empirische Befunde zu zugewanderungsbezogenen Disparitäten notwendig, um bestehende Benachteiligungen sichtbar zu machen. Ein fortlaufendes Monitoring von Disparitäten erfordert dabei eine konsistente und kontinuierliche Verwendung von Indikatoren. Für die Analysen in diesem Kapitel werden daher wie in den vergangenen Studienzyklen Schülerinnen und Schüler dann als Kinder mit Migrationshintergrund beschrieben, wenn mindestens ein Elternteil nicht im Teilnehmerstaat geboren wurde. Diese Frage wurde in Deutschland bereits seit 2007, international jedoch erst ab 2015 in TIMSS erhoben. Ein Vergleich mit den vorangegangenen Studienzyklen von 2007 und 2011 kann daher nur auf nationaler Ebene vorgenommen werden. Darüber hinaus wird aufgrund der Umstellung bei den abgefragten Kategorien nach dem heimischen Sprachgebrauch in den internationalen Hintergrundfragebögen für Schülerinnen und Schüler zwischen 2007, 2011, 2015 und 2019 ebenfalls auf einen Vergleich dieses Merkmals im Trend verzichtet. In diesem Kapitel wird für den internationalen Vergleich der Frage nach dem Geburtsland der Eltern der Vorzug vor dem heimischen Sprachgebrauch gegeben, da das Geburtsland zuverlässiger über Migrationserfahrungen in den Familien Auskunft gibt als der heimische Sprachgebrauch.

Aufgrund des Erhebungsdesigns ist eine darüber hinausgehende Differenzierung nach Herkunftsgruppen oder Generationen nicht möglich. Hier wird verwiesen auf den IQB-Bildungstrend 2016 (Rjosk et al., 2017), der bereits die im Rahmen von IGLU 2006 festgestellten höheren Anteile von Personen mit Migrationshintergrund in Stadtstaaten bestätigt (Haag, Böhme & Stanat, 2012; Schwippert, Hornberg & Goy, 2008) und aufzeigt, dass sich die Disparitäten deutlich zwischen unterschiedlichen Geburtsstaaten der Kinder oder Eltern unter-

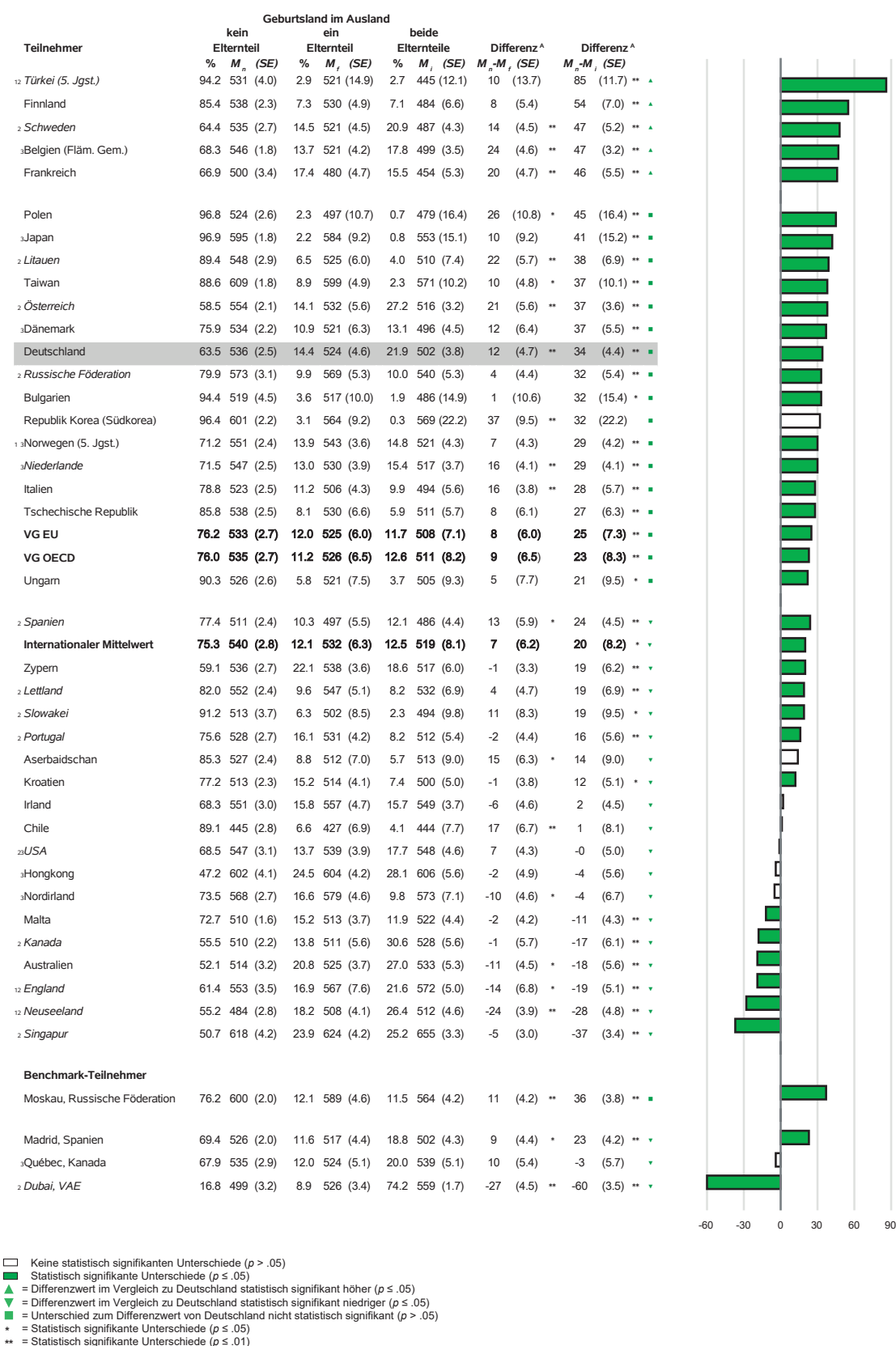
scheiden. Ebenso ist das Stichprobendesign von TIMSS nicht dafür ausgelegt, verlässliche Aussagen zu Leistungen von vor oder während der Grundschulzeit nach Deutschland zugewanderten Kindern zu treffen (Wendt et al., 2016). In TIMSS 2019 machen Schülerinnen und Schüler, die wie ihre Eltern nicht in Deutschland geboren wurden, etwa 7 bis 9 Prozent aller Viertklässlerinnen und Viertklässler aus. Etwa 2 Prozent von ihnen nahmen aus Testbefreiungsgründen nicht an TIMSS teil. Aufgrund der geringen Fallzahlen werden ihre Ergebnisse aber nicht gesondert ausgewiesen, sondern zählen zu der Gruppe von Schülerinnen und Schülern mit Migrationshintergrund (beide Elternteile). In diesem Kapitel wurde zudem für die nationalen Analysen von naturwissenschaftlichen Disparitäten im Trend die Stichprobe um 81 Schülerinnen und Schüler reduziert, die zu wenige Testaufgaben bearbeitet haben, als dass eine reliable und valide Schätzung ihrer naturwissenschaftlichen Kompetenzen möglich wäre und damit mögliche Unterschiede zwischen den Schülerinnen- und Schülergruppen über- oder unterschätzt werden würden.

Um die Übersichtlichkeit des internationalen Vergleichs zu erhöhen, fokussieren die nachfolgend berichteten Gegenüberstellungen auf die TIMSS-Teilnehmerstaaten, die Mitglieder der *Europäischen Union* (EU) sind, die der *Organisation for Economic Co-operation and Development* (OECD) angehören und schließlich auch auf die Teilnehmerstaaten, deren Leistungswerte auf der Gesamtskala Mathematik beziehungsweise Naturwissenschaften statistisch signifikant über den für Deutschland berichteten Mittelwerten liegen oder sich von diesen nicht signifikant unterscheiden.

10.4 Mathematische und naturwissenschaftliche Kompetenzen der Schülerinnen und Schüler nach Migrationshintergrund in den Familien im internationalen Vergleich

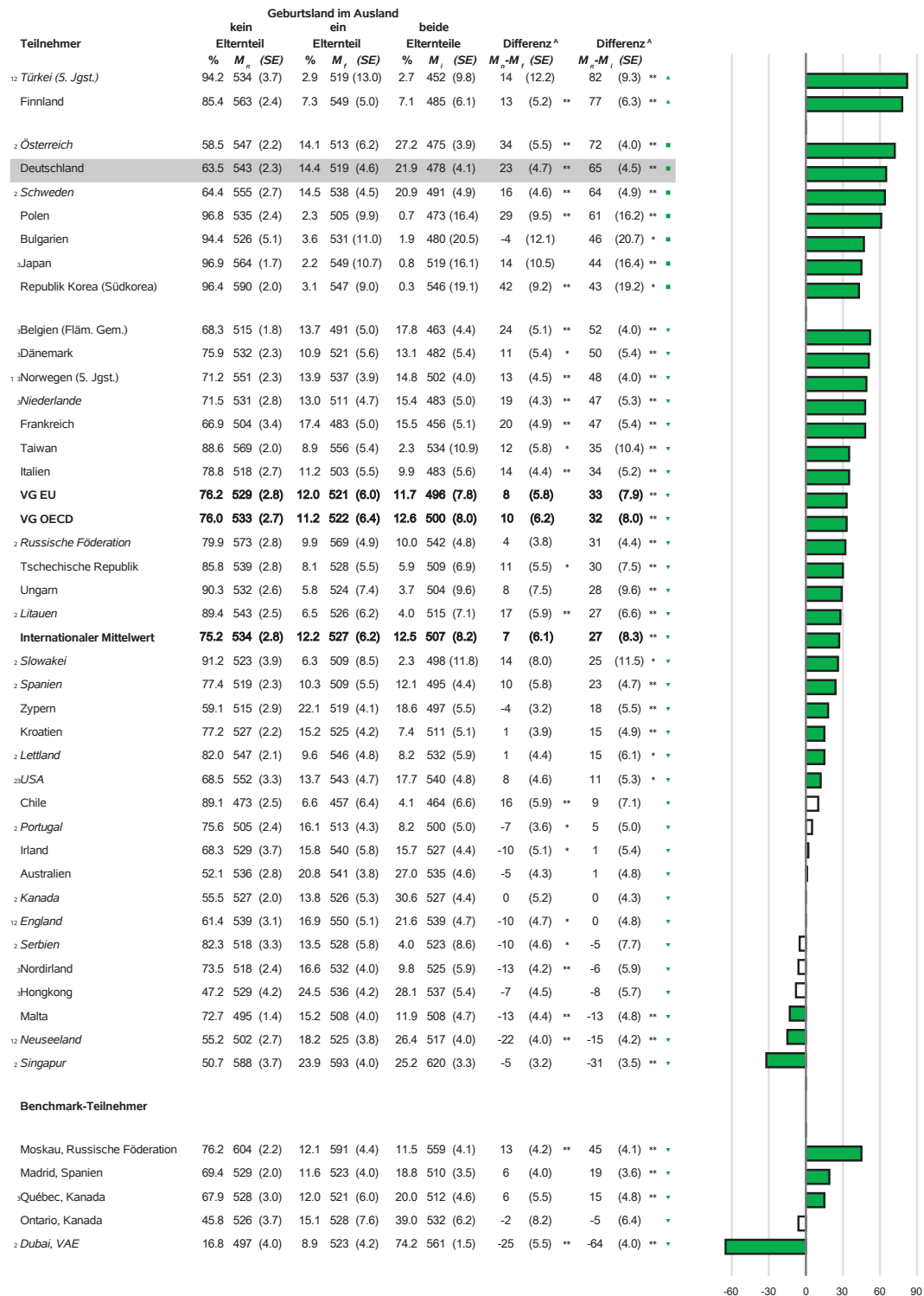
Dass der schulische Erfolg und der Zuwanderungshintergrund der Schülerinnen und Schüler zusammenhängen, ist in vielen Teilnehmerstaaten ein immer wieder festzustellender Befund. Um Schülerinnen- und Schülergruppen mit und ohne Migrationshintergrund vergleichend erfassen zu können, haben sich die an TIMSS beteiligten Staaten darauf verständigt, die Eltern der Viertklässlerinnen und Viertklässler nach ihren Geburtsstaaten zu fragen. So können Schülerinnen und Schüler unterschieden werden, deren Eltern angeben, dass (1) beide im Ausland, (2) ein Elternteil im Ausland und eines im Teilnehmerstaat oder (3) beide im jeweiligen Teilnehmerstaat (kein Elternteil im Ausland) geboren wurden. In Abbildung 10.1 werden die Testleistungen der Schülerinnen und Schüler in Mathematik nach Migrationshintergrund der Eltern im internationalen Vergleich dargestellt. Im linken Teil der Abbildung sind die Verteilung des Geburtslandes der Eltern und die Leistungen der jeweiligen Gruppen im internationalen Vergleich dargestellt. Für Deutschland zeigt sich, dass mit 36,5 Prozent mehr als jedes dritte Grundschulkind in einer Familie mit Zuwanderungshintergrund aufwächst. Vergleichbare Bevölkerungszusammensetzungen lassen sich auch in vielen anderen europäischen Staaten wie beispielsweise Österreich, der Flämischen Gemeinschaft in Belgien, Frankreich, Schweden, England oder Irland beobachten.

Im rechten Teil der Abbildung 10.1 ist der Leistungsvorsprung in Mathematik für die Kinder visualisiert, deren Eltern angeben, beide im Teilnehmerstaat ge-

Abbildung 10.1: Testleistungen der Schülerinnen und Schüler in Mathematik nach Migrationshintergrund der Eltern im internationalen Vergleich

boren worden zu sein, vor jenen, deren Eltern angeben, beide im Ausland geboren worden zu sein. Die Staaten sind nach Höhe der Leistungsdisparitäten zwischen Kindern ohne und mit Migrationshintergrund (kein Elternteil vs. beide Elternteile) geordnet. Diese Leistungsunterschiede sind auch im grafischen Teil der Abbildung illustriert, wobei signifikante Unterschiede durch die grüne Färbung der Differenzbalken kenntlich gemacht sind. Die für Deutschland festgestellte Differenz liegt mit 34 Punkten in Mathematik signifikant über den mittleren Differenzen der TIMSS-Teilnehmerstaaten (internationaler Mittelwert: 20 Punkte). Kein signifikanter Unterschied liegt zwischen der Differenz in Deutschland und dem Durchschnitt der Mitglieder der EU (25 Punkte) und dem Durchschnitt der Mitglieder der OECD (23 Punkte) vor. Nur in fünf der Teilnehmerstaaten ist der Leistungsunterschied zwischen Schülerinnen und Schülern aus Familien ohne Migrationshintergrund und Schülerinnen und Schülern, bei denen beide Eltern Migrationserfahrungen haben, signifikant größer als in Deutschland. Diese Differenz beträgt in der Türkei 85 Punkte, in Finnland 54 Punkte, in Schweden und der Flämischen Gemeinschaft in Belgien jeweils 47 Punkte und in Frankreich 46 Punkte. Abbildung 10.1 kann ebenfalls entnommen werden, dass sich für 15 Teilnehmerstaaten, darunter beispielsweise die Nachbarstaaten Polen, Dänemark, Österreich und die Niederlande, vergleichbar hohe migrationsbedingte Leistungsdisparitäten in Mathematik beobachten lassen wie in Deutschland. In zwölf Staaten fallen entsprechende migrationsbedingte Disparitäten signifikant niedriger aus als in Deutschland. In sechs Staaten zeigen Kinder aus Familien mit Migrationshintergrund (beide Elternteile) signifikant bessere Leistungen in Mathematik als Kinder aus Familien ohne Migrationshintergrund (kein Elternteil). Den meisten dieser Teilnehmerstaaten ist eine besonders selektive Einwanderungspolitik gemein (Keeley, 2009).

Verglichen mit den Kompetenzunterschieden in Mathematik zwischen Schülerinnen und Schülern, deren Elternteile beide nicht im Ausland geboren wurden mit jenen, die zwei im Ausland geborene Elternteile haben, sind in den naturwissenschaftlichen Kompetenzen insgesamt größere Unterschiede festzustellen (siehe Abbildung 10.2). Die in Deutschland festzustellende Differenz liegt in den Naturwissenschaften bei 65 Punkten. Dabei weisen nur zwei Staaten, namentlich die Türkei (82 Punkte) und Finnland (77 Punkte), signifikant größere Leistungsdisparitäten auf und nur in sechs weiteren Teilnehmerstaaten, darunter die EU-Mitglieder Österreich, Polen und Schweden, lassen sich vergleichbar große Leistungsdisparitäten beobachten. Hingegen fallen in den meisten Teilnehmerstaaten die migrationsbezogenen Leistungsdisparitäten in den naturwissenschaftlichen Kompetenzen signifikant geringer aus als in Deutschland. Entsprechend liegen die Mittelwerte aller Vergleichsgruppen mit 33 Punkten für die EU, 32 Punkten für die OECD und 27 Punkten im internationalen Durchschnitt signifikant unter den für Deutschland festzustellenden Leistungsdisparitäten.

Abbildung 10.2: Testleistungen der Schülerinnen und Schüler in Naturwissenschaften nach Migrationshintergrund der Eltern im internationalen Vergleich□ Keine statistisch signifikanten Unterschiede ($p > .05$)■ Statistisch signifikante Unterschiede ($p \leq .05$)▲ = Differenzwert im Vergleich zu Deutschland statistisch signifikant höher ($p \leq .05$)▼ = Differenzwert im Vergleich zu Deutschland statistisch signifikant niedriger ($p \leq .05$)■ = Unterschied zum Differenzwert von Deutschland nicht statistisch signifikant ($p > .05$)* = Statistisch signifikante Unterschiede ($p \leq .05$)** = Statistisch signifikante Unterschiede ($p \leq .01$)

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

¹ = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.² = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.³ = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.^A = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

10.5 Lebensbedingungen von Schülerinnen und Schülern mit und ohne Migrationshintergrund in Deutschland

Ein rezenter Befund für Deutschland ist, dass sich die Einkommens- und Lebensbedingungen von Familien mit und ohne Migrationshintergrund in Deutschland deutlich unterscheiden. Für die Interpretation von beobachtbaren Leistungsdisparitäten sind entsprechende Unterschiede zu berücksichtigen. In Tabelle 10.1 sind Überblicksstatistiken dargestellt. Der Anteil von Viertklässlerinnen und Viertklässlern aus Familien mit Migrationshintergrund liegt in Deutschland bei 36.5 Prozent, wobei für 21.9 Prozent der Viertklässlerinnen und Viertklässler gilt, dass beide Elternteile im Ausland geboren wurden.

Erwartungsgemäß lassen sich Unterschiede im familiären Sprachgebrauch beobachten. Für den überwiegenden Anteil der Schülerinnen und Schüler mit Eltern ohne Migrationshintergrund wird angegeben, dass *immer* (85.5%) beziehungsweise *fast immer* (11.5%) zu Hause Deutsch gesprochen wird. Auch in 39.4 Prozent aller Familien, in denen ein Elternteil in Deutschland und eines im Ausland geboren wurde, wird nach Angabe der Eltern immer Deutsch gesprochen, in weiteren 28.6 Prozent fast immer Deutsch. Auch in mehr als jeder dritten Familie, in der beide Elternteile nicht in Deutschland geboren wurden, wird *immer* (14.5%) oder *fast immer* (23.1%) Deutsch gesprochen. In knapp zwei Drittel aller Familien, in denen beide Elternteile im Ausland geboren wurden, ist Deutsch nicht die primäre Sprache familiärer Kommunikation. Sehr wenige Kinder geben an, nie zu Hause Deutsch zu sprechen – unabhängig davon, ob nur ein Elternteil oder beide Elternteile im Ausland geboren wurden. Die Daten weisen auf eine gelungene sprachliche Integration hin.

Sehr deutliche Unterschiede lassen sich zudem in der *Armutsgefährdung* beobachten, deren Erfassung genauer in Kapitel 9 in diesem Band beschrieben ist. Während mit 14.4 Prozent etwa jedes sechste Kind ohne Migrationshintergrund in Deutschland unter Bedingungen der *Armutsgefährdung* aufwächst, ist der Anteil von Kindern mit Migrationshintergrund (beide Elternteile) mit 44.5 Prozent mehr als dreimal so hoch. 20.8 Prozent aller Familien mit einem im Ausland geborenen Elternteil leben unter Bedingungen der *Armutsgefährdung*.

Auch für die Erwerbstätigkeit der Eltern und die Dauer des Kindergartenbesuchs lassen sich Unterschiede zwischen Familien mit und ohne Migrationshintergrund beobachten. Eltern ohne Migrationshintergrund sind häufiger erwerbstätig als Eltern mit Migrationshintergrund. Die migrationsspezifischen Unterschiede fallen zwischen Vätern geringer aus (92.5% vs. 85.8%) als zwischen Müttern (86.4% vs. 64.2%). Der Anteil von Kindern, die vor der Grundschule mindestens drei Jahre in den Kindergarten gegangen sind, ist in den Familien, in denen beide Eltern im Ausland geboren wurden, am niedrigsten (63.6%) und in den Familien mit einem im Ausland und einem in Deutschland geborenen Elternteil am höchsten (80.0%). 73.7 Prozent der Familien mit zwei in Deutschland geborenen Elternteilen haben einen mindestens dreijährigen Kindergartenbesuch realisiert.

Tabelle 10.1: Angaben zur Verteilung von Viertklässlerinnen und Viertklässlern im Hinblick auf den Gebrauch der deutschen Sprache, den Erwerbsstatus von Mutter und Vater sowie den Kindergartenbesuch und die Armutsgefährdung der Familien nach Migrationshintergrund (Anteile in Prozent)

Geburtsland im Ausland	Viertklässlerinnen und Viertklässler %	Gebrauch der deutschen Sprache zu Hause				Mind. 3 Jahre Kindergarten- besuch %	Erwerbstätigkeit ^A		Armuts- gefährdung der Familie %
		immer %	fast immer %	manchmal %	nie %		Vater %	Mutter %	
beide Elternteile	21.9	14.5	23.1	52.7	9.7	63.6	85.8	64.2	44.5
ein Elternteil	14.4	39.4	28.6	30.5	1.5	80.0	89.5	74.4	20.8
kein Elternteil	63.5	85.5	11.5	3.0	0.0	73.7	92.5	86.4	14.4

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.
A = Voll- oder Teilzeit

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

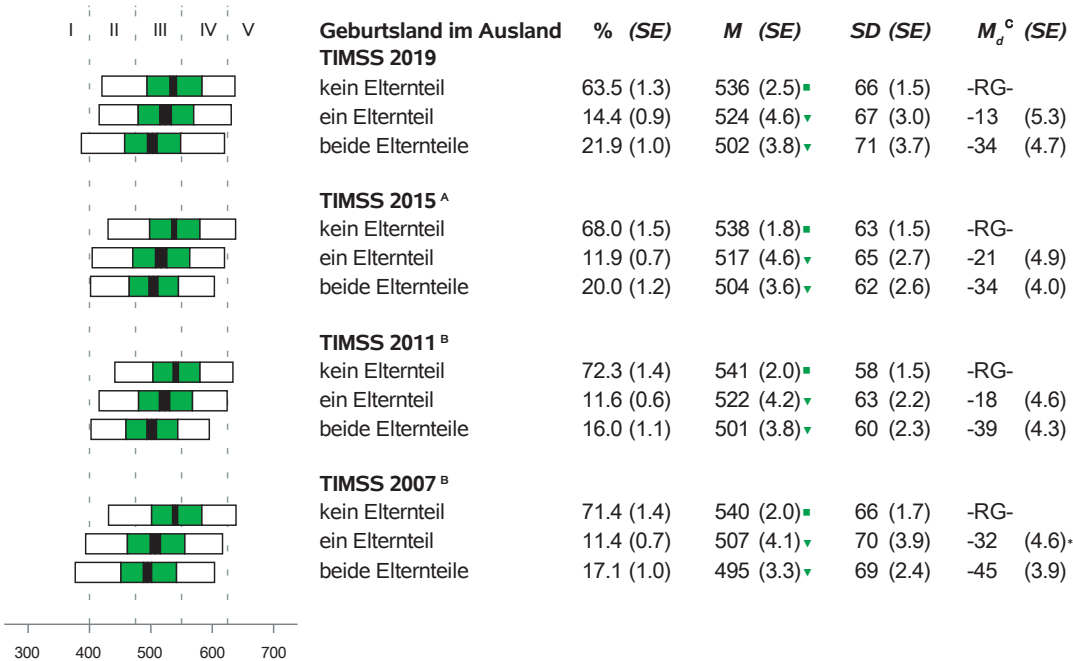
10.6 Kompetenzunterschiede von Schülerinnen und Schülern in Deutschland im Trend

Im internationalen Vergleich zeigt sich, dass die Leistungsunterschiede zwischen Schülerinnen und Schülern aus Familien mit und ohne Migrationshintergrund in Deutschland vor allem in den Naturwissenschaften signifikant höher ausfallen als in einer Reihe von anderen Teilnehmerstaaten, die Mitglieder der EU oder der OECD sind und vergleichbare Bevölkerungszusammensetzungen haben; ein Befund, der sich bereits in TIMSS 2007 (Bonsen et al., 2008), TIMSS 2011 (Tarelli et al., 2012) und TIMSS 2015 (Wendt et al., 2016) feststellen ließ. In Abbildung 10.3 sind die mathematischen Kompetenzen von Schülerinnen und Schülern aus Familien mit unterschiedlichem Migrationshintergrund im Trend dargestellt. Auch werden die Veränderungen in den prozentualen Anteilen von Schülerinnen und Schülern nach dem Geburtsland der Eltern berichtet.

Zunächst ist festzustellen, dass der Anteil von Kindern mit Migrationshintergrund in Deutschland von 29 Prozent in 2007 auf 36 Prozent in 2019 gestiegen ist. Entsprechend ist von 2007 zu 2019 der Anteil an Kindern, deren Eltern beide in Deutschland geboren wurden, von 71 Prozent auf 64 Prozent zurückgegangen. Die Veränderungen sind signifikant (siehe Tabelle 12.1 in Kapitel 12 in diesem Band).

In allen Studienzyklen ließen sich für Kinder, deren Eltern beide in Deutschland geboren wurden, signifikant höhere Mathematikleistungen beobachten als für Kinder mit Migrationshintergrund. In 2019 zeigen Schülerinnen und Schüler mit Migrationshintergrund im Vergleich zu ihren Mitschülerinnen und Mitschülern, deren Eltern beide in Deutschland geboren wurden, Mathematikleistungen, die mit 13 (ein Elternteil im Ausland geboren) beziehungsweise 34 (beide Elternteile) Leistungspunkten signifikant geringer ausfallen. Für Kinder, deren Eltern im Ausland geboren wurden, liegt der Leistungsunterschied im Rahmen dessen, was Kinder in mehr als einem halben Schuljahr dazulernen (Wendt et al., 2017). Erfreulich ist, dass die Leistungsdisparitäten in Mathematik zwischen Kindern mit einem im Ausland geborenen Elternteil und jenen ohne Migrationshintergrund in 2019 signifikant geringer sind als in 2007, hier lagen die Leistungsunterschiede noch bei 32 Punkten. Nominell zeigen sich auch geringere Leistungsdisparitäten zwischen Kindern mit beiden im Ausland geborenen

Abbildung 10.3: Testleistungen der Schülerinnen und Schüler in Mathematik in Deutschland nach Migrationshintergrund der Eltern – TIMSS 2007, 2011, 2015 und 2019 im Vergleich



Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

★ = Unterschied zu 2019 statistisch signifikant ($p \leq .05$)

▲ = Wert im Vergleich zu „kein Elternteil“ statistisch signifikant höher ($p \leq .05$)

▼ = Wert im Vergleich zu „kein Elternteil“ statistisch signifikant niedriger ($p \leq .05$)

■ = Unterschied zu „kein Elternteil“ nicht statistisch signifikant ($p > .05$)

A = Abweichungen zu den Ergebnissen von TIMSS 2015 (Wendt et al., 2016) ergeben sich aus einem veränderten Umgang mit fehlenden Werten (siehe Kapitel 2 in diesem Band).

B = Abweichungen zu den in Tarelli, Schwippert und Stubbe (2012) berichteten Standardfehlern begründen sich in einem optimierten Berechnungsverfahren.

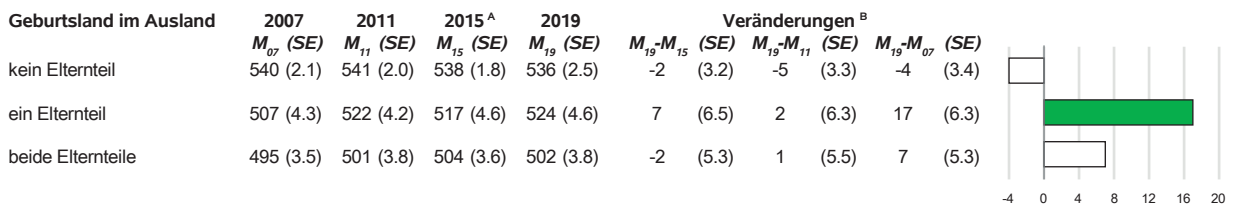
C = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

RG= Referenzgruppe

Elternteilen und Kindern ohne Migrationshintergrund, diese Veränderung lässt sich jedoch nicht zufallskritisch absichern.

Vergleicht man die mittleren Mathematikleistungen innerhalb der Gruppen mit verschiedenem Migrationshintergrund, zeigen sich für die Gruppe der Kinder ohne Migrationshintergrund sowie der Kinder mit beiden im Ausland geborenen Elternteilen zwischen 2007 und 2019 keine signifikanten Unterschiede (siehe Abbildung 10.4). Für Kinder mit einem in Deutschland und einem im Ausland geborenen Elternteil lassen sich mit einem Zuwachs von 17 Leistungspunkten deutliche und signifikante Verbesserungen in den Mathematikleistungen zwischen 2007 und 2019 feststellen – eine positive Entwicklung, die sich bereits mit den Ergebnissen von TIMSS 2011 und 2015 andeutete (Tarelli et al., 2012; Wendt et al., 2016). Nominell zeigen auch Kinder mit beiden im Ausland geborenen Elternteilen bessere mathematische Leistungen in 2019 als in 2007, diese Differenz fällt jedoch nicht signifikant aus.

In den Naturwissenschaften fallen auch in TIMSS 2019 die Leistungsunterschiede zwischen den Schülerinnen und Schülern mit unterschiedlichem Migrationshintergrund im Vergleich zu Mathematik deutlich höher aus (Abbildung 10.5). Schülerinnen und Schüler ohne Migrationshintergrund erzielen mit 545 Punkten Leistungsvorsprünge von 22 beziehungsweise 60 Punkten im Vergleich zu ihren Mitschülerinnen und Mitschülern aus Familien mit Migrationshintergrund. Abweichungen in den Differenzwerten zur Abbildung 10.2 ergeben sich aus einer leicht differenten Datengrundlage (siehe Abschnitt 10.3). Dieser Leistungsunterschied von 60 Punkten zwischen Kindern ohne Migrationshinter-

Abbildung 10.4: Unterschiede in den mittleren Mathematikleistungen im Trend in Deutschland nach Migrationshintergrund der Eltern – TIMSS 2007, 2011, 2015 und 2019 im Vergleich

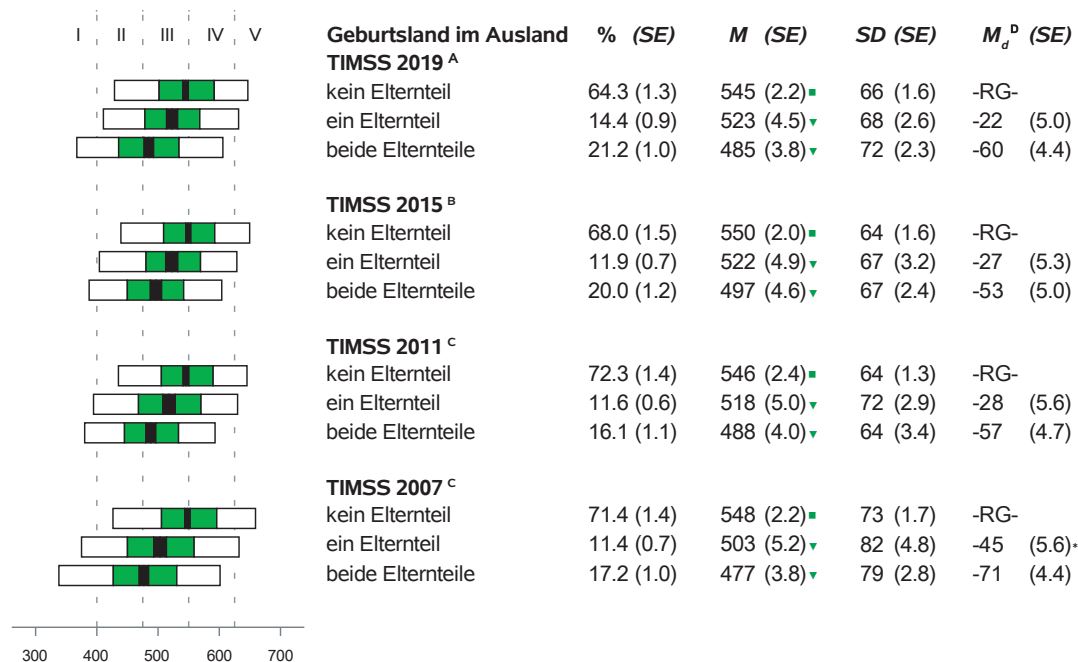
Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

Keine statistisch signifikante Veränderung zwischen 2007 und 2019 ($p > .05$)Statistisch signifikante Veränderung zwischen 2007 und 2019 ($p \leq .05$)^A = Abweichungen zu den Ergebnissen von TIMSS 2015 (Wendt et al., 2016) ergeben sich aus einem veränderten Umgang mit fehlenden Werten (siehe Kapitel 2 in diesem Band).^B = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

grund und solchen, deren Eltern beide im Ausland geboren wurden, ist beachtlich und entspricht mehr als eineinhalb Schuljahren (Wendt & Schwippert, 2017). Im Vergleich zu 2007 liegt diese Leistungsdisparität auf einem ähnlichen Niveau. Positive Entwicklungen, die sich in TIMSS 2015 abzeichneten (Wendt, Schwippert & Stubbe, 2016), lassen sich für migrationsspezifische Leistungsdisparitäten für die Gruppe der Kinder mit beiden im Ausland geborenen Elternteilen für die naturwissenschaftlichen Kompetenzen 2019 nicht mehr nachweisen. Eine signifikante Reduktion der Leistungsdisparitäten zwischen

Abbildung 10.5: Testleistungen der Schülerinnen und Schüler in Naturwissenschaften in Deutschland nach Migrationshintergrund der Eltern – TIMSS 2007, 2011, 2015 und 2019 im Vergleich

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

* = Unterschied zu 2019 statistisch signifikant ($p \leq .05$)▲ = Wert im Vergleich zu „kein Elternteil“ statistisch signifikant höher ($p \leq .05$)▼ = Wert im Vergleich zu „kein Elternteil“ statistisch signifikant niedriger ($p \leq .05$)■ = Unterschied zu „kein Elternteil“ nicht statistisch signifikant ($p > .05$)^A = Abweichungen zu Abbildung 10.2 sind in einer minimal veränderten Stichprobenszusammensetzung begründet.^B = Abweichungen zu den Ergebnissen von TIMSS 2015 (Wendt et al., 2016) ergeben sich aus einem veränderten Umgang mit fehlenden Werten (siehe Kapitel 2 in diesem Band).^C = Abweichungen zu den in Tarelli, Schwippert und Stubbe (2012) berichteten Standardfehlern begründen sich in einem optimierten Berechnungsverfahren.^D = Inkonsistenzen in den berichteten Differenzen sind im Rundungsverfahren begründet.

RG = Referenzgruppe

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Kindern mit einem und beiden in Deutschland geborenen Elternteilen von 45 auf 22 Leistungspunkte, lässt sich wie bereits in 2011 und 2015 auch weiterhin in 2019 feststellen. Eine Leistungsdisparität von 22 Punkten entspricht dem mittleren Lernzuwachs von mehr als einem halben Schuljahr und bleibt trotz positiver Entwicklungstendenz weiterhin beachtenswert.

Wie für die mathematischen Leistungen zeigen sich auch für den Bereich der Naturwissenschaften keine signifikanten Veränderungen zwischen 2007 und 2019 in den mittleren Leistungen der Kinder mit beiden in Deutschland oder beiden im Ausland geborenen Elternteilen (siehe Abbildung 10.6). Für Kinder mit einem in Deutschland und einem im Ausland geborenen Elternteil lassen sich – wie bereits in TIMSS 2011 und 2015 beobachtet (Tarelli et al., 2012; Wendt et al., 2016) – mit einem Zuwachs von 20 Leistungspunkten deutliche und signifikante Verbesserungen in den naturwissenschaftlichen Leistungen feststellen.

Abbildung 10.6: Unterschiede in den mittleren Testleistungen in Naturwissenschaften in Deutschland im Trend nach Migrationshintergrund der Eltern – TIMSS 2007, 2011, 2015 und 2019 im Vergleich

Geburtsland im Ausland	2007	2011	2015 ^A	2019 ^B	Veränderungen ^C			
	M_{07} (SE)	M_{11} (SE)	M_{15} (SE)	M_{19} (SE)	$M_{19}-M_{15}$ (SE)	$M_{19}-M_{11}$ (SE)	$M_{19}-M_{07}$ (SE)	
kein Elternteil	548 (2.2)	546 (2.4)	550 (2.0)	545 (2.2)	-5 (3.0)	-1 (3.3)	-3 (3.1)	
ein Elternteil	503 (5.2)	518 (5.0)	522 (4.9)	523 (4.5)	1 (6.7)	5 (6.8)	20 (6.9)	
beide Elternteile	477 (3.8)	488 (4.0)	497 (4.6)	485 (3.8)	-12 (6.0)	-3 (5.6)	8 (5.4)	

-5 -1 3 7 11 15 19

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

Keine statistisch signifikante Veränderung zwischen 2007 und 2019 ($p > .05$)

Statistisch signifikante Veränderung zwischen 2007 und 2019 ($p \leq .05$)

A = Abweichungen zu den Ergebnissen von TIMSS 2015 (Wendt et al., 2016) ergeben sich aus einem veränderten Umgang mit fehlenden Werten (siehe Kapitel 2 in diesem Band).

B = Abweichungen zu Abbildung 10.2 sind in einer minimal veränderten Stichprobenzusammensetzung begründet.

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

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

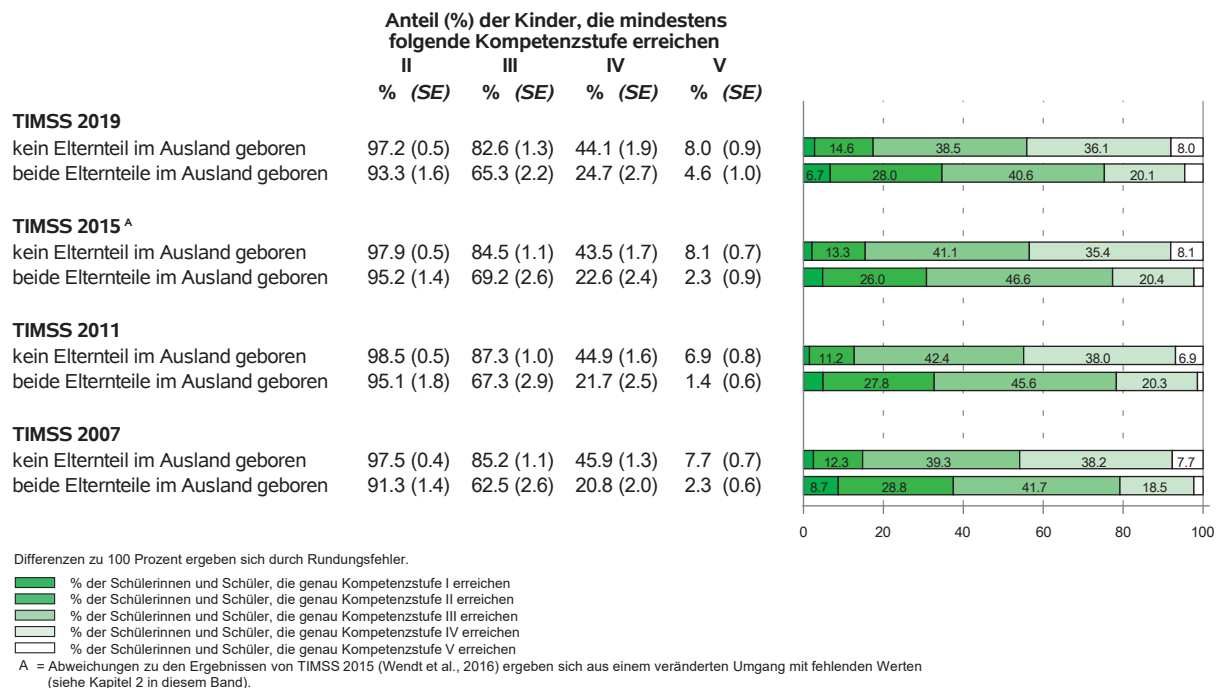
10.7 Leistungen auf unterschiedlichen Kompetenzstufen im Trend

In TIMSS werden für jede Kompetenzdomäne jeweils fünf Kompetenzstufen unterschieden, die in den Kapiteln 3 und 4 in diesem Band ausführlicher dargestellt sind. Mithilfe der Kompetenzstufen können die Leistungen der Schülerinnen und Schüler inhaltlich beschrieben werden. In Deutschland ist es ein bildungspolitisches Ziel, möglichst allen Schülerinnen und Schülern grundlegende Kompetenzen auf dem Niveau der Kompetenzstufe III zu vermitteln und gleichzeitig möglichst viele Schülerinnen und Schüler mit höheren Kompetenzen auszustatten. Abbildung 10.7 ist die prozentuale Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen der Mathematik im Trend zu entnehmen. Für alle Erhebungen zeigt sich, dass sich Kinder aus Familien ohne Migrationshintergrund systematisch häufiger auf den höheren mathematischen Kompetenzstufen verorten lassen und Kinder mit Migrationshintergrund häufiger Leistungen auf einem niedrigen beziehungsweise rudimentären Kompetenzniveau zeigen. An diesem grundlegenden Befund und den prozentualen Verteilungen auf die Kompetenzstufen zeigen sich zwischen 2007 und 2019 keine deutlichen Verschiebungen.

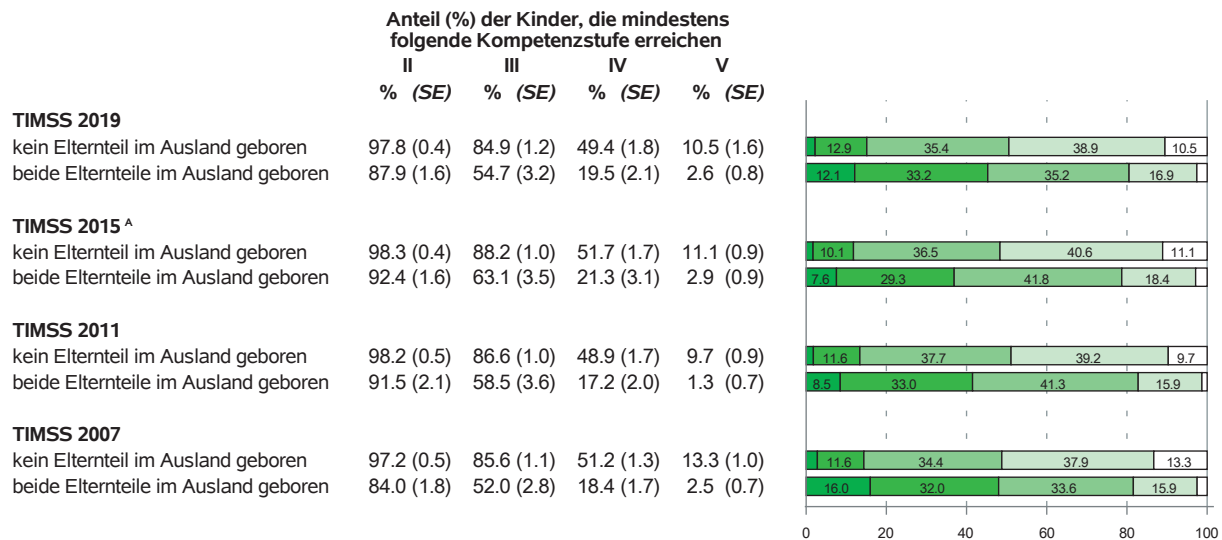
Während mit 17.4 Prozent in 2019 nur jede sechste Schülerin beziehungsweise jeder sechste Schüler ohne Migrationshintergrund niedrige Mathematikleistungen auf einem Niveau unter der Kompetenzstufe III zeigt, sind es bei Kindern, deren Eltern beide im Ausland geboren sind, 34.7 Prozent, also mehr als jedes drit-

te Kind dieser Gruppe. Auch wenn insgesamt die Leistungsunterschiede zwischen Kindern mit und ohne Migrationshintergrund in Mathematik signifikant seit 2007 gesunken sind (siehe Abbildung 10.4), so stellen diese Anteile für den zukünftigen Lernprozess in Mathematik weiterhin eine besondere Herausforderung dar, da diese Schülerinnen und Schüler mit hoher Wahrscheinlichkeit erhebliche Schwierigkeiten beim Lernen in der Sekundarstufe I haben werden (siehe Kapitel 3 in diesem Band). Besonders gute Leistungen auf dem Niveau der Kompetenzstufe V erzielen in 2019 lediglich 4,6 Prozent der Schülerinnen und Schüler, deren Eltern beide im Ausland geboren wurden, während es 8 Prozent der Schülerinnen und Schüler mit Eltern ohne Migrationshintergrund sind.

Abbildung 10.7: Prozentuale Verteilung der Schülerinnen und Schüler in Deutschland auf die fünf Kompetenzstufen (Mathematik) nach Migrationshintergrund der Eltern – 2007, 2011, 2015 und 2019 im Vergleich



Für die Naturwissenschaften (siehe Abbildung 10.8) zeigt sich die Ungleichverteilung von Leistungen der Kinder auf die unterschiedlichen Kompetenzstufen zwischen Kindern mit zwei im Ausland geborenen Elternteilen und Kindern mit keinem im Ausland geborenen Elternteil noch deutlicher. Während im Jahr 2019 Kinder aus Familien ohne Migrationshintergrund zu 10,5 Prozent die Kompetenzstufe V erreichen, ist dieser Anteil mit 2,6 Prozent signifikant geringer bei Kindern, deren Eltern beide im Ausland geboren wurden. Während am unteren Leistungsspektrum mit 45,3 Prozent fast jedes zweite Kind aus einer Familie mit Migrationshintergrund die Kompetenzstufe III (grundlegendes Alltagswissen) nicht erreicht, liegt dieser Anteil mit 15,1 Prozent bei den Schülerinnen und Schülern, deren Eltern beide in Deutschland geboren wurden, deutlich und signifikant darunter. Schülerinnen und Schüler, die nicht Kompetenzstufe III erreichen, vermögen es nur in Ansätzen, naturwissenschaftliches Wissen produktiv einzusetzen (siehe Kapitel 4 in diesem Band). Mit hoher Wahrscheinlichkeit haben diese Schülerinnen und Schüler erhebliche Schwierigkeiten beim naturwissenschaftlichen Lernen in der Sekundarstufe I.

Abbildung 10.8: Prozentuale Verteilung der Schülerinnen und Schüler in Deutschland auf die fünf Kompetenzstufen (Naturwissenschaften) nach Migrationshintergrund der Eltern – TIMSS 2007, 2011, 2015 und 2019 im Vergleich

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

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

^A = Abweichungen zu den Ergebnissen von TIMSS 2015 (Wendt et al., 2016) ergeben sich aus einem veränderten Umgang mit fehlenden Werten (siehe Kapitel 2 in diesem Band).

10.8 Zusammenhänge des Migrationshintergrundes mit Sprache und Merkmalen der sozialen Herkunft

Wie in Abschnitt 10.3 beschrieben, unterscheiden sich die Sprachkompetenzen sowie die sozioökonomischen Verhältnisse in Familien mit und ohne Migrationshintergrund bedeutsam. Vor dem Hintergrund von Studien, die die sprachlichen Kompetenzen und die sozioökonomische Herkunft als bedeutsame Faktoren zur Erklärung migrationsspezifischer Leistungsunterschiede belegen, wird im Folgenden untersucht, inwieweit sich auch in TIMSS 2019 Zusammenhänge mit dem Lernerfolg von Schülerinnen und Schülern in Mathematik und Naturwissenschaften feststellen lassen. Hierzu wurden verschiedene Regressionsanalysen gerechnet. Die Ergebnisse sind in Tabelle 10.2 dargestellt. Die Modelle 1 zeigen die bereits in den Abbildungen 10.3 und 10.5 berichteten Leistungsunterschiede zwischen Kindern mit und ohne Migrationshintergrund. Unter Berücksichtigung der in der Familie gesprochenen Sprache verringern sich die Effekte des Geburtslandes der Eltern sowohl für die mathematische als auch die naturwissenschaftliche Kompetenz (Modelle 2). Bei weiterer Berücksichtigung von Indikatoren der sozialen Herkunft (Modelle 3 bis 5) erweist sich der Migrationsstatus der Familie für die mathematische Kompetenz nicht mehr als signifikant bedeutsam für die Erklärung der Schülerinnen- und Schülerleistungen. Migrationsspezifische Leistungsdisparitäten in Mathematik sind folglich sowohl auf sprachliche Aspekte als auch auf sozioökonomische Benachteiligungen zurückzuführen (Modell 5). Für die Naturwissenschaften bleibt hingegen ein eigenständiger Erklärungsbeitrag des Migrationsstatus bestehen. Auch unter Kontrolle des familiären Sprachgebrauchs und der Indikatoren des ökonomischen und kulturellen Kapitals in den Familien erzielen

Schülerinnen und Schüler mit Migrationshintergrund im Vergleich zu ihren Mitschülerinnen und Mitschülern, deren Eltern beide in Deutschland geboren wurden, naturwissenschaftliche Leistungen, die mit 14.7 (ein Elternteil im Ausland geboren) beziehungsweise 28.8 (beide Elternteile) Leistungspunkten geringer ausfallen. Für beide Leistungsdomänen zeigt sich deutlich, dass auch Kinder – unter Kontrolle von Indikatoren der sozialen Herkunft und des Migrationsstatus –, die immer oder fast immer zu Hause Deutsch sprechen, mit 22.4 (Mathematik) beziehungsweise 27.5 (Naturwissenschaften) Leistungspunkten in beiden Kompetenzbereichen deutliche und signifikant höhere Leistungen zeigen als ihre Mitschülerinnen und Mitschüler, die nie oder nur manchmal zu Hause Deutsch sprechen. Zudem zeigen Kinder aus Familien mit hohen bildungsrelevanten sozioökonomischen Ressourcen erheblich bessere mathematische und naturwissenschaftliche Leistungen als Kinder, die in Familien mit geringem ökonomischem und kulturellem Kapital aufwachsen. Die beschriebenen Unterschiede zwischen den beiden Kompetenzdomänen sollten Gegenstand weiterer vertiefender Analysen sein.

Tabelle 10.2: Regressionsmodell zur Erklärung von Unterschieden in mathematischen und naturwissenschaftlichen Kompetenzen zwischen Kindern mit und ohne Migrationshintergrund für TIMSS 2019

	Mathematik									
	Modell 1		Modell 2		Modell 3		Modell 4		Modell 5	
	b	(SE)	b	(SE)	b	(SE)	b	(SE)	b	(SE)
Regressionskonstante	536.5	(2.5)**	537.3	(2.6)**	525.1	(2.9)**	523.3	(3.6)**	522.3	(3.6)**
Migrationshintergrund (ein Elternteil) ^A	-12.5	(4.7)**	-4.6	(4.8) ^{ns}	-3.4	(4.7) ^{ns}	-5.6	(5.1) ^{ns}	-4.5	(5.1) ^{ns}
Migrationshintergrund (beide Elternteile) ^B	-34.1	(4.4)**	-18.9	(4.9)**	-14.4	(4.9)**	-11.1	(6.3) ^{ns}	-7.3	(6.2) ^{ns}
Familiensprache ist nicht Deutsch ^C			-25.2	(5.0)**	-22.7	(4.9)**	-22.9	(5.9)**	-22.4	(6.1)**
Bedeutsamer Buchbesitz in der Familie ^D					34.7	(3.5)**	27.1	(4.2)**	23.7	(4.3)**
Familie mit hohem Bildungsniveau ^E							28.3	(3.9)**	17.0	(4.5)**
Familie mit hohem Berufsstatus ^F									18.2	(5.4)**
R ²	0.04		0.05		0.11		0.13		0.13	
	Naturwissenschaften									
	Modell 1		Modell 2		Modell 3		Modell 4		Modell 5	
	b	(SE)	b	(SE)	b	(SE)	b	(SE)	b	(SE)
Regressionskonstante	544.9	(2.2)**	545.8	(2.2)**	533.4	(2.6)**	531.6	(3.1)**	530.8	(3.0)**
Migrationshintergrund (ein Elternteil) ^A	-21.9	(4.5)**	-13.8	(4.6)**	-12.5	(4.5)**	-14.6	(5.7)*	-14.7	(5.6)**
Migrationshintergrund (beide Elternteile) ^B	-59.9	(4.2)**	-42.4	(5.0)**	-38.0	(5.1)**	-34.8	(7.3)**	-28.8	(7.8)**
Familiensprache ist nicht Deutsch ^C			-28.8	(5.5)**	-25.8	(5.5)**	-28.3	(7.0)**	-27.5	(7.6)**
Bedeutsamer Buchbesitz in der Familie ^D					34.2	(3.2)**	28.7	(3.7)**	24.7	(3.7)**
Familie mit hohem Bildungsniveau ^E							25.5	(3.7)**	14.6	(4.8)**
Familie mit hohem Berufsstatus ^F									18.4	(5.1)**
R ²	0.11		0.12		0.17		0.18		0.17	

b = Regressionsgewicht (unstandardisiert)

Signifikanzniveau: ns = nicht signifikant; * = signifikant ($p \leq .05$); ** = signifikant ($p \leq .01$)

A = Migrationshintergrund nach Geburtsland der Eltern (0 = kein Elternteil im Ausland geboren; 1 = ein Elternteil im Ausland geboren; 2 = beide Elternteile im Ausland geboren)

B = Migrationshintergrund nach Geburtsland der Eltern (0 = kein Elternteil im Ausland geboren; 1 = ein Elternteil im Ausland geboren; 2 = beide Elternteile im Ausland geboren)

C = Familiensprache nach Angabe der Schülerinnen und Schüler (0 = Deutsch: immer oder fast immer; 1 = Deutsch: manchmal oder nie)

D = Heimischer Buchbesitz nach Angabe der Eltern (0 = maximal 100 Bücher; 1 = mehr als 100 Bücher)

E = Bildungsniveau nach höchstem Bildungsabschluss der Eltern (0 = kein Elternteil mit mindestens Fachhochschulabschluss; 1 = mindestens ein Elternteil mit mindestens Fachhochschulabschluss)

F = Berufsstatus: Höchster ISEI (International Socio-Economic Index of Occupational Status) im Haushalt (0 = Werte unter 65 Punkte; 1 = Werte über 65 Punkte [z.B. Akademiker])

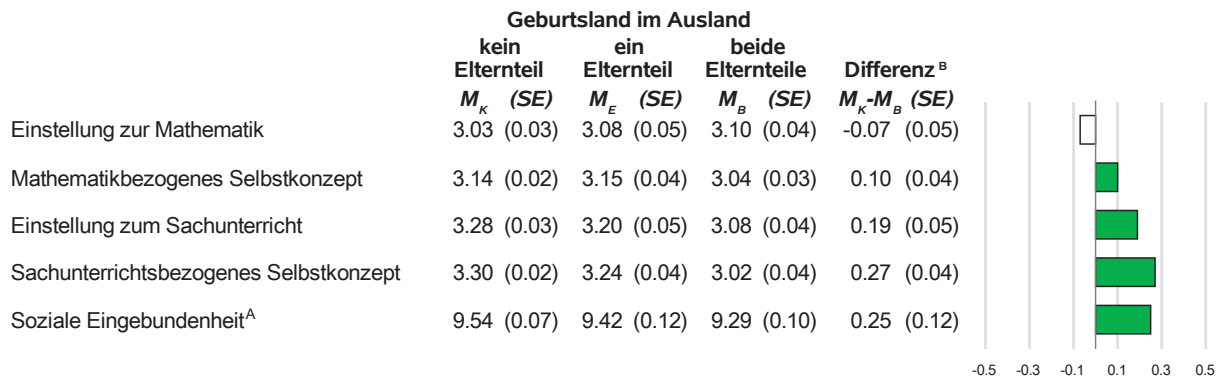
10.9 Fachbezogene Selbstkonzepte, Motivation und soziale Eingebundenheit

Neben den in der TIMSS-Konzeption beschriebenen Bereichen mathematischer und naturwissenschaftlicher Kompetenzen stellen auch positive Einstellungen gegenüber dem Lernen im Fach und ein positives Selbstkonzept der eigenen Fähigkeiten wichtige Zielbereiche des Unterrichts dar (siehe Kapitel 3 und 4 in diesem Band). Schulische Integration bedeutet, dass Schülerinnen und Schüler nicht nur am Lernen Freude haben, sondern sich auch angenommen fühlen. In Abbildung 10.9 sind Unterschiede in mathematik- und naturwissenschaftsbezogenen Selbstkonzepten sowie Einstellungen zu Mathematik und Naturwissenschaften nach Migrationsstatus der Schülerinnen- und Schülerfamilien sowie der sozialen Eingebundenheit dargestellt. Im linken Teil der Abbildung sind in tabellarischer Form die Mittelwerte sowie die zugehörigen Standardfehler zu den fünf Konstrukten getrennt nach Migrationsstatus dargestellt. Der Differenzwert für Kinder, deren Eltern beide in Deutschland beziehungsweise beide im Ausland geboren sind, wird zudem im rechten Teil der Abbildung grafisch illustriert. Statistisch signifikante Unterschiede sind grün gefärbt.

In Bezug auf die betrachteten motivationalen Maße ist festzuhalten, dass ein Großteil der Viertklässlerinnen und Viertklässler in Deutschland unabhängig von ihrem Migrationsstatus in allen vier betrachteten Indikatoren (mathematik- und naturwissenschaftsbezogene Selbstkonzepte, Einstellungen zu Mathematik und Naturwissenschaften) hohe positive Selbsteinschätzungen zum Ausdruck bringt (siehe Kapitel 3 und 4 in diesem Band). Wie Abbildung 10.9 verdeutlicht, haben Kinder aus Familien ohne Migrationshintergrund positivere Werte in drei von vier motivationalen Merkmalen. Sie berichten, dass sie über eine höhere Motivation im Sachunterricht verfügen (0.19) und selbstbewusster hinsichtlich ihrer mathematik- und naturwissenschaftsbezogenen Selbstkonzepte (0.10 bzw. 0.27) als ihre Mitschülerinnen und Mitschüler sind, deren Eltern beide im Ausland geboren sind. Die Unterschiede zwischen den Gruppen sind zwar statistisch signifikant, dennoch handelt es sich um sehr kleine Effekte. Festzustellen ist, dass die Effekte für den Sachunterricht größer ausfallen als für Mathematik. Keine signifikanten Unterschiede lässt sich für die Einstellung zum Mathematikunterricht beobachten.

In Bezug auf die soziale Eingebundenheit zeigt sich, dass sich die meisten Viertklässlerinnen und Viertklässler in Deutschland in der Schule sozial eingebunden fühlen. 89 Prozent aller Schülerinnen und Schüler bringen hier hohe bis sehr hohe Zustimmungswerte zum Ausdruck (keine grafische Darstellung). Abbildung 10.9 verdeutlicht jedoch, dass Kinder aus Familien mit Migrationshintergrund signifikant niedrigere Zustimmungswerte zum Ausdruck bringen als ihre Mitschülerinnen und Mitschüler, deren Eltern beide in Deutschland geboren sind. Insgesamt verdeutlichen die Ergebnisse, dass Ansätze der individuellen Förderung von Kindern mit Migrationshintergrund nicht nur sprachliche und strukturelle Voraussetzungen für fachliches Lernen berücksichtigten sollten, sondern auch motivationale sowie Aspekte der sozialen Eingebundenheit in der Schule stärker in den Blick zu nehmen sind.

Abbildung 10.9: Mittelwerte und Differenzen von mathematik- und naturwissenschaftsbezogenen Selbstkonzepten, Einstellungen zu Mathematik und Naturwissenschaften sowie sozialer Eingebundenheit nach Migrationshintergrund



Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

□ Unterschied nicht statistisch signifikant ($p > .05$)

■ Unterschied statistisch signifikant ($p \leq .05$)

A = Die Skala umfasst fünf Fragen (z.B. An dieser Schule habe ich das Gefühl, dazuzugehören.) mit vierstufigem Antwortformat (1 = Stimme stark zu, ..., 4 = Stimme überhaupt nicht zu).

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

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

10.10 Zusammenfassung

Die Befunde aus TIMSS 2019 bestätigen, dass Schülerinnen und Schüler mit Migrationshintergrund in vielen Teilnehmerstaaten und -regionen niedrigere mathematische und naturwissenschaftliche Leistungen zeigen als ihre Mitschülerinnen und Mitschüler, deren Eltern beide im jeweiligen Staat geboren wurden. Dies ist auch in Deutschland nicht anders. Hier liegt der Unterschied in den Mathematikleistungen bei 34 Punkten und in den naturwissenschaftlichen Leistungen sogar bei 60 Punkten und damit in der Größenordnung dessen, was Kinder im Laufe eines Jahres (Mathematik) beziehungsweise im Laufe von zwei Jahren (Naturwissenschaften) durchschnittlich dazulernen. Deutlich zeigt sich in international vergleichender Perspektive, dass die Leistungsunterschiede erheblich variieren. Einige Teilnehmerstaaten und -regionen, darunter auch einige Nachbarstaaten mit vergleichbaren Bevölkerungszusammensetzungen, weisen im Vergleich zu Deutschland deutlich niedrigere migrationsbezogene Disparitäten in ihrem Bildungssystem auf. Weitergehend zu prüfen wäre, ob dies durch Maßnahmen in den Bildungssystemen oder durch die Zusammensetzung der Migrantinnen- und Migrantengruppen in den jeweiligen Staaten zu erklären ist.

Die dargestellten Befunde erlauben es auch für 2019 nicht, ein positives Fazit zu ziehen. Auf Basis der TIMSS-Ergebnisse lassen sich keine Hinweise erkennen, dass es in Deutschland trotz vielfältiger Bemühungen bisher gelungen ist, dem bildungspolitischen Ziel – der systematischen Reduktion von zuwanderungsbezogenen Disparitäten – näher zu kommen. Die migrationsbezogenen Leistungsdisparitäten sind in Deutschland seit 2007 praktisch unverändert. Ergebnisse wie aus TIMSS 2015, die auf eine Reduktion der Leistungsdisparitäten hoffen ließen, lassen sich in 2019 nicht nachweisen. Die in Kapitel 12 in diesem Band dargestellten komplexen Trendanalysen geben keine Hinweise darauf, dass Veränderungen in der Schülerschaft den Erfolg von verbessertem Unterricht und inner- oder außerschulischer Fördermaßnahmen überschatten. Die in diesem Kapitel dargestellten Regressionsanalysen zeigen deutlich,

dass vor allem Sprachkompetenzen und sozioökonomische Benachteiligung wichtige Faktoren sind, um migrationspezifische Leistungsdisparitäten zu erklären. Die Förderung von Kindern entsprechend ihrer unterschiedlichen individuellen und dabei insbesondere sprachlichen sowie sozioökonomischen Voraussetzungen sollte noch stärker als bisher in den Fokus bildungspolitischer sowie praktischer Bemühungen rücken.

Gleichwohl verweisen die in diesem Kapitel vorgenommenen Analysen zu motivationalen Merkmalen und der sozialen Eingebundenheit darauf, dass sich Kinder mit Migrationshintergrund an Schulen nicht im gleichen Maße eingebunden fühlen wie ihre Mitschülerinnen und Mitschüler ohne Migrationshintergrund. Auch sind sie, was ihre fachlichen Leistungen angeht, weniger selbstbewusst und haben auch weniger Freude am Lernen naturwissenschaftlicher Inhalte. Um dem Ziel der Reduktion von migrationsbezogenen Disparitäten in Deutschland näher zu kommen, sollten zukünftig verstärkt auch Merkmale der Institution Schule in den Blick rücken. Hier ist kritisch zu fragen, wie die Qualität schulischer Sozialbeziehungen und des Fachunterrichts verbessert werden kann, damit sich alle Kinder unabhängig von ihren sprachlichen und soziokulturellen Voraussetzungen gleich eingebunden fühlen und selbstbewusste und motivierte Lernende sind.

Literatur

- Autorengruppe Bildungsberichterstattung. (2016). *Bildung in Deutschland 2016. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Bielefeld: Bertelsmann. Verfügbar unter: <https://www.bildungsbericht.de/de/bildungsberichte-seit-2006/bildungsbericht-2016/pdf-bildungsbericht-2016/bildungsbericht-2016>
- Bonsen, M., Kummer, N. & Bos, W. (2008). Schülerinnen und Schüler mit Migrationshintergrund. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 157–175). Münster: Waxmann.
- Esser, H. (2006). *Sprache und Integration. Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt a. M.: Campus.
- Faber, G. (2013). Klassenzusammensetzung als Kontextfaktor für die Leistungs- und Selbstkonzeptentwicklung. In R. Becker und A. Schulze (Hrsg.) *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen* (S. 325–352). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-18985-7_12
- Faber, G., Tiedemann, J. & Billmann-Mahecha, E. (2011). Selbstkonzept und Lernfreude in der Grundschulmathematik: Die Bedeutung von Migration und Geschlecht. Längsschnittliche Ergebnisse aus der Hannoverschen Grundschulstudie. *Heilpädagogische Forschung*, 37, 127–143.
- Gisdakis, B. (2007). Oh wie wohl ist mir in der Schule... Schulisches Wohlbefinden – Veränderungen und Einflussfaktoren im Laufe der Grundschulzeit. In C. Alt (Hrsg.), *Kinderleben – Start in die Grundschule. Band 3: Ergebnisse aus der zweiten Welle* (S. 81–136). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gogolin, I. (1994). *Der monolinguale Habitus der multilingualen Schule*. Münster: Waxmann.
- Gogolin, I., Neumann, U. & Roth, H.-J. (2003). *Förderung von Kindern und Jugendlichen mit Migrationshintergrund* (Materialien zur Bildungsplanung und Forschungsförderung, Bd. 107). Berlin: BLK.
- Gomolla, M. & Radtke, F.-O. (2007). *Institutionelle Diskriminierung. Die Herstellung ethnischer Differenz in der Schule* (2. Aufl.) Wiesbaden: VS Verlag für Sozialwissenschaften.
- Haag, N., Böhme, K. & Stanat, P. (2012). Zuwanderungsbezogene Disparitäten. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen*

- und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik. *Ergebnisse des IQB-Ländervergleichs 2011* (S. 209–235). Münster: Waxmann.
- Harazd, B. & Schürer, S. (2006). Veränderungen der Schulfreude von der Grundschule zur weiterführenden Schule. In A. Schröder-Lenzen (Hrsg.), *Risikofaktoren kindlicher Entwicklung. Migration, Leistungsangst und Schulübergang* (S. 208–222). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90075-9_11
- Hasse, R. & Schmidt, L. (2012). Institutionelle Diskriminierung. In U. Bauer, U. H. Bittlingmayer & A. Scherr (Hrsg.), *Handbuch Bildungs- und Erziehungssoziologie* (S. 883–899). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-18944-4_52
- Heinze, A., Herwartz-Emden, L. & Reiss, K. (2007). Mathematikkenntnisse und sprachliche Kompetenz bei Kindern mit Migrationshintergrund zu Beginn der Grundschulzeit. *Zeitschrift für Pädagogik*, 53(4) 562–581
- Hofmann, F., Bonitz, M., Lippert, N. & Gläser-Zikuda, M. (2018). Wohlbefinden von Grundschülerinnen und Grundschülern. Zur Bedeutung individueller, sozialer und schulischer Faktoren. In G. Hagenauer & T. Hascher (Hrsg.), *Emotionen und Emotionsregulation in Schule und Hochschule* (S. 121–135) Münster: Waxmann.
- Keeley, B. (2009). Internationale Migration. Die menschliche Seite der Globalisierung. Paris: OECD Publishing. <https://doi.org/10.1787/9789264075740-de>
- Kempert, S., Edele, A., Rauch, D., Wolf, K. M., Paetsch, J., Darsow, A., Maluch, J. & Stanat, P. (2016). Die Rolle der Sprache für zugewandungsbezogene Ungleichheiten im Bildungserfolg. In C. Diehl, C. Hunkler & C. Kristen (Hrsg.), *Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten* (S. 157–241). Wiesbaden: Springer VS.
- Kigel, R. M., McElvany, N. & Becker, M. (2015). Effects of immigrant background on text comprehension, vocabulary, and reading motivation: A longitudinal study. *Learning and Instruction*, 35, 73–84. <https://doi.org/10.1016/j.learninstruc.2014.10.001>
- KMK (Ständige Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland). (2002). *Bericht ‚Zuwanderung‘. Beschluss der Kultusministerkonferenz vom 24.05.2002 i. d. F. vom 16.11.2006*. Verfügbar unter: https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2002/2002_05_24-Zuwanderung.pdf
- KMK (Ständige Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland). (2004). *Bericht ‚Standards für die Lehrerbildung: Bildungswissenschaften‘. Beschluss der Kultusministerkonferenz vom 16.12.2004*. Verfügbar unter: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- KMK (Ständige Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland). (2015). *Darstellung von kultureller Vielfalt, Integration und Migration in Bildungsmedien – Gemeinsame Erklärung der Kultusministerkonferenz, der Organisationen von Menschen mit Migrationshintergrund und der Bildungsmedienverlage. Beschluss der Kultusministerkonferenz vom 08.10.2015*. Verfügbar unter: https://www.globales-lernen.de/sites/default/files/files/link-elements/kmk-top-003-sg-migrantenverbaende_a2-erkl.pdf
- Kohl, K., Striegler, K., Peters, K. & Leyendecker, B. (2011). Positive Schuleinstellung, Lernfreude und respektvolle Schüler-Lehrer-Beziehung – die Situation von Kindern aus zugewanderten Familien in der Grundschule. In A. Ittel, H. Merckens & L. Stecher (Hrsg.), *Jahrbuch Jugendforschung* (S. 46–73). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-93116-6_2
- Kristen, C. & Olczyk, M. (2013). Ethnische Einbettung und Bildungserfolg. In R. Becker & A. Schulze (Hrsg.), *Bildungskontexte. Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen* (S. 353–403). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-18985-7_13
- Kunz, R. (2008). *Die schulische Versorgung zugewanderter Kinder und Jugendlicher in Deutschland. Organisation, Förderung und psycho-soziale Betreuung*. Hamburg: Dr. Kovač.
- Lehmann, R. H., Peek, R. & Gänsfuß, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Bericht über die Untersuchung im September 1996*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.

- Lossen, K., Tillmann, K., Holtappels, H. G., Rollett, W. & Hannemann, J. (2016). Entwicklung der naturwissenschaftlichen Kompetenzen und des sachunterrichtsbezogenen Selbstkonzepts bei Schüler/-innen in Ganztagsgrundschulen. Ergebnisse der Längsschnittstudie StEG-P zu Effekten der Schülerteilnahme und der Angebotsqualität. *Zeitschrift für Pädagogik*, 62(6), 760–779.
- McElvany, N., Jungermann, A., Bos, W. & Holtappels, H. G. (2017). *Ankommen in der Schule. Chancen und Herausforderungen der Integration von Kindern und Jugendlichen mit Fluchterfahrung*. Münster: Waxmann.
- Mehringer, V. (2013). *Weichenstellungen in der Grundschule. Sozial-Integration von Kindern mit Migrationshintergrund* (Interkulturelle Bildungsforschung, Bd. 22). Münster: Waxmann.
- Mielke, R., Goy, M. & Pietsch, M. (2006). Das Leseselbstkonzept am Ende der Grundschulzeit. In W. Bos & M. Pietsch (Hrsg.), *KESS 4 – Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 4 in Hamburger Grundschulen* (S. 87–109). Münster: Waxmann.
- Miyamoto, A., Pfof, M. & Artelt, C. (2017). Reciprocal relations between intrinsic reading motivation and reading competence: A comparison between native and immigrant students in Germany. *Journal of Research in Reading*, 41(1), 176–196. <https://doi.org/10.1111/1467-9817.12113>
- Nicklaussen J. (2012). Das Wohlbefinden von Grundschulkindern. In F. Hellmich, F. Hoya & S. Förster (Hrsg.), *Bedingungen des Lehrens und Lernens in der Grundschule* (Jahrbuch Grundschulforschung, Bd. 16). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-19137-9_11
- Prediger, S. (2013). Sprachmittel für mathematische Verstehensprozesse – Einblicke in Probleme, Vorgehensweisen und Ergebnisse von Entwicklungsforschungsstudien. In A. Pallack (Hrsg.), *Impulse für eine zeitgemäße Mathematiklehrer-Ausbildung. MNU-Dokumentation der 16. Fachleitertagung Mathematik*. (S. 26–36). Neuss: Seeberger.
- Rauer, W. & Schuck, K. D. (2003). *Fragebogen zur Erfassung emotionaler und sozialer Schulerfahrungen von Grundschulkindern dritter und vierter Klassen. FEESS 3–4*. Göttingen: Beltz.
- Rjosk, C., Haag, N., Heppt, B. & Stanat, P. (2017). Zuwanderungsbezogene Disparitäten. In P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich & N. Haag (Hrsg.), *IQB Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich* (S. 237–275). Münster: Waxmann.
- Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., Klieme, E. & Stanat, P. (2014). Socio-economic and language minority classroom composition and individual reading achievement: The mediating role of instructional quality. *Learning and Instruction*, 32, 63–72. <https://doi.org/10.1016/j.learninstruc.2014.01.007>
- Roebers, C., Mecheril, A. & Schneider, W. (1998). Migrantenkindern in deutschen Schulen. Eine Studie zur Persönlichkeitsentwicklung. *Zeitschrift für Pädagogik*, 44(5), 723–736.
- Scherr, A. & Niermann, D. (2012). Migration und Kultur im schulischen Kontext. In U. Bauer, U. H. Bittlingmayer & A. Scherr (Hrsg.), *Handbuch Bildungs- und Erziehungssoziologie* (S. 863–882). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-531-18944-4_51
- Schurig, M., Wendt, H., Kasper, D. & Bos, W. (2015). Fachspezifische Stärken und Schwächen von Viertklässlerinnen und Viertklässlern in Deutschland im europäischen Vergleich. In H. Wendt, T. C. Stubbe, K. Schwippert & W. Bos (Hrsg.), *IGLU & TIMSS: 10 Jahre internationale vergleichende Schulleistungsforschung in der Grundschule. Vertiefende Analysen zu IGLU und TIMSS 2001 bis 2011* (S. 35–54). Münster: Waxmann.
- Schwippert, K., Bos, W. & Lankes, E.-M. (2003). Heterogenität und Chancengleichheit am Ende der vierten Jahrgangsstufe im internationalen Vergleich. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 265–302). Münster: Waxmann.
- Schwippert, K., Hornberg, S. & Goy, M. (2008). Lesekompetenzen von Kindern mit Migrationshintergrund im nationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU-E 2006. Die*

- Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (S. 111–125). Münster: Waxmann.
- Schwippert, K., Wendt, H. & Tarelli, I. (2012). Lesekompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 191–208). Münster: Waxmann.
- Stanat, P. (2008). Heranwachsende mit Migrationshintergrund im deutschen Bildungswesen. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland* (S. 683–745). Reinbek: Rowohlt.
- Stanat, P. & Christensen, G. (2006). *Schulerfolg von Jugendlichen mit Migrationshintergrund im internationalen Vergleich. Eine Analyse von Voraussetzungen und Erträgen schulischen Lernens im Rahmen von PISA 2003* (Bildungsforschung, Bd. 19). Verfügbar unter: <http://docplayer.org/docview/34/13104811/„l„file=/storage/34/13104811/13104811.pdf>
- Stanat, P., Schwippert, K. & Gröhlich, C. (2010). Der Einfluss des Migrantenanteils in Schulklassen auf den Kompetenzerwerb: Längsschnittliche Überprüfung eines umstrittenen Effekts. In C. Allemann-Ghionda, P. Stanat, K. Göbel & C. Röhner (Hrsg.), *Migration, Identität, Sprache und Bildungserfolg* (Zeitschrift für Pädagogik, 55. Beiheft, S. 147–164). Weinheim: Beltz.
- Tarelli, I., Schwippert, K. & Stubbe, T. C. (2012). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In W. Bos, H. Wendt, O. Köller & C. Selter (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 247–267). Münster: Waxmann.
- van Ophuysen, S. (2008). Zur Veränderung der Schulfreude von Klasse 4 bis 7. Eine Längsschnittanalyse schulformspezifischer Effekte von Ferien und Grundschulübergang. *Zeitschrift für Pädagogische Psychologie*, 22(3/4), 293–306.
<https://doi.org/10.1024/1010-0652.22.34.293>
- Vennemann, M. (2018). *Individual- und Kompositionseffekte und der Kompetenzzuwachs in Mathematik und Naturwissenschaft am Ende der Grundschule*. Wiesbaden: Springer Spektrum. <https://doi.org/10.1007/978-3-658-23506-2>
- Walsen, J. C. (2013). *Das Wohlbefinden von Grundschulkindern. Soziale und emotionale Schulerfahrungen in der Primarstufe*. Verfügbar unter: <http://oops.uni-oldenburg.de/1929/1/walwoh13.pdf>
- Wendt, H., Kasper, D., Bos, W., Vennemann, M. & Goy, M. (2017). Wie viele Punkte auf der TIMSS-Metrik entsprechen einem Lernjahr? Leistungszuwächse in Mathematik und Naturwissenschaften am Ende der Grundschulzeit. In T. Eckert & B. Gniewosz (Hrsg.), *Bildungsgerechtigkeit* (S. 121–153). Wiesbaden: Springer VS.
- Wendt, H. & Schwippert, K. (2017). Lesekompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 219–234). Münster: Waxmann.
- Wendt, H., Schwippert, K. & Stubbe, T. C. (2016). Mathematische und naturwissenschaftliche Kompetenzen von Schülerinnen und Schülern mit Migrationshintergrund. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 317–331). Münster: Waxmann.

Kapitel 11

Schullaufbahnpräferenzen am Übergang in die Sekundarstufe und der Zusammenhang mit leistungsrelevanten und sozialen Merkmalen

Tobias C. Stubbe, Daniel Kasper und Armin Jentsch

11.1 Einleitung

Der Übergang von der Grundschule auf eine Schulform der Sekundarstufe I hat für den zukünftigen Bildungserfolg von Schülerinnen und Schülern in Deutschland eine zentrale Bedeutung (Bellenberg & Klemm, 2000). Diese im internationalen Vergleich frühe Verteilung von Kindern auf unterschiedliche Bildungsgänge wird von vielen Expertinnen und Experten kritisiert. Zum einen ist eine verlässliche Vorhersage der zukünftigen Leistungsentwicklung zu diesem frühen Zeitpunkt stark fehlerbehaftet (Schuchart & Weishaupt, 2004). Zum anderen weisen sowohl die Empfehlungen von Grundschulen als auch die Entscheidungen von Erziehungsberechtigten deutliche Zusammenhänge mit dem sozialen Hintergrund und einer möglichen Migrationsgeschichte der Familie auf (im Überblick Maaz & Nagy, 2009).

Von Befürworterinnen und Befürworthern dieser frühen Trennung wird als Legitimierung häufig die Möglichkeit der nachträglichen Korrektur der ursprünglichen Schullaufbahnentscheidung im Verlauf der Sekundarstufe angeführt. Schulformwechsel werden dabei als horizontale Durchlässigkeit bezeichnet. Die Möglichkeit, einen höherwertigen Schulabschluss auf einer Schulart zu erwerben, die herkömmlich nicht zu diesem Abschluss führt (z.B. Real-schulabschluss an einer Hauptschule), wird vertikale Durchlässigkeit genannt. Empirische Daten zeigen, dass Schulformwechsel vor allem in Form von Abstiegen vorkommen und dass diese darüber hinaus – wie auch die ursprünglichen Schullaufbahnentscheidungen – vom sozialen Hintergrund der Erziehungsberechtigten abhängen (im Überblick Stubbe, 2009). Die Möglichkeiten der vertikalen Durchlässigkeit werden im deutschen Sekundarschulsystem in den vergangenen Jahren zwar verstärkt genutzt, es muss aber bezweifelt werden, ob Abschlüsse, die an Schulen erworben werden, die eigentlich nicht zu diesem

Abschluss führen, auf dem Arbeitsmarkt denselben Wert besitzen wie auf der dafür vorgesehenen Schulform erworbene Abschlüsse (Schuchart, 2007, 2013).

Traditionell werden in der Bundesrepublik Deutschland in der Sekundarstufe I drei Schulformen unterschieden: Hauptschule, Realschule und Gymnasium. In den 1970er-Jahren wurden in einigen Ländern zudem Gesamtschulen eingeführt, die diese drei Schulformen in einer Institution integrieren (Köller, 2008). Das klassische dreigliedrige Schulsystem erfuhr in den vergangenen Jahren in allen Ländern zahlreiche Reformen, durch die bestehende Schulformen abgeschafft und neue Schulformen eingeführt wurden. Dabei hat das föderale System der Bundesrepublik zu einer sehr uneinheitlichen Struktur der Schullandschaft geführt.

Das Gymnasium ist mittlerweile die einzige Schulform, die in allen Ländern der Bundesrepublik Deutschland existiert. Seit etwa 15 Jahren besuchen relativ konstant 35 Prozent der Sekundarschülerinnen und -schüler ein Gymnasium, womit diese Schulform hinsichtlich der Besuchsquoten die erfolgreichste ist (Rabenstein, Horn, Keßler & Stubbe, 2020, Abbildung 1.1). In Ländern mit einem zweigliedrigen Schulsystem erweist sich jeweils die zweite Schulform neben dem Gymnasium als diejenige mit den meisten Schülerinnen und Schülern (Mahler, Schipolowski & Weirich, 2019, Tabelle 3.3).

Tabelle 11.1 gibt einen Überblick über die allgemeinbildenden Schulformen der Sekundarstufe (ohne Förderschulen) in den einzelnen Ländern für das Schuljahr der Erhebung der *Trends in International Mathematics and Science Study* (TIMSS) 2019 (Schuljahr 2018/19). Im linken Teil der Tabelle finden sich die drei ‚klassischen‘ Schulformen, wobei auf den ersten Blick zu erkennen ist, dass Haupt- und Realschule in der Mehrzahl der Länder nicht mehr vorhanden sind und in den übrigen Ländern – abgesehen von Bayern – von Schulen mit mehreren Bildungsgängen ergänzt werden. Diese neuen Schulformen – sowie die Gesamtschule – finden sich im rechten Teil der Tabelle. Unterschieden wird dabei zusätzlich zwischen Schulen ohne beziehungsweise mit Gymnasialbildung und bei Letzteren zusätzlich zwischen Schulen, die keine eigene gymnasiale Oberstufe haben und solchen, bei denen das der Fall ist.

Eine konsequente Zweigliedrigkeit findet sich aktuell in den Stadtstaaten Berlin und Hamburg, im Saarland sowie in Sachsen und Schleswig-Holstein (seit dem Schuljahr 2019/20). Auch Bremen setzt nach der Grundschule zunächst auf ein zweigliedriges System, das ab der neunten Jahrgangsstufe aber durch eine Schulform für lernbenachteiligte Schülerinnen und Schüler ergänzt wird. Die Benennung von Schulformen verläuft in den Ländern weitgehend autonom und folgt keiner bundesweit einheitlichen Regelung. So finden sich für neue Schulformen ohne Gymnasialbildung in den neun Ländern, in denen diese Schulform existiert, sieben unterschiedliche Bezeichnungen.¹

Auch die gesetzlichen Vorgaben zum Übergang von der Primar- in die Sekundarstufe unterscheiden sich zwischen den Ländern zum Teil erheblich (KMK, 2015). Während teilweise von der Grundschule eine verbindliche Empfehlung für eine weiterführende Schule ausgesprochen wird (z.B. Bayern), werden von den Grundschulen anderer Länder zwar Empfehlungen gegeben, die endgültige Entscheidung liegt jedoch bei den Erziehungsberechtigten (z.B. Nordrhein-Westfalen) (siehe Tabelle 8.2 im Berichtsband zu IGLU 2016: Stubbe, Bos & Schurig, 2017).

1 Die Regionalschule in Schleswig-Holstein ist zum Schuljahr 2018/19 ausgelaufen und wird hier entsprechend nicht mehr mitgezählt.

Tabelle 11.1: Übersicht über die Schulformen der Sekundarstufe I in den Ländern der Bundesrepublik Deutschland (Schuljahr 2018/19)

Länder	Schulen mit einem Bildungsgang			Schulen mit mehreren Bildungsgängen		
	Hauptschule	Realschule	Gymnasium	ohne Gymnasialbildung	mit Gymnasialbildung	
					ohne gymn. Oberstufe	mit gymn. Oberstufe
Baden-Württemberg	Hauptschule Werkrealschule ^A	Realschule	Gymnasium	–	–	Gemeinschaftsschule*
Bayern	Mittelschule ^A	Realschule	Gymnasium	–	(Gesamtschule) ^B	–
Berlin	–	–	Gymnasium	–	–	Integrierte Sekundarschule* Gemeinschaftsschule*
Brandenburg	–	–	Gymnasium	Oberschule ^{CD}	–	Gesamtschule
Bremen	–	–	Gymnasium	Werksschule ^E	–	Oberschule
Hamburg	–	–	Gymnasium	–	–	Stadtteilschule
Hessen	Hauptschule	Realschule	Gymnasium	Mittelstufenschule	–	Gesamtschule*
Meckl.-Vorpommern	–	–	Gymnasium	Regionale Schule ^D	–	Gesamtschule*
Niedersachsen	Hauptschule	Realschule	Gymnasium	Oberschule	Oberschule	Gesamtschule*
Nordrhein-Westfalen	Hauptschule	Realschule	Gymnasium	–	Sekundarschule	Gesamtschule*
Rheinland-Pfalz	–	–	Gymnasium	Realschule plus ^{CD}	–	Gesamtschule*
Saarland	–	–	Gymnasium	–	–	Gemeinschaftsschule*
Sachsen	–	–	Gymnasium	Oberschule ^{CD}	–	–
Sachsen-Anhalt	–	–	Gymnasium	Sekundarschule	–	Gesamtschule Gemeinschaftsschule*
Schleswig-Holstein	–	–	Gymnasium	(Regionalschule) ^{DF}	–	Gemeinschaftsschule
Thüringen	–	–	Gymnasium	Regelschule ^D	–	Gesamtschule* Gemeinschaftsschule

* = Eine gymnasiale Oberstufe kann an der jeweiligen Schule oder als Kooperation mit anderen Schulen angeboten werden.

A = An dieser Schulform kann neben dem Hauptschulabschluss auch ein mittlerer Schulabschluss erworben werden, der jedoch nicht dem Realschulabschluss entspricht.

B = Insgesamt 5 Schulen, die als ‚Schulen besonderer Art‘ geführt werden.

C = An dieser Schulform ist der Erwerb der Fachoberschulreife möglich.

D = Wechsel an ein Gymnasium/Übertritt in die gymnasiale Oberstufe bei entsprechendem Notendurchschnitt oder dem Besuch spezieller Kurse (Sachsen) explizit möglich.

E = Schulform für lernbenachteiligte Jugendliche (Jahrgangsstufe 9 bis 11).

F = Schulform ist zum Schuljahr 2018/19 ausgelaufen. Bestehende Regionalschulen werden zu Gemeinschaftsschulen umgebaut.

Die Angaben basieren auf verfügbaren Veröffentlichungen der Kultusministerien der Länder der Bundesrepublik Deutschland (Stand: Schuljahr 2018/19).

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Um diesen Unterschieden zu entsprechen, müssten Analysen zum Thema Grundschulübergang grundsätzlich auf Länderebene durchgeführt werden. Da dies mit den Daten aus TIMSS 2019 jedoch nicht möglich ist, werden in diesem Kapitel alle Befunde auf Bundesebene dargestellt.

Die Analysen in diesem Kapitel schließen – ebenso wie bereits bei TIMSS 2015 – an die Berichtslegung im Rahmen der bisherigen Zyklen der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) 2001, 2006, 2011 und 2016 an, wobei im Jahr 2011 durch die zeitgleiche Durchführung der beiden Studien

Zusammenhänge mit allen drei Kompetenzbereichen (Lesen, Mathematik und Naturwissenschaften) untersucht werden konnten (Arnold, Bos, Richert & Stubbe, 2007, 2010; Bos et al., 2004; Milek, Lüdtke, Trautwein, Maaz & Stubbe, 2009; Stubbe & Bos, 2008; Stubbe, Bos & Euen, 2012; Stubbe et al., 2017; Stubbe, Lorenz, Bos & Kasper, 2016).

Wichtig ist, an dieser Stelle darauf hinzuweisen, dass für alle Trendanalysen in diesem Kapitel die Werte für TIMSS 2015 neu berechnet wurden, da damals alle Analysen mit imputierten Daten durchgeführt worden sind (Wendt et al., 2016), während dies für die verschiedenen IGLU-Zyklen sowie für TIMSS 2019 nicht gemacht wurde (siehe Kapitel 2 in diesem Band). Somit sind die Ergebnisse aus den einzelnen Jahren in diesem Kapitel gut miteinander zu vergleichen. Bei der Interpretation muss aber bedacht werden, dass ohne Imputation beispielsweise der Anteil der Gymnasialpräferenz der Erziehungsberechtigten tendenziell überschätzt wird.

Im zweiten Abschnitt dieses Kapitels werden zunächst deskriptiv die Schullaufbahnpräferenzen von Lehrkräften und Erziehungsberechtigten dargestellt. Es folgt im dritten Abschnitt die Analyse des Zusammenhangs zwischen den Schullaufbahnpräferenzen der Lehrkräfte und den leistungsrelevanten Merkmalen von Schülerinnen und Schülern. Schließlich werden im vierten Abschnitt Ergebnisse zum Zusammenhang zwischen den Schullaufbahnpräferenzen von Lehrkräften und Erziehungsberechtigten sowie dem sozialen Hintergrund der Schülerinnen und Schüler präsentiert. Eine Zusammenfassung der Befunde erfolgt im abschließenden fünften Abschnitt.

11.2 Schullaufbahnpräferenzen von Lehrkräften und Erziehungsberechtigten: Verteilung und Ausmaß der Übereinstimmung

Lehrkräfte wurden bei TIMSS 2019 wie schon bei TIMSS 2015 und in Übereinstimmung mit dem IGLU-Vorgehen nach dem erwarteten Schulabschluss der einzelnen Viertklässlerinnen und Viertklässler gefragt. Dieses Vorgehen ist der Abfrage der ausgesprochenen Übergangsempfehlung vorzuziehen, da die Empfehlungspraktiken und deren Verbindlichkeit zwischen den Ländern erheblich variieren (siehe Abschnitt 11.1). Mit der Frage nach dem erwarteten Schulabschluss (im Folgenden als ‚Schullaufbahnpräferenzen der Lehrkräfte‘ bezeichnet) erhält man einen guten Hinweis auf die prognostizierte Entwicklung der einzelnen Schülerinnen und Schüler aus Sicht ihrer Grundschullehrkräfte.

Wie in den vergangenen Erhebungen zu IGLU und TIMSS wurden die ‚Schullaufbahnpräferenzen der Erziehungsberechtigten‘ über deren Angabe zu der Schulform erfasst, die ihre Kinder im kommenden Schuljahr voraussichtlich besuchen werden. Da die Erziehungsberechtigten in fast allen Ländern die Schulanmeldungen ihrer Kinder vornehmen, kann die im Anschluss an die Grundschule besuchte Schule als Indikator für deren Präferenz interpretiert werden.

In Tabelle 11.2 finden sich die relativen Häufigkeiten der Schullaufbahnpräferenzen der Lehrkräfte und der Erziehungsberechtigten für die vier Erhebungszeitpunkte von IGLU (2011 inkl. TIMSS) sowie für TIMSS 2015 und TIMSS 2019. Hinsichtlich der Angaben der Lehrkräfte fällt auf, dass die Werte seit 2011 relativ konstant ausfallen. Für jeweils rund 40 Prozent der Viertklässlerinnen und Viertklässler erwarten die befragten Lehrkräfte einen Realschulabschluss

Tabelle 11.2: Schullaufbahnpräferenzen von Lehrkräften und Erziehungsberechtigten von IGLU 2001 bis TIMSS 2019 (Angaben in Zeilenprozent)

Schullaufbahnpräferenz	Hauptschule	Realschule	Gymnasium	Schulen mit mehreren Bildungsgängen
Lehrkräfte				
2019 (TIMSS)	19.1	40.1	40.8	–
2016 (IGLU)	17.4	39.6	43.0	–
2015 (TIMSS)	20.4	39.3	40.3	–
2011 (IGLU/TIMSS)	21.6	36.6	41.8	–
2006 (IGLU)	24.8	35.5	39.7	–
2001 (IGLU)	29.3	35.7	34.9	–
Erziehungsberechtigte				
2019 (TIMSS)	6.8	25.2	46.9	21.0
2016 (IGLU)	6.1	21.3	47.6	25.0
2015 (TIMSS)	7.0	26.3	46.0	20.7
2011 (IGLU/TIMSS)	11.9	28.3	45.5	14.3
2006 (IGLU)	14.7	27.0	47.3	11.0
2001 (IGLU)	22.1	29.2	40.8	7.8

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

Abweichungen zu den im Rahmen von TIMSS 2015 berichteten Werten (Stubbe et al., 2016) ergeben sich daraus, dass dort für 2015 mit imputierten Daten gearbeitet wurde.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

beziehungsweise das Abitur als Schulabschluss. Für knapp 20 Prozent wird ein Hauptschulabschluss erwartet. Blickt man bis ins Jahr der ersten IGLU-Erhebung 2001 zurück, zeigt sich ein Rückgang der Hauptschulpräferenzen um 10.3 Prozentpunkte, während der Anteil der Gymnasialpräferenzen um 5.9 Prozentpunkte gestiegen ist.²

Bei den Erziehungsberechtigten sind die Werte seit 2015 relativ konstant: Etwas mehr als 45 Prozent präferieren das Gymnasium, jeweils 20 bis 25 Prozent eine Realschule beziehungsweise eine Schule mit mehreren Bildungsgängen und nur knapp über 5 Prozent eine Hauptschule. Innerhalb der vergangenen 18 Jahre hat es bei diesen Präferenzen aber deutlichere Veränderungen gegeben als bei den Lehrkräften. Grund dafür ist vor allem die zunehmende Verbreitung von Schulen mit mehreren Bildungsgängen, bei denen entsprechend ein Anstieg um das 2.7-Fache zu verzeichnen ist. Bei den Gymnasialpräferenzen hat es nach 2001 zwar zunächst einen Anstieg um 6.5 Prozentpunkte gegeben; seit 2006 gibt es bei diesem Wert aber kaum Veränderungen. Einen Rückgang bei den Präferenzen hat es vor allem hinsichtlich der Hauptschulen gegeben (um 69.1 % seit 2001), aber auch die Realschulpräferenzen sind etwas rückläufig. Dass es nach wie vor relativ viele Übergänge auf Realschulen gibt, liegt darin begründet, dass es diese Schulform zwar nur noch in wenigen, dafür aber in den bevölkerungsreichen Ländern gibt.

Die Kreuztabelle 11.3 zeigt für die drei Präferenzen der Lehrkräfte den jeweiligen Anteil der Schülerinnen und Schüler nach Präferenz der Erziehungsberechtigten. Vergleicht man diese Werte mit jenen aus dem Jahr 2011 (Tabelle 8.4 aus dem Berichtsband zu IGLU 2011: Stubbe et al., 2012), zeigen sich relativ große Unterschiede, die sich wiederum primär aus der steigenden Bedeutung der Schulen mit mehreren Bildungsgängen ergeben. Nur 36.2 Prozent der Viertklässlerinnen und Viertklässler, deren Lehrkräfte für sie einen Hauptschulabschluss erwarten, werden den Angaben der Erziehungsberechtigten

2 Inkonsistenzen zwischen den im Text und in den Abbildungen berichteten Werten sind im Rundungsverfahren begründet.

Tabelle 11.3: Übereinstimmung der Schullaufbahnpräferenzen von Lehrkräften und Erziehungsberechtigten (Angaben in Zeilenprozent)

		Schullaufbahnpräferenz der Erziehungsberechtigten				
		Hauptschule	Realschule	Gymnasium	Schule mit mehreren Bildungsgängen	n
Schullaufbahnpräferenz der Lehrkräfte	Hauptschule	36.2	27.7	4.4	31.7	267
	Realschule	2.7	46.6	20.4	30.3	710
	Gymnasium	0.2	7.9	83.4	8.6	928
	n	117	478	931	379	1905

Differenzen zu 100 Prozent ergeben sich durch Rundungsfehler.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

zufolge tatsächlich auf diese Schulform wechseln (2011: 54.5%). Etwas höher als vor acht Jahren sind die Prozentwerte für den Wechsel auf eine Realschule (2019: 27.7%; 2011: 22.1%) oder auf ein Gymnasium (2019: 4.4%; 2011: 1.2%). Deutlich höher fällt erwartungskonform der Wert für Schulen mit mehreren Bildungsgängen aus (2019: 31.7%; 2011: 22.2%).

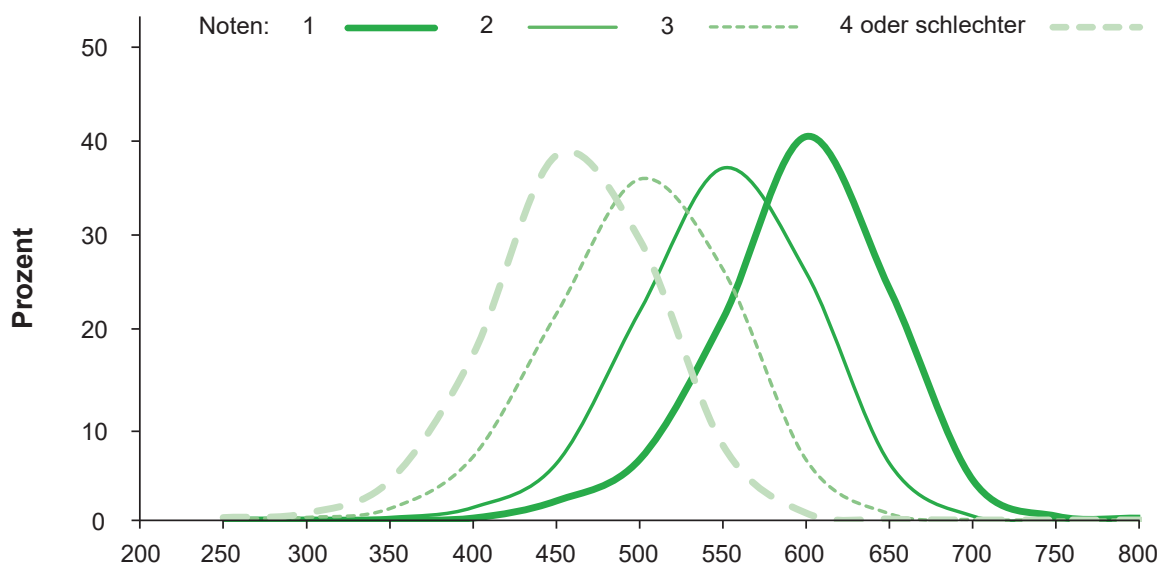
Auch bei den Schülerinnen und Schülern, die eine Lehrkräftepräferenz für die Realschule aufweisen, fällt der Anteil derer, die auf eine Realschule wechseln werden, mit 46.6 Prozent geringer aus als 2011 (59.5%). Besonders deutlich hat sich auch in dieser Gruppe der Wert für die Schulen mit mehreren Bildungsgängen verändert (2019: 30.3%; 2011: 18.8%).

Für Kinder, für die die Grundschullehrkräfte das Abitur erwarten, sind Schulen mit mehreren Bildungsgängen hingegen deutlich weniger bedeutsam. Der Anteil, der auf eine Schule mit mehreren Bildungsgängen wechselt, ist moderat um 1.7 Prozentpunkte auf 8.6 Prozent gestiegen. Mehr als 80 Prozent der Erziehungsberechtigten entscheiden sich für ein Gymnasium für ihr Kind, wenn die jeweilige Grundschullehrkraft erwartet, dass das Abitur erreicht wird.

11.3 Schullaufbahnpräferenzen von Lehrkräften: Leistungsrelevante Merkmale der Schülerinnen und Schüler

Wie bereits betont, unterscheiden sich die Regelungen zum Übergang von der Primar- in die Sekundarstufe zwischen den einzelnen Ländern der Bundesrepublik Deutschland. Gemein ist den jeweiligen Vorgaben aber, dass unterschiedliche leistungsrelevante Merkmale der Schülerinnen und Schüler die Schullaufbahnempfehlung bestimmen sollen (KMK, 2015). Das wichtigste Kriterium sind dabei die Noten (insbesondere in den Hauptfächern), ergänzt durch weitere Aspekte wie Lernentwicklung sowie Arbeits- und Sozialverhalten.

Da aus unterschiedlichen Studien bekannt ist, dass die Korrelation zwischen Schulnoten und standardisierten Leistungstests relativ gering ausfällt (im Überblick Lintorf, 2012), zeigt Abbildung 11.1 zunächst die Mathematikkompetenz der Viertklässlerinnen und Viertklässler differenziert nach den Mathematiknoten, bevor der Zusammenhang zwischen ausgewählten leistungsrelevanten Merkmalen und der Schullaufbahnpräferenz der Lehrkräfte betrachtet wird.

Abbildung 11.1: Testleistungen der Schülerinnen und Schüler differenziert nach Mathematiknoten – Gesamtskala Mathematik

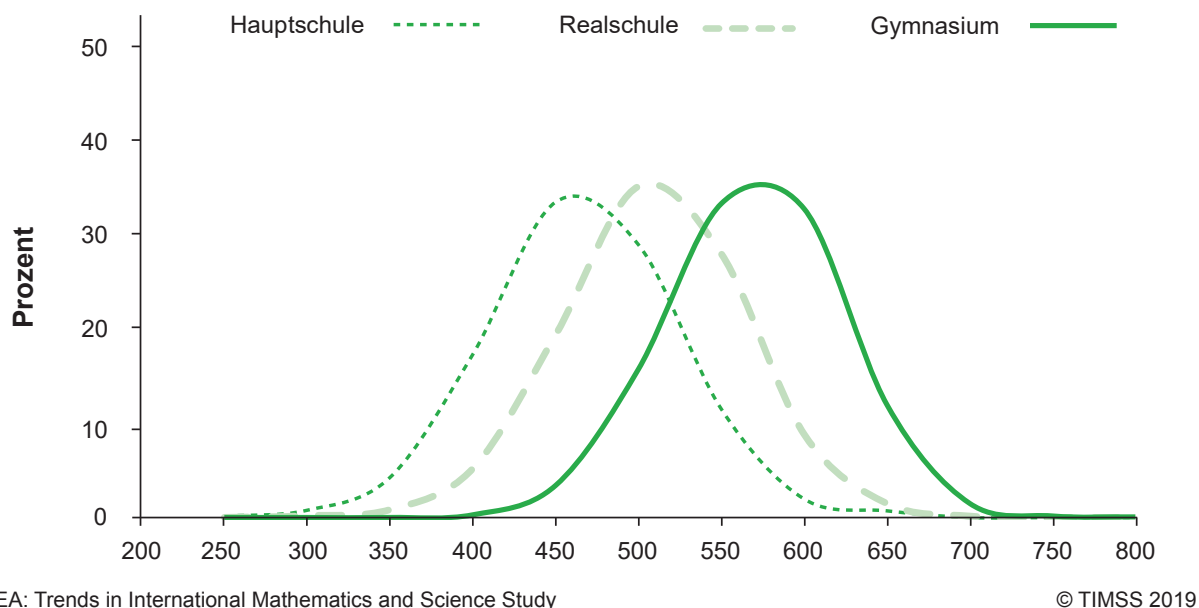
IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Zunächst wird erwartungskonform deutlich, dass sich die mittleren Mathematikkompetenzen je nach erhaltener Mathematiknote substanziell unterscheiden. Zugleich zeigen die Überlappungen der vier Kurven aber auch, dass Schülerinnen und Schüler mit mittleren Testleistungen (circa 500–550 Punkte) je nach Lehrkraft alle Noten von ausreichend (oder schlechter) bis sehr gut erhalten können. Noten hängen somit nicht nur von den Leistungen der Schülerinnen und Schüler ab, sondern auch von den Bewertungsmaßstäben der jeweiligen Lehrkräfte. In den folgenden Analysen muss also immer bedacht werden, dass sich Noten im Gegensatz zu Tests nur bedingt für den Leistungsvergleich zwischen unterschiedlichen Klassen eignen.

In Abbildung 11.2 wird die Mathematikkompetenz der Viertklässlerinnen und Viertklässler differenziert nach der Schullaufbahnpräferenz der Lehrkräfte dargestellt. Auch in dieser Abbildung erkennt man neben den offensichtlichen Mittelwertdifferenzen die deutlichen Überlappungen der drei Kurven, was darauf hinweist, dass die Lehrkräfte von Schülerinnen und Schülern mit mittleren Testleistungen – insbesondere von denjenigen mit Kompetenzstufe III (475–550 Punkte) – Präferenzen für alle drei Schulformen aufweisen.

Tabelle 11.4 zeigt den Zusammenhang der Schullaufbahnpräferenz der Lehrkräfte mit den leistungsrelevanten Merkmalen der Viertklässlerinnen und Viertklässler, die im Rahmen von TIMSS 2019 erhoben wurden: die mathematische und die naturwissenschaftliche Kompetenz, die Anstrengungsbereitschaft und die Leistungsangst (erhoben über den Elternfragebogen), die Noten in den Fächern Deutsch, Mathematik und Sachunterricht (erhoben über die Schülerteilnahmeliste) sowie das Selbstkonzept in Mathematik und im Sachunterricht (erhoben über den Schülerfragebogen). Berichtet werden standardisierte Regressionskoeffizienten (β) mit der dichotomen Variable *Gymnasialpräferenz der Lehrkräfte* als abhängige Variable. Bei den Modellen in der ersten Spalte wird jeweils nur eine unabhängige Variable berücksichtigt. Bei den Modellen in der zweiten Spalte werden zusätzlich zu der jeweiligen Zeilenvariable die Kompetenzen in den Domänen Mathematik und Naturwissenschaften kontrolliert.

Abbildung 11.2: Testleistungen der Schülerinnen und Schüler differenziert nach Schullaufbahnpräferenz der Lehrkräfte – Gesamtskala Mathematik

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Entsprechend werden bei den Modellen in der dritten Spalte zusätzlich zu der jeweiligen Zeilenvariable die Noten in den Fächern Deutsch, Mathematik und Sachunterricht kontrolliert.

Es zeigt sich, dass fast alle Effekte (abgesehen vom Selbstkonzept Sachunterricht unter Kontrolle der Noten) signifikant ausfallen. Als besonders eng erweist sich (ohne Kontrolle weiterer Variablen) erwartungskonform der Zusammenhang mit den Schulnoten aber auch mit den standardisierten Leistungstests, gefolgt von Anstrengungsbereitschaft, Selbstkonzept Mathematik, Leistungsangst und Selbstkonzept Sachunterricht.

Unter Kontrolle der mathematischen und der naturwissenschaftlichen Kompetenz lassen sich für die übrigen Merkmale weiterhin relativ große Effekte feststellen, wobei die Schulnoten wiederum den engsten Zusammenhang mit der Schullaufbahnpräferenz aufweisen. Deutlich kleinere Regressionskoeffizienten ergeben sich für die Modelle, in denen die Noten kontrolliert werden. Neben den Kompetenzen zeigen sich für die Anstrengungsbereitschaft und für das Selbstkonzept Mathematik die engsten Zusammenhänge mit der abhängigen Variable.

In vorangegangenen Studienzyklen wurden diese Zusammenhänge mit Korrelationskoeffizienten statt mit standardisierten Regressionskoeffizienten berichtet. Mit Einschränkungen lassen sich die Befunde dennoch vergleichen und es zeigen sich insgesamt kaum Veränderungen (Arnold et al., 2007, Tabelle X.6; Stubbe et al., 2012, Tabelle 8.7; Stubbe et al., 2017, Tabelle 8.5).

Tabelle 11.4: Zusammenhang zwischen der Schullaufbahnpräferenz der Lehrkräfte und leistungsrelevanten Merkmalen der Schülerinnen und Schüler (ohne und mit Berücksichtigung der Kompetenzen und Schulnoten)

	Zusammenhang mit der Gymnasialpräferenz der Lehrkräfte		Zusammenhang mit der Gymnasialpräferenz der Lehrkräfte unter Kontrolle der Mathematik- und Naturwissenschaftskompetenz		Zusammenhang mit der Gymnasialpräferenz der Lehrkräfte unter Kontrolle der Deutsch-, Mathematik- und Sachunterrichtsnoten	
	beta	(SE)	beta	(SE)	beta	(SE)
Mathematische Kompetenz	0.55**	(0.01)			0.16**	(0.02)
Naturwissenschaftliche Kompetenz	0.49**	(0.02)			0.11**	(0.02)
Anstrengungsbereitschaft	0.43**	(0.02)	0.27**	(0.02)	0.11**	(0.02)
Leistungsangst	- 0.30**	(0.02)	- 0.13**	(0.02)	- 0.05**	(0.02)
Note in Deutsch	- 0.64**	(0.01)	- 0.49**	(0.02)		
Note in Mathematik	- 0.62**	(0.01)	- 0.46**	(0.02)		
Note im Sachunterricht	- 0.56**	(0.02)	- 0.38**	(0.02)		
Selbstkonzept Mathematik	0.39**	(0.02)	0.19**	(0.02)	0.09**	(0.02)
Selbstkonzept Sachunterricht	0.25**	(0.02)	0.11**	(0.02)	— ^{ns}	— ^{ns}

Signifikanzniveau: ns = nicht signifikant; * = signifikant ($p \leq .05$); ** = signifikant ($p \leq .01$)

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

11.4 Schullaufbahnpräferenzen von Lehrkräften und Erziehungsberechtigten: Soziale Merkmale der Schülerinnen und Schüler

Während die Berücksichtigung leistungsrelevanter Merkmale der Schülerinnen und Schüler beim Übergang in die Sekundarstufe bildungspolitisch vorgegeben ist, dürften Hintergrundmerkmale der Kinder sowie ihrer Familien, die nicht unmittelbar mit Leistung in Verbindung stehen, keinen Zusammenhang mit den Präferenzen der Grundschullehrkräfte sowie der eigenen Erziehungsberechtigten aufweisen. Dies widerspräche dem gesellschaftlichen Anspruch auf gleiche Chancen für höhere Bildungsabschlüsse unabhängig vom sozialen Status des Elternhauses. Zahlreiche Studien weisen diese Zusammenhänge jedoch regelmäßig nach (siehe Kapitel 9 in diesem Band), sodass im Folgenden Merkmale des sozialen Status analysiert werden. Die soziale Lage von Familien wird dabei mit Hilfe der EGP-Klassen (Erikson, Goldthorpe & Portocarero, 1979) beschrieben, die Personen ausgehend von Angaben zu ihrer beruflichen Stellung distinkten Gruppen zuordnen (siehe Kapitel 9 in diesem Band).

In Tabelle 11.5 sind die relativen Chancen (*odds ratios*) für eine Gymnasialpräferenz der Lehrkräfte und der Erziehungsberechtigten in Abhängigkeit von der EGP-Klasse der Familie dargestellt. Als Referenzgruppe dienen jeweils Kinder, deren Lehrkräfte beziehungsweise Erziehungsberechtigte eine Präferenz für eine andere Schulform als das Gymnasium haben und deren Erziehungsberechtigte der Gruppe der (Fach-)Arbeiter angehören.

Ohne Berücksichtigung weiterer Einflussfaktoren (Modell I) ist die Chance einer Schülerin beziehungsweise eines Schülers aus der oberen Dienstklasse, eine Gymnasialpräferenz der Lehrkraft zu erhalten, knapp viermal so hoch wie die eines (Fach-)Arbeiterkindes. Alle übrigen Effekte sind nicht signifikant. Trotzdem lassen sich allein mit der EGP-Klassenzugehörigkeit 13 Prozent der Varianz in der abhängigen Variable erklären. Da ein Zusammenhang zwischen dem sozialen Status von Familien und den schulischen Leistungen der Kinder besteht (siehe Kapitel 9 in diesem Band), werden in Modell II zunächst

Tabelle 11.5: Relative Chancen (*odds ratios*) für eine Gymnasialpräferenz der Lehrkräfte bzw. der Erziehungsberechtigten nach sozialer Lage (EGP) der Schülerinnen- und Schülerfamilien

Sozioökonomische Stellung der Familie (EGP-Klassen)	Gymnasialpräferenz der Lehrkräfte			Gymnasialpräferenz der Erziehungsberechtigten		
	Modell I	Modell II	Modell III	Modell I	Modell II	Modell III
Obere Dienstklasse (I)	3.87**	3.30*	— ^{ns}	3.92**	3.31*	— ^{ns}
Untere Dienstklasse (II)	— ^{ns}	— ^{ns}	— ^{ns}	2.50**	2.32**	2.05*
Routinedienstleistungen (III)	— ^{ns}	— ^{ns}	— ^{ns}	— ^{ns}	— ^{ns}	— ^{ns}
Selbstständige (IV)	— ^{ns}	— ^{ns}	— ^{ns}	— ^{ns}	— ^{ns}	— ^{ns}
(Fach-)Arbeiter (V, VI)	Referenzgruppe (<i>odds ratio</i> = 1)					
Un- und angelernte Arbeiter (VII)	— ^{ns}	— ^{ns}	— ^{ns}	— ^{ns}	— ^{ns}	— ^{ns}
Nagelkerkes R^2	0.13	0.30	0.45	0.12	0.24	0.36

Signifikanzniveau: ns = nicht signifikant; * = signifikant ($p \leq .05$); ** = signifikant ($p \leq .01$)

Modell I: Ohne Kontrolle von Kovariaten

Modell II: Kontrolle von kognitiven Fähigkeiten

Modell III: Kontrolle von kognitiven Fähigkeiten und der Mathematik- und Naturwissenschaftskompetenzen

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

die kognitiven Fähigkeiten³ und in Modell III zudem die Mathematik- und Naturwissenschaftskompetenz der Viertklässlerinnen und Viertklässler kontrolliert, um zu prüfen, ob ungleiche Chancen in ungleichen Leistungen begründet liegen. In Modell II lässt sich noch ein signifikanter Effekt feststellen; in Modell III aber nicht. Durch die Hinzunahme von Kontrollvariablen erklären diese Modelle 30 beziehungsweise 45 Prozent der Varianz der Gymnasialpräferenz der Lehrkräfte.

Bei den Gymnasialpräferenzen der Erziehungsberechtigten zeigen sich sowohl in Modell I als auch in Modell II signifikante Effekte bei den beiden Dienstklassen. In Modell III kann nur für die untere Dienstklasse ein signifikanter Effekt festgestellt werden. Die Erklärungskraft dieser Modelle fällt etwas geringer aus als bei den Modellen für die Lehrkräfte.

Für den Trend seit 2011 werden in Tabelle 11.6 die relativen Chancen (*odds ratios*) für die sogenannte *service class* (EGP I und II) im Vergleich zur *working class* (EGP V, VI, VII) für eine Gymnasialpräferenz der Lehrkräfte und der Erziehungsberechtigten berichtet. Ohne Berücksichtigung weiterer Einflussfaktoren (Modell I) zeigten sich sowohl bei den Lehrkräften als auch bei den Erziehungsberechtigten 2011 die größten sozialen Disparitäten und 2015 die geringsten. Die Werte für die aktuelle Studie liegen dazwischen. Unter Kontrolle der kognitiven Fähigkeiten (Modell II) fallen alle Effekte geringer aus, sind aber weiterhin signifikant. Werden zusätzlich die mathematischen und naturwissenschaftlichen Kompetenzen kontrolliert (Modell III), lassen sich für 2011 weder für die Präferenzen der Lehrkräfte noch für die Präferenzen der Erziehungsberechtigten signifikante Effekte des sozialen Status feststellen. Ebenso ist es 2015 im Modell der Erziehungsberechtigten. Die übrigen drei Koeffizienten sind signifikant, aber geringer als die entsprechenden Werte aus Modell II.

Eine alternative und anschauliche Möglichkeit den Zusammenhang zwischen Hintergrundmerkmalen und Schullaufbahnpräferenzen darzustellen, ist die Berechnung der Kompetenzwerte, die notwendig sind, damit Viertklässlerinnen und Viertklässler mit einer Wahrscheinlichkeit von mindestens 50 Prozent eine Präferenz für das Gymnasium erhalten. Diese Kompetenzwerte wurden in frühe-

3 Die erhobenen Subskalen des kognitiven Fähigkeitstests (KFT) wurden seit TIMSS 2007 unter der Annahme der Messinvarianz skaliert.

Tabelle 11.6: Relative Chancen (*odds ratios*) für eine Gymnasialpräferenz der Lehrkräfte bzw. der Erziehungsberechtigten für Kinder aus Familien der *service class* (EGP I und II) im Vergleich mit Kindern aus Familien der *working class* (EGP V, VI und VII) bei TIMSS 2011, 2015 und 2019

	Gymnasialpräferenz der Lehrkräfte			Gymnasialpräferenz der Erziehungsberechtigten		
	Modell I	Modell II	Modell III	Modell I	Modell II	Modell III
2019	3.91**	3.37**	2.41*	3.86**	3.38**	2.53*
2015	3.67**	3.36**	2.65**	3.04**	2.65**	— ^{ns}
2011	4.48**	3.79*	— ^{ns}	4.61**	3.86*	— ^{ns}

Signifikanzniveau: ns = nicht signifikant; * = signifikant ($p \leq .05$); ** = signifikant ($p \leq .01$)

Modell I: Ohne Kontrolle von Kovariaten

Modell II: Kontrolle der kognitiven Fähigkeiten

Modell III: Kontrolle der kognitiven Fähigkeiten und der Mathematik- und Naturwissenschaftskompetenzen

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

ren IGLU- und TIMSS-Zyklen als ‚kritische Werte‘ beziehungsweise gruppenspezifische Standards bezeichnet. Da für TIMSS 2019 ein geringfügig abweichendes Berechnungsverfahren gewählt wurde, wird im Folgenden von Schwellenwerten gesprochen.

Tabelle 11.7 stellt diese Schwellenwerte getrennt für Lehrkräfte und Erziehungsberechtigte sowie getrennt für Mathematik und Naturwissenschaften dar. Als Referenz werden zunächst die Schwellenwerte für die Gesamtpopulation berichtet. Dabei zeigt sich, dass die Viertklässlerinnen und Viertklässler für eine wahrscheinliche Gymnasialpräferenz ihrer Lehrkräfte in Mathematik eine Kompetenz benötigen, die 28 Punkte über dem nationalen Mittelwert liegt, und in Naturwissenschaften eine Kompetenz, die 34 Punkte über dem nationalen Mittelwert liegt. Die Schwellenwerte der Erziehungsberechtigten liegen nur 3 Punkte (Mathematik) beziehungsweise 5 Punkte (Naturwissenschaften) unter denen der Lehrkräfte. Es kann also nicht grundsätzlich festgestellt werden, dass die Erziehungsberechtigten geringere Leistungen für eine Gymnasialpräferenz erwarten als Lehrkräfte.

Betrachtet man die Werte getrennt nach EGP-Klasse, zeigt sich, dass Kinder aus der oberen Dienstklasse eine Mathematikkompetenz von nur 518 Punkten (3 Punkte unter dem nationalen Mittelwert) benötigen, damit eine fünfzigprozentige Wahrscheinlichkeit für eine Gymnasialpräferenz der Lehrkräfte besteht. Im Bereich Naturwissenschaften liegt dieser Wert sogar bei nur 506 Punkten (12 Punkte unter dem nationalen Mittelwert). Kinder von (Fach-)Arbeitern benötigen für eine entsprechende Präferenz in Mathematik eine 43 Punkte höhere Kompetenz und in Naturwissenschaften sogar eine 61 Punkte höhere Kompetenz⁴. Die Differenzen zwischen diesen beiden Gruppen liegen somit deutlich über dem, was Grundschulkinder durchschnittlich in der vierten Klasse dazulernen (Wendt, Kasper, Bos, Vennemann & Goy, 2017).

Noch etwas höher fallen diese Differenzen aus, wenn man die Schwellenwerte der Erziehungsberechtigten betrachtet: 51 Punkte im Bereich Mathematik und 74 Punkte in Naturwissenschaften. Dies liegt in ungünstigeren Präferenzen der unteren sozialen Lagen begründet. Während sich die Schwellenwerte in den beiden Dienstklassen zwischen Lehrkräften und Erziehungsberechtigten praktisch nicht unterscheiden und in der Gruppe der Routinedienstleistungen bei den Erziehungsberechtigten sogar etwas geringer ausfallen als bei den Lehrkräften,

4 Inkonsistenzen zwischen den im Text und in den Abbildungen berichteten Werten sind im Rundungsverfahren begründet.

Tabelle 11.7: Schwellenwerte der Mathematik- und Naturwissenschaftskompetenz für eine fünfzigprozentige Wahrscheinlichkeit für eine Gymnasialpräferenz der Lehrkräfte beziehungsweise der Erziehungsberechtigten

	Schwellenwert für eine Gymnasialpräferenz der Lehrkräfte		Schwellenwert für eine Gymnasialpräferenz der Erziehungsberechtigten	
	Mathematik	Naturwissenschaften	Mathematik	Naturwissenschaften
Gesamt	549	552	546	547
Obere Dienstklasse (I)	518	506	519	504
Untere Dienstklasse (II)	529	526	529	526
Routinedienstleistungen (III)	550	556	542	545
Selbstständige (IV)	547	556	566	579
(Fach-)Arbeiter (V,VI)	562	567	571	578
Un- und angelernte Arbeiter (VII)	566	579	579	603
kein Elternteil im Ausland geboren	547	554	551	558
ein Elternteil im Ausland geboren	541	539	535	531
beide Elternteile im Ausland geboren	557	543	530	507
Mädchen	535	540	538	542
Jungen	562	563	554	552

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

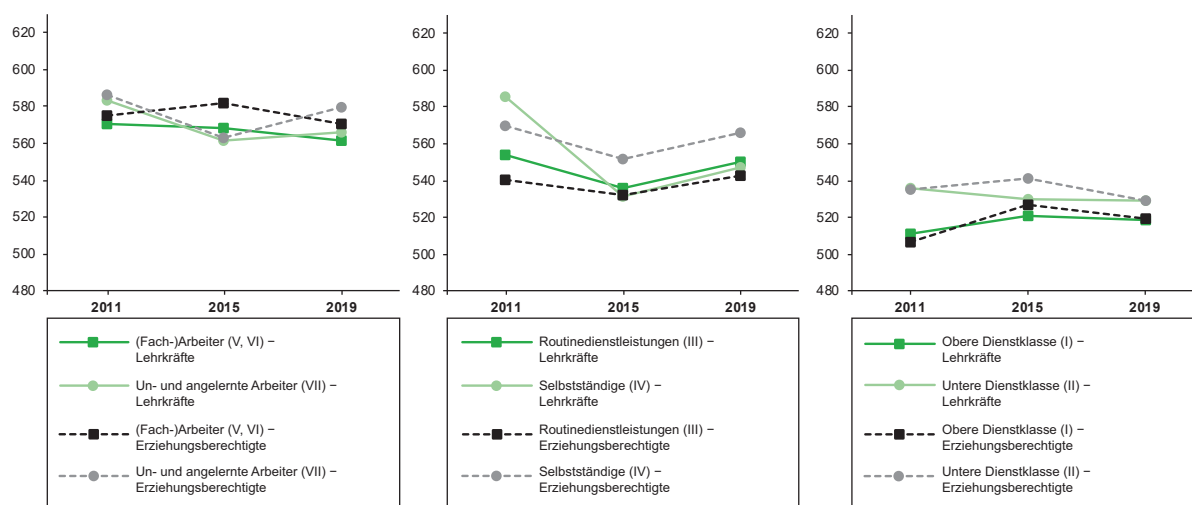
liegen die Werte in den übrigen EGP-Klassen bei den Erziehungsberechtigten höher als bei den Lehrkräften.

Für den Kompetenzbereich Mathematik sind in Abbildung 11.3 die Schwellenwerte der Lehrkräfte und der Erziehungsberechtigten für TIMSS 2011, 2015 und 2019 getrennt nach EGP-Klasse dargestellt. Betrachtet man das gesamte Intervall von acht Jahren, fallen nur drei Werte 2019 höher aus als 2011: Erziehungsberechtigte/Routinedienstleistungen (2 Punkte), Lehrkräfte/Obere Dienstklasse (8 Punkte) und Erziehungsberechtigte/Obere Dienstklasse (13 Punkte). Von Kindern aus der oberen Dienstklasse erwarten 2019 also sowohl Lehrkräfte als auch Erziehungsberechtigte eine etwas höhere Kompetenz als acht Jahre zuvor, damit eine Gymnasialpräferenz entsprechend wahrscheinlich wird. Die übrigen Werte liegen in TIMSS 2019 mit zwei Ausnahmen 3 bis 9 Punkte unter denen aus 2011. Höhere Differenzen zeigen sich für die Präferenzen der Lehrkräfte in der Gruppe der un- und angelernten Arbeiter (17 Punkte) und in der Gruppe der Selbstständigen (38 Punkte). Abgesehen von der oberen Dienstklasse sind die Erwartungen der Lehrkräfte und der Erziehungsberechtigten 2019 im Vergleich zu 2011 also praktisch unverändert oder sogar etwas geringer.

Der direkte Vergleich der Schwellenwerte für die obere Dienstklasse mit denen der Gruppe der (Fach-)Arbeiter offenbart für 2011 etwas größere Unterschiede als in den beiden Folgestudien. Bei den Präferenzen der Lehrkräfte lag die Differenz 2011 bei 60 Punkten, 2015 bei 47 Punkten und 2019 bei 43 Punkten. Die Werte für die Erziehungsberechtigten betragen 69 Punkte (2011), 55 Punkte (2015) und 51 Punkte (2019). Der Grad der sozialen Disparitäten scheint also leicht rückläufig zu sein, da die Anforderungen für eine Gymnasialpräferenz nicht mehr ganz so stark zwischen den sozialen Lagen variieren. Dennoch benötigen Viertklässlerinnen und Viertklässler aus der *working class* weiterhin einen Leistungsvorsprung von mehr als einem Lernjahr vor Kindern aus der *service class*, um eine hinreichende Wahrscheinlichkeit für eine Gymnasialpräferenz zu erhalten.

Ergänzend zu den sozialen Disparitäten nach EGP-Klasse sind in Tabelle 11.7 auch die Schwellenwerte in Abhängigkeit vom Migrationshintergrund und vom

Abbildung 11.3: Schwellenwerte der Mathematikkompetenz für eine fünfzigprozentige Wahrscheinlichkeit für eine Gymnasialpräferenz der Lehrkräfte beziehungsweise der Erziehungsberechtigten bei TIMSS 2011, 2015 und 2019



Abweichungen zu den im Rahmen von IGLU/TIMSS 2011 und TIMSS 2015 berichteten Werten (Stubbe et al., 2012; Stubbe et al., 2016) ergeben sich aus dem veränderten Berechnungsverfahren und daraus, dass dort für 2015 mit imputierten Daten gearbeitet wurde.

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Geschlecht dargestellt. Hinsichtlich der migrationsbedingten Unterschiede bestätigen sich für den Bereich Mathematik die aus anderen Studien bekannten Befunde (Becker & Gresch, 2016). Lehrkräfte erwarten von Kindern, deren Eltern beide im Ausland geboren wurden, eine 10 Punkte höhere Kompetenz als von Schülerinnen und Schülern ohne Migrationshintergrund. Bei den Präferenzen der Erziehungsberechtigten ist es umgekehrt: Wurden beide Elternteile im Ausland geboren, beträgt der Schwellenwert 530 Punkte und wenn beide Elternteile in Deutschland geboren wurden 551 Punkte. Dieses Ergebnis weist somit auf die relativ hohen Bildungsaspirationen von Familien mit Migrationshintergrund hin.

Im Bereich Naturwissenschaften liegen sowohl die Schwellenwerte der Lehrkräfte als auch die der Erziehungsberechtigten für Kinder mit Migrationshintergrund unter denen von Kindern ohne Migrationshintergrund. Mädchen benötigen sowohl in Mathematik als auch in Naturwissenschaften eine geringere Kompetenz als Jungen, um eine Gymnasialpräferenz zu erhalten. Diese Differenzen liegen zwischen 10 Punkten (Erziehungsberechtigte/Naturwissenschaften) und 27 Punkten (Lehrkräfte/Mathematik).

11.5 Zusammenfassung

Die vergangenen Jahre waren durch zahlreiche Reformen der Sekundarschulstruktur geprägt. Bemerkenswert ist, dass es den Ländern der Bundesrepublik Deutschland nach wie vor nicht gelingt, sich auf einheitliche Namen für die neuen Schulformen zu einigen. Namentlich lassen sich derzeit 15 Schulformen (inkl. Hauptschule, Realschule und Gymnasium) unterscheiden, wobei zum Teil sogar ein Name in verschiedenen Ländern für unterschiedliche Schultypen steht. Es ist anzunehmen, dass die Wahl einer weiterführenden Schule durch diese Tatsache erschwert wird – und zwar insbesondere in bildungsferneren Haushalten.

Die in diesem Kapitel vorgelegten Trendanalysen zum Übergang von der Grundschule in die Sekundarstufe bei TIMSS 2011, 2015 und 2019, die an die

Ergebnisse der vier bisherigen IGLU-Zyklen anschließen, zeigen ein insgesamt relativ konstantes Bild. Veränderungen ergeben sich insbesondere durch die erwähnten Schulstrukturreformen. Fast jedes vierte Kind wechselt den Angaben der Erziehungsberechtigten zufolge 2019 auf eine Schule mit mehreren Bildungsgängen – damit ist dieser Anteil rund dreimal so hoch wie 18 Jahre zuvor. Knapp die Hälfte der Viertklässlerinnen und Viertklässler wechselt auf ein Gymnasium. Seit 2006 ist dieser Wert relativ konstant. Hinsichtlich der Schullaufbahnpräferenzen der Lehrkräfte befinden sich die Werte seit 2011 auf einem gleichbleibenden Niveau.

Es zeigt sich, dass leistungsrelevante Merkmale der Schülerinnen und Schüler entsprechend der bildungspolitischen Vorgaben einen relativ engen Zusammenhang mit der Schullaufbahnpräferenz der Lehrkräfte aufweisen.

Bedenklich ist hingegen der auch unter Kontrolle der Kompetenzen hohe Zusammenhang zwischen der sozialen Herkunft der Schülerinnen und Schüler und den Schullaufbahnpräferenzen sowohl der Lehrkräfte als auch der Erziehungsberechtigten. Viertklässlerinnen und Viertklässler aus der *working class* benötigen sowohl in Mathematik als auch in den Naturwissenschaften einen Leistungsvorsprung von mehr als einem Lernjahr vor Kindern aus der *service class*, um eine hinreichende Wahrscheinlichkeit für eine Gymnasialpräferenz zu erhalten.

Literatur

- Arnold, K.-H., Bos, W., Richert, P. & Stubbe, T. C. (2007). Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 271–297). Münster: Waxmann.
- Arnold, K.-H., Bos, W., Richert, P. & Stubbe, T. C. (2010). Der Übergang von der Grundschule in die Sekundarstufe: Schullaufbahnpräferenzen von Lehrkräften und Eltern im Ländervergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert, I. Tarelli & R. Valtin (Hrsg.), *IGLU 2006 – die Grundschule auf dem Prüfstand. Vertiefende Analysen zu Rahmenbedingungen schulischen Lernens* (S. 13–32). Münster: Waxmann.
- Becker B. & Gresch C. (2016). Bildungsaspirationen in Familien mit Migrationshintergrund. In C. Diehl, C. Hunkler & C. Kristen (Hrsg.), *Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten* (S. 73–115). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-04322-3_3
- Bellenberg, G. & Klemm, K. (2000). Scheitern im System, Scheitern des Systems? Ein etwas anderer Blick auf Schulqualität. In H.-G. Rolff, W. Bos, K. Klemm, H. Pfeiffer & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung*. (Bd. 11, S. 51–75). Weinheim: Juventa.
- Bos, W., Voss, A., Lankes, E.-M., Schwippert, K., Thiel, O. & Valtin, R. (2004). Schullaufbahneempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hrsg.), *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich* (S. 191–228). Münster: Waxmann.
- Erikson, R., Goldthorpe, H. J. & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology*, 30(4), 415–441. <https://doi.org/10.2307/589632>
- KMK (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland). (2015). *Übergang von der Grundschule in Schulen des Sekundarbereichs I und Förderung, Beobachtung und Orientierung in den Jahrgangsstufen 5 und 6 (sog. Orientierungsstufe)*. Verfügbar unter <http://www.kmk.org/fileadmin/Dateien/>

- veroeffentlichungen_beschluesse/2015/2015_02_19-Uebergang_Grundschole-SI-Orientierungsstufe.pdf
- Köller, O. (2008). Gesamtschule – Erweiterung statt Alternative. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (S. 437–465). Reinbek: Rowohlt.
- Lintorf, K. (2012). *Wie vorhersagbar sind Grundschulnoten? Prädiktionskraft individueller und kontextspezifischer Merkmale*. Wiesbaden: Springer VS.
<https://doi.org/10.1007/978-3-531-94339-8>
- Maaz, K. & Nagy, G. (2009). Der Übergang von der Grundschule in die weiterführenden Schulen des Sekundarschulsystems. Definition, Spezifikation und Quantifizierung primärer und sekundärer Herkunftseffekte. In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen, Zeitschrift für Erziehungswissenschaft, Sonderheft 12* (S. 153–182). Wiesbaden: VS Verlag für Sozialwissenschaften.
https://doi.org/10.1007/978-3-531-92216-4_7
- Mahler, N., Schipolowski, S. & Weirich S. (2019). Anlage, Durchführung und Auswertung des IQB-Bildungstrends 2018. Anlage und Durchführung. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S. Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 99–124). Münster: Waxmann.
- Milek, A., Lüdtke, O., Trautwein, U., Maaz, K. & Stubbe, T. C. (2009). Wie konsistent sind Referenzgruppeneffekte bei der Vergabe von Schulformempfehlungen? Bundeslandspezifische Analysen von Daten der IGLU-Studie. In J. Baumert, K. Maaz & U. Trautwein (Hrsg.), *Bildungsentscheidungen, Zeitschrift für Erziehungswissenschaft, Sonderheft 12* (S. 282–301). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92216-4_12
- Rabenstein, K., Horn, K.-P., Keßler, C. I. & Stubbe, T. C. (2020). Inklusion als Herausforderung für das Gymnasium. Zur Einleitung in das Thema und in den Band. In K. Rabenstein, T. C. Stubbe & K.-P. Horn (Hrsg.), *Inklusion und Gymnasium. Studien zu Perspektiven von Lehrkräften und Studierenden* (S. 7–34). Göttingen: Universitätsverlag Göttingen.
- Schuchart, C. (2007). Schulabschluss und Ausbildungsberuf. Zur Bedeutung der schulartbezogenen Bildungsbiografie. *Zeitschrift für Erziehungswissenschaft*, 10(3), 381–398. <https://doi.org/10.1007/s11618-007-0042-7>
- Schuchart, C. (2013). Kein Abschluss ohne Anschluss? Durchlässigkeit und Vergleichbarkeit von Bildungswegen in der Sekundarstufe II. *Die Deutsche Schule*, 105(4), 345–363.
- Schuchart, C. & Weishaupt, H. (2004). Die prognostische Qualität der Übergangsempfehlungen der niedersächsischen Orientierungsstufe. *Zeitschrift für Pädagogik*, 50(6), 882–902.
- Stubbe, T. C. (2009). *Bildungsentscheidungen und sekundäre Herkunftseffekte. Soziale Disparitäten bei Hamburger Schülerinnen und Schülern der Sekundarstufe I*. Münster: Waxmann.
- Stubbe, T. C. & Bos, W. (2008). Schullaufbahneempfehlungen von Lehrkräften und Schullaufbahnentscheidungen von Eltern am Ende der vierten Jahrgangsstufe. *Empirische Pädagogik*, 22(1), 49–63.
- Stubbe, T. C., Bos, W. & Euen, B. (2012). Der Übergang von der Primar- in die Sekundarstufe. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.), *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 209–226). Münster: Waxmann.
- Stubbe, T. C., Bos, W. & Schurig, M. (2017). Der Übergang von der Primar- in die Sekundarstufe. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 235–250). Münster: Waxmann.
- Stubbe, T. C., Lorenz, J., Bos, W. & Kasper, D. (2016). Der Übergang von der Primar- in die Sekundarstufe. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 351–365). Münster: Waxmann.

- Wendt, H., Bos, W., Kasper, D., Walzebug, A., Goy, M. & Jusufi, D. (2016). Ziele, Anlage und Durchführung der *Trends in International Mathematics and Science Study* (TIMSS 2015). In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 31–77). Münster: Waxmann.
- Wendt, H., Kasper, D., Bos, W., Vennemann, M. & Goy, M. (2017). Wie viele Punkte auf der TIMSS-Metrik entsprechen einem Lernjahr? Leistungszuwächse in Mathematik und Naturwissenschaften am Ende der Grundschulzeit. In T. Eckert & B. Gniewosz (Hrsg.), *Bildungsgerechtigkeit* (S. 121–153). Wiesbaden: Springer VS.
https://doi.org/10.1007/978-3-658-15003-7_8

Kapitel 12

Trends in Schülerzusammensetzungen und mathematischen und naturwissenschaftlichen Kompetenzen

Daniel Kasper, Heike Wendt, Knut Schwippert und Olaf Köller

12.1 Einleitung

Mit der *Trends in International Mathematics and Science Study* (TIMSS) 2019 ist es möglich, Entwicklungen von Leistungsständen in mathematischen und naturwissenschaftlichen Kompetenzen von Viertklässlerinnen und Viertklässlern über die Jahre 2007, 2011, 2015 bis 2019 nachzuzeichnen und diese in international vergleichender Perspektive einzuordnen. Wie in den Kapiteln zuvor berichtet, lassen sich für Deutschland aus der Perspektive von TIMSS 2019 im Vergleich zu Befunden der vorangegangenen Studienzyklen folgenden Ergebnisse feststellen:

- Mathematik: Mit einem Leistungsmittelwert von 521 Punkten sind die Ergebnisse stabil, während sich in vielen anderen Staaten (darunter auch Nachbarstaaten) positive Trends beobachten lassen. Hingegen ist der Anteil der Schülerinnen und Schüler, die lediglich die unteren beiden Kompetenzstufen erreichen, in TIMSS 2019 leicht gestiegen und es lässt sich eine leichte Zunahme der Leistungsstreuung beobachten (siehe Kapitel 3 in diesem Band).
- Naturwissenschaften: Mit einem Leistungsmittelwert von 518 Punkten liegen die Ergebnisse 10 Punkte unter dem Mittelwert vorheriger Studienzyklen. Diese Veränderung ist statistisch signifikant. Mit Hongkong und Italien gehört Deutschland zu den wenigen Staaten mit einem entsprechend ungünstigen Befundmuster. Auch lässt sich eine leichte Zunahme der Leistungsstreuung sowie ein gestiegener Anteil der Schülerinnen und Schüler, die lediglich die unteren Kompetenzstufen erreichen, feststellen (siehe Kapitel 4 in diesem Band).

Zu denken geben insbesondere die in 2019 auf 25.3 (Mathematik) beziehungsweise 27.7 Prozent (Naturwissenschaften) gestiegenen Anteile an Schülerinnen und Schülern, die am Ende der vierten Klasse unter Kompetenzstufe III liegen (siehe Kapitel 3 und 4 in diesem Band). Es ist davon auszugehen, dass sie in der Sekundarstufe I mit erheblichen Schwierigkeiten konfrontiert sein werden, dem weiteren Unterricht in Mathematik und den naturwissenschaftlichen Fächern zu folgen. Insbesondere im Hinblick auf die Förderung dieser Kinder besteht somit dringender Handlungsbedarf.

Wie in Kapitel 5 in diesem Band beschrieben, lassen sich die Veränderungen nicht durch die computerbasierte Testform erklären. Selbst wenn die computerbasierte Testform im Vergleich zur papierbasierten Testung für die Schülerinnen und Schüler etwas schwieriger ist, so zeigt sich auch in der Stichprobe, die papierbasiert getestet wurde, eine leichte Abnahme der naturwissenschaftlichen Kompetenzen.

Das Design der Studie lässt es jedoch nicht zu, gesicherte Aussagen darüber zu treffen, worauf solche Veränderungen kausal zurückzuführen sind. Als Ursachen ließen sich unter anderem gesellschaftliche und demografische Entwicklungen sowie strukturelle und inhaltliche Reformen im Bildungswesen diskutieren. Im Rahmen der Berichtslegung von TIMSS 2015 wurden bereits zentrale Entwicklungen im Grundschulbereich seit der Jahrtausendwende skizziert (Kasper, Wendt, Bos & Köller, 2016). In der Zusammenschau von Bildungsreformen der letzten 15 Jahre, die auf das Primarschulwesen in Deutschland gerichtet waren, offenbaren sich vielfältige und vielschichtige Veränderungsprozesse, die auf unterschiedlichen Ebenen des Bildungssystems angegangen wurden. Inwiefern diesen Maßnahmen auch Steuerungsfunktionen zuzuschreiben sind und ob diesen Reformen auch tatsächliche Veränderungen des mathematischen beziehungsweise naturwissenschaftlichen Lehrens und Lernens in der Grundschule gefolgt sind, lässt sich nur schwer trennen. Offensichtlich ist hingegen, dass sich in den letzten Jahren die Zusammensetzung der Schülerschaft verändert hat (siehe Tabelle 12.1; Kapitel 9 und Kapitel 10 in diesem Band sowie die Ergebnisse von TIMSS 2015: Kasper et al., 2016 und der *Internationalen Grundschul-Lese-Untersuchung* (IGLU) 2016: Bremerich-Vos, Wendt & Bos, 2017). Die Interpretationen der Ergebnisse des IQB-Bildungstrends (Stanat, Schipolowski, Rjosk, Weirich & Haag, 2017) liefern Hinweise darauf, dass gestiegene Anteile von Kindern mit Migrationshintergrund, sonderpädagogischem Förderbedarf sowie niedrigem sozioökonomischen Status und damit veränderte Rahmenbedingungen schulischen Lernens, als mögliche Erklärungen für die veränderten Leistungsmittelwerte herangezogen werden können. In TIMSS 2015 konnten bei Modellierungen unter Berücksichtigung von Veränderungen der Struktur der Schülerschaft positive Leistungstrends zu 2011 in Form eines Zuwachses von 12 Punkten (Mathematik) beziehungsweise 22 Punkten (Naturwissenschaften) beobachtet werden (Kasper et al., 2016). In IGLU 2016 ließ sich hingegen auch unter Berücksichtigung der Veränderung relevanter Merkmale der Schülerschaft über die IGLU-Erhebungszyklen hinweg – anders als bei TIMSS – kein signifikant positiver Leistungstrend beobachten (Bremerich-Vos, Wendt & Bos, 2017).

In diesem Beitrag wird zunächst der Frage nachgegangen, inwieweit sich die Schülerzusammensetzung im Querschnitt der untersuchten Schülerpopulationen von TIMSS 2007, 2011, 2015 und 2019 verändert hat. Darauf aufbauend wird geklärt, inwiefern sich Trends in mathematischen und naturwissenschaftlichen Leistungen unter Kontrolle von Veränderungen zentraler Merkmale der

Schülerpopulation darstellen lassen. Es wird untersucht, inwieweit sich die mittleren Schülerinnen- und Schülerleistungen in Deutschland von TIMSS 2007, 2011, 2015 zu 2019 auf veränderte Schülerschaften zurückführen lassen. Weiter wird dargestellt, inwieweit sich eine etwas stärker von Vielfalt geprägte Schülerschaft auch in Klassenzusammensetzungen widerspiegelt.

12.2 Charakteristika der Viertklässlerinnen und Viertklässler in TIMSS 2007, 2011, 2015 und 2019

Demografischer Wandel, Europäisierung, Migration sowie schulstrukturelle Veränderungen (wie etwa Inklusion) bedingen eine veränderte Komposition der Schülerschaft an Grundschulen in Deutschland (Autorengruppe Bildungsberichterstattung, 2016, 2020; Statistisches Bundesamt, 2017). In Tabelle 12.1 sind Entwicklungen für die Zyklen 2007, 2011, 2015 und 2019 dargestellt, die sich anhand zentraler Hintergrundmerkmale der Schülerinnen und Schüler beobachten lassen. In Bezug auf die Komposition der Schülerschaft zeigt sich, dass der Anteil an Kindern aus armutsgefährdeten Familien signifikant gesunken ist. Im Vergleich zu den vorangegangenen Erhebungsrunden ist der Anteil von Kindern mit Migrationshintergrund (beide Eltern im Ausland geboren) in TIMSS 2019 mit 22 Prozent statistisch signifikant höher und spiegelt die realen Veränderungen in der Komposition der Schülerschaft an Grundschulen in Deutschland wider (Statistisches Bundesamt, 2017). Gleiches gilt für den ebenfalls im Vergleich zu 2007 statistisch signifikant höheren Anteil an Schülerinnen und Schülern mit besonderen Unterstützungsbedarfen, dieser liegt in TIMSS 2019 bei 5.9 Prozent. Hierzu zählen Kinder in TIMSS 2015 und 2019, wenn sie nach Angabe der

Tabelle 12.1: Charakteristika von Viertklässlerinnen und Viertklässlern in TIMSS 2007, 2011, 2015 und 2019

	TIMSS 2007 % (SE)	TIMSS 2011 % (SE)	TIMSS 2015 % (SE)	TIMSS 2019 % (SE)
Geschlecht				
Mädchen	49.0 (0.6)	49.3 (0.8)	48.0 (0.7)	49.6 (0.8)
Jungen	51.0 (0.6)	50.7 (0.8)	52.0 (0.7)	50.4 (0.8)
Status				
Regelschülerinnen und -schüler	96.6 (0.7)	94.1 (0.8)	94.1 (0.5)	94.1 (0.5) ▼
Kinder mit bes. Unterstützungsbedarfen	3.4 (0.7)	5.9 (0.8)	5.9 (0.5) ▲	5.9 (0.5) ▲
Migrationshintergrund				
kein Elternteil im Ausland geboren	71.4 (1.4)	72.3 (1.4)	68.0 ¹ (1.5) *	63.6 (1.3) ▼*-
ein Elternteil im Ausland geboren	11.4 (0.7)	11.6 (0.6)	11.9 ¹ (0.7)	14.4 (0.9) ▲*+
beide Elternteile im Ausland geboren	17.2 (1.0)	16.1 (1.1)	20.0 ¹ (1.2) ▲*	22.0 (1.0) ▲*
Kinder aus armutsgefährdeten Familien	33.9 (1.3)	25.4 (1.2) ▼	28.9 (1.2) ▼	21.5 (1.0) ▼*-

Signifikanzniveau:

+/- = Unterschied zu TIMSS 2015 statistisch signifikant ($p \leq .05$)

* = Unterschied zu TIMSS 2011 statistisch signifikant ($p \leq .05$)

▲/▼ = Unterschied zu TIMSS 2007 statistisch signifikant ($p \leq .05$)

¹ = Abweichungen zu den Ergebnissen von TIMSS 2015 (Wendt, Schwippert & Stubbe, 2016) ergeben sich aus einem veränderten Umgang mit fehlenden Werten (siehe Kapitel 2 in diesem Band).

Schule über einen nach Verfahren ihres jeweiligen Landes der Bundesrepublik Deutschland diagnostizierten sonderpädagogischen Förderbedarf verfügen. Da für TIMSS 2007 und 2011 keine schülerinnen- und schülerspezifischen Angaben zu sonderpädagogischen Förderbedarfen vorliegen, wurden hier Schülerinnen und Schüler dann zu den Kindern mit besonderen Förderbedarfen gezählt, wenn es sich bei ihnen nach Angabe der Schule um Inklusionskinder handelte (Wendt, Bos, Goy & Jusufi, 2017). Die Ergebnisse bilden die Praxis ab, dass im Zuge von Inklusion zunehmend mehr Kinder mit sonderpädagogischen Förderbedarfen an Regelschulen unterrichtet werden.

12.3 Trends in den Leistungen der Schülerinnen und Schüler

Im internationalen Vergleich zeigte sich, dass sich für mehr als die Hälfte aller Staaten, die – wie Deutschland – an TIMSS 2007, 2011, 2015 und 2019 teilgenommen haben, zwischen dem Jahr 2007 und dem Jahr 2019 in Mathematik und/oder Naturwissenschaften positive Veränderungen der mittleren Leistungswerte beobachten ließen (siehe Kapitel 3 und 4 in diesem Band). Für Deutschland ließen sich in Mathematik – wie auch für fünf weitere Teilnehmerstaaten (Niederlande, Dänemark, Australien, Hongkong und Neuseeland) – hingegen keine signifikanten Veränderungen in den Schülerinnen- und Schülerleistungen zwischen TIMSS 2007 und 2019 feststellen. Für die Naturwissenschaften zeigt sich zwischen dem Jahr 2007 und dem Jahr 2019 ein statistisch signifikanter Leistungsrückgang von 10 Punkten auf der TIMSS-Gesamtskala. Nur in zwei weiteren Teilnehmerstaaten (Italien und Hongkong) lässt sich ebenfalls ein negativer Trend beobachten. Gleichzeitig lassen sich für den gleichen Zeitraum sowohl in der amtlichen Statistik (Statistisches Bundesamt, 2017) als auch den TIMSS-Stichproben Veränderungen in zentralen Schülerinnen- und Schülermerkmalen feststellen (siehe Abschnitt 12.1).

12.3.1 Trends in den mittleren Schülerinnen- und Schülerleistungen unter Berücksichtigung der Mehrebenenstruktur

Komplexe Trendanalysen bieten die Möglichkeit, Veränderungen der Leistungen zwischen TIMSS 2007, 2011, 2015 und 2019 zu modellieren, die sich ergeben hätten, wenn die Zusammensetzung der Schülerschaft hinsichtlich verschiedener soziodemografischer Merkmale gleichgeblieben wäre. Entsprechende Analysen können Hinweise darauf liefern, inwieweit der bei TIMSS 2019 im Vergleich zu 2007 beobachtbare gestiegene Anteil von Kindern mit Migrationshintergrund beziehungsweise sonderpädagogischem Förderbedarf sowie geringere Anteile von Kindern aus armutsgefährdeten Familien und damit veränderte Rahmenbedingungen schulischen Lernens, als Erklärung für die veränderten Leistungsmittelwerte herangezogen werden sollte. Eine Möglichkeit, die Bedeutung dieser Veränderungen für die mathematischen und naturwissenschaftlichen Leistungsentwicklungen von Viertklässlerinnen und Viertklässlern abzuschätzen und damit quasi einen Nettoeffekt der Veränderung zu ermitteln, bietet die Anwendung von Trendmodellen. In diesen Modellen wird die Leistungsdifferenz zwischen den Studienzyklen ermittelt, die sich ergeben hätte, wenn sich die soziodemografischen Merkmale in der Schülerschaft nicht über die Zeit verändert hätten. Als erklärende Faktoren der Leistung wurden der

Messzeitpunkt, die soziodemografischen Merkmale und die Interaktionen zwischen den soziodemografischen Merkmalen und dem Messzeitpunkt berücksichtigt. Für die Analysen kommt ein lineares Modell mit Zufallseffekten in Form eines Mehrebenenmodells zum Einsatz (McCulloch, Searle & Neuhaus, 2008). Da die Schülerinnen und Schüler in der TIMSS-Stichprobe im Klassenverbund getestet werden (siehe Kapitel 2 in diesem Band), ist davon auszugehen, dass ihre Leistungen (ohne Kontrolle der Klassen- und Schulzugehörigkeit) voneinander abhängig sind. Abhängigkeiten zwischen den Leistungen, die sich aus klassenbedingten Gemeinsamkeiten ergeben können, werden im Mehrebenenmodell durch die Aufnahme eines klassenspezifischen *Intercept*-Wertes kontrolliert. Da die in TIMSS untersuchten Klassen außerdem eine Zufallsauswahl von allen zur definierten Zielpopulation gehörenden Klassen in Deutschland darstellen, wird der *Intercept*-Wert als zufällig angenommen, das heißt, es wurde ein sogenanntes *random intercept model* berechnet (Raudenbush & Bryk, 2002).

Die Tabellen 12.2 und 12.3 zeigen, wie sich die Leistungstrends von Viertklässlerinnen und Viertklässlern in Deutschland in TIMSS unter Kontrolle von Schülerinnen- und Schülercharakteristika darstellen. Als ein Ergebnis dieser Analyse ist zunächst festzuhalten, dass sich ein Viertel aller Unterschiede (Mathematik: 26.6% und Naturwissenschaften: 24.6%) der beobachteten Schülerinnen- und Schülerleistungen auf Unterschiede zwischen den Schulen beziehungsweise Klassen zurückführen lässt und die verbleibenden etwa drei Viertel auf individuelle Unterschiede zwischen den einzelnen Schülerinnen und Schülern innerhalb der Klassen zurückgehen. Diese Varianzanteile entsprechen Werten aus vergleichbaren Studien (Bremerich-Vos, Wendt & Bos, 2017; Kasper et al., 2016; siehe Tabelle 6.1 in diesem Band). Ein durchaus beachtlicher Anteil der Differenzen bei den mathematischen und naturwissenschaftlichen Kompetenzen ist also auf Unterschiede zwischen den Schulen beziehungsweise Klassen zurückzuführen.

In den Modellen 1 wurden die in den Abbildungen 3.6 und 4.6 (siehe Kapitel 3 und 4 in diesem Band) dargestellten Leistungstrends unter Berücksichtigung der Mehrebenenstruktur der Daten geschätzt. Unter Berücksichtigung der Mehrebenenstruktur zeigt sich in Modell 1 in Bezug auf die Leistungstrends, dass sich sowohl für Mathematik als auch die Naturwissenschaften keine signifikanten Unterschiede in den mittleren Schülerinnen- und Schülerleistungen zwischen 2007 und 2019 nachweisen lassen. Dies widerspricht jedoch nicht der generellen Befundlage in Kapitel 4 in diesem Band, da zum einen Mehrebenenmodelle unter bestimmten Bedingungen zu Über- beziehungsweise Unterschätzungen führen können (McCulloch et al., 2008), zum anderen aber auch die Betrachtungsperspektive eine andere ist. Die in Kapitel 4 in diesem Band beschriebene Leistungsdifferenz von 10 Punkten von 2007 zu 2019 für die Naturwissenschaften dokumentiert einen generellen (globalen) Unterschied zwischen den Grundschulkindern, ohne dabei zunächst ihre Zugehörigkeit zu ihren Schulen beziehungsweise Klassen zu berücksichtigen. Wird differenzierter die gruppierte Zugehörigkeit der Schülerinnen und Schüler zu ihren Klassen berücksichtigt, werden die Befunde auf die mittleren Klassenleistungen bezogen, die – so modelliert – nicht signifikant zwischen den Erhebungen differieren, obwohl die generellen Unterschiede auf globaler Ebene weiterhin bestehen. In den Modellen 2 bis 6 werden die Leistungstrends unter Berücksichtigung zentraler Schülerinnen- und Schülercharakteristika analysiert. In Modell 2 wird zunächst das Geschlecht berücksichtigt. In Modell 3 werden Indikatoren für den Migrationshintergrund und Kinder mit besonderem Unterstützungsbedarf mit aufge-

Tabelle 12.2: Veränderungen der Schülerinnen- und Schülerleistungen in Mathematik zwischen TIMSS 2007, 2011, 2015 und 2019 unter Kontrolle von Schülerinnen- und Schülermerkmalen (Mehrebenenmodell)

	Modell 1		Modell 2		Modell 3		Modell 4		Modell 5		Modell 6	
	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)	<i>b</i>	(SE)
Konstante	521.6	(3.7)**	514.0	(3.4)**	427.9	(9.9)**	-770.8	(43.6)**	-776.2	(44.6)**	-1116.5	(74.4)**
Testzeitpunkt												
TIMSS 2011 ^A	3.9	(4.8) ^{ns}	1.7	(5.5) ^{ns}	6.9	(4.9) ^{ns}	-6.1	(3.8) ^{ns}	-2.1	(6.7) ^{ns}	-2.6	(6.4) ^{ns}
TIMSS 2015 ^B	-1.6	(4.2) ^{ns}	7.2	(4.7) ^{ns}	6.4	(5.0) ^{ns}	0.4	(3.9) ^{ns}	3.3	(6.8) ^{ns}	3.1	(6.9) ^{ns}
TIMSS 2019 ^C	-2.2	(4.6) ^{ns}	3.8	(4.4) ^{ns}	2.6	(4.9) ^{ns}	-7.4	(4.1) [*]	-2.4	(7.3) ^{ns}	-3.3	(7.1) ^{ns}
Schülerebene												
Geschlecht ^D			10.5	(4.6)**	12.5	(4.5)**	18.7	(3.6)**	18.7	(3.6)**	18.6	(3.6)**
MGH (ein Elternteil) ^E					13.1	(4.5)**	4.3	(3.7) ^{ns}	3.9	(6.4) ^{ns}	3.6	(6.3) ^{ns}
MGH (kein Elternteil) ^F					28.4	(3.4)**	10.8	(2.7)**	15.0	(4.8)**	13.2	(4.7)**
Kein SPF ^G					73.0	(8.9)**	45.4	(6.8)**	48.9	(10.1)**	43.2	(10.3)**
Geschlecht × T11			0.8	(6.2) ^{ns}	-3.2	(6.1) ^{ns}	-5.3	(5.5) ^{ns}	9.6	(15.6) ^{ns}	11.9	(14.3) ^{ns}
Geschlecht × T15			-2.0	(6.1) ^{ns}	-6.9	(6.7) ^{ns}	-8.5	(5.5) ^{ns}	-4.2	(17.0) ^{ns}	-3.7	(17.2) ^{ns}
Geschlecht × T19			-6.3	(6.1) ^{ns}	0.3	(6.8) ^{ns}	-1.0	(5.8) ^{ns}	-4.8	(19.2) ^{ns}	-2.1	(18.4) ^{ns}
MGH (ein Elternteil) × T11									-1.1	(9.1) ^{ns}	-1.2	(8.9) ^{ns}
MGH (kein Elternteil) × T11									-5.4	(7.4) ^{ns}	-5.9	(7.2) ^{ns}
MGH (ein Elternteil) × T15									0.4	(9.4) ^{ns}	0.8	(9.3) ^{ns}
MGH (kein Elternteil) × T15									-4.1	(6.9) ^{ns}	-4.1	(6.8) ^{ns}
MGH (ein Elternteil) × T19									1.8	(10.7) ^{ns}	2.3	(10.7) ^{ns}
MGH (kein Elternteil) × T19									-7.6	(8.2) ^{ns}	-7.8	(8.0) ^{ns}
Geschlecht × Kein SPF × T11									-15.8	(15.5) ^{ns}	-18.0	(14.1) ^{ns}
Geschlecht × Kein SPF × T15									-4.4	(17.1) ^{ns}	-5.0	(17.2) ^{ns}
Geschlecht × Kein SPF × T19									4.1	(18.8) ^{ns}	1.2	(18.1) ^{ns}
Familien mit hohem sozioökonomischen Status ^H							5.0	(0.3)**	5.0	(0.3)**	4.7	(0.3)**
Kognitive Fähigkeiten							7.4	(0.3)**	7.4	(0.3)**	7.3	(0.3)**
Schulebene												
Mittelwert des sozioökonomischen Status											3.5	(0.7)**
Anteil der Schülerinnen und Schüler ohne sonderpädagogischen Förderbedarf											24.2	(15.6) ^{ns}
Anteil der Schülerinnen und Schüler mit Migrationshintergrund (ein Elternteil)											-10.6	(11.4) ^{ns}
Anteil der Schülerinnen und Schüler ohne Migrationshintergrund (kein Elternteil)											10.2	(4.7) [*]
Erklärte Varianzanteile												
Zwischen den Schulen (26.6%)	0.5		0.5		0.5		0.5		0.5		72.1	
Innerhalb der Schulen (73.3%)	0.0		0.6		5.4		31.2		31.3		31.3	
Gesamt	0.1		0.5		4.1		23.0		23.1		42.2	

Anmerkungen: MGH = Migrationshintergrund; SPF = sonderpädagogischer Förderbedarf; T11 = TIMSS 2011; T15 = TIMSS 2015; T19 = TIMSS 2019

b = Regressionsgewicht (unstandardisiert)

Signifikanzniveau: ns = nicht signifikant; * = signifikant ($p \leq .05$); ** = signifikant ($p \leq .01$)

A = Testzeitpunkt (0 = TIMSS 2007; 1 = TIMSS 2011; 0 = TIMSS 2015; 0 = TIMSS 2019)

B = Testzeitpunkt (0 = TIMSS 2007; 0 = TIMSS 2011; 1 = TIMSS 2015; 0 = TIMSS 2019)

C = Testzeitpunkt (0 = TIMSS 2007; 0 = TIMSS 2011; 0 = TIMSS 2015; 1 = TIMSS 2019)

D = Geschlecht (0 = Mädchen; 1 = Jungen)

E = Migrationshintergrund nach Geburtsland der Eltern (0 = beide Elternteile im Ausland geboren; 1 = ein Elternteil im Ausland geboren; 0 = kein Elternteil im Ausland geboren)

F = Migrationshintergrund nach Geburtsland der Eltern (0 = beide Elternteile im Ausland geboren; 0 = ein Elternteil im Ausland geboren; 1 = kein Elternteil im Ausland geboren)

G = Sonderpädagogischer Förderbedarf (0 = sonderpädagogischer Förderbedarf; 1 = kein sonderpädagogischer Förderbedarf)

H = Sozioökonomischer Status der Familie: Sozialindex (unter der Annahme der Messinvarianz seit TIMSS 2007 skaliert)

Tabelle 12.3: Veränderungen der Schülerinnen- und Schülerleistungen in Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 unter Kontrolle von Schülerinnen- und Schülermerkmalen (Mehrebenenmodell)

	Modell 1		Modell 2		Modell 3		Modell 4		Modell 5		Modell 6	
	b	(SE)	b	(SE)	b	(SE)	b	(SE)	b	(SE)	b	(SE)
Konstante	524.8	(3.8)**	515.2	(3.5)**	414.8	(10.7)**	-853.9	(44.5)**	-864.7	(45.0)**	-1169.5	(81.7)**
Testzeitpunkt												
TIMSS 2011 ^A	0.6	(5.1) ^{ns}	2.3	(5.4) ^{ns}	3.9	(5.3) ^{ns}	-8.8	(4.2)*	2.5	(7.9) ^{ns}	1.9	(7.7) ^{ns}
TIMSS 2015 ^B	1.8	(4.6) ^{ns}	4.0	(5.3) ^{ns}	13.8	(5.1)**	7.8	(4.0)*	15.1	(8.7)*	14.8	(8.7)*
TIMSS 2019 ^C	-4.3	(4.4) ^{ns}	10.1	(5.1)*	6.8	(5.1) ^{ns}	-4.2	(4.4) ^{ns}	-1.5	(9.5) ^{ns}	-2.5	(9.3) ^{ns}
Schülerebene												
Geschlecht ^D			3.3	(5.3) ^{ns}	15.0	(5.1)**	21.7	(4.4)**	21.7	(4.4)**	21.6	(4.4)**
MGH (ein Elternteil) ^E					26.1	(4.6)**	16.1	(4.0)**	16.4	(8.4)*	16.0	(8.4)*
MGH (kein Elternteil) ^F					50.7	(3.7)**	31.6	(3.1)**	38.8	(6.4)**	36.7	(6.4)**
Kein SPF ^G					70.3	(9.2)**	40.6	(7.3)**	46.5	(10.9)**	41.7	(11.0)**
Geschlecht × T11			10.4	(7.1) ^{ns}	-2.4	(6.9) ^{ns}	-5.0	(6.3) ^{ns}	5.6	(15.8) ^{ns}	7.8	(14.9) ^{ns}
Geschlecht × T15			8.5	(6.7) ^{ns}	-11.9	(7.4) ^{ns}	-14.0	(6.5)*	-4.0	(20.1) ^{ns}	-3.7	(19.9) ^{ns}
Geschlecht × T19			-1.0	(6.8) ^{ns}	-8.9	(7.5) ^{ns}	-10.6	(6.8) ^{ns}	-1.9	(20.0) ^{ns}	-0.1	(19.7) ^{ns}
MGH (ein Elternteil) × T11									-7.0	(11.8) ^{ns}	-7.2	(11.7) ^{ns}
MGH (kein Elternteil) × T11									-14.6	(8.2)*	-15.1	(8.2)*
MGH (ein Elternteil) × T15									0.0	(11.7) ^{ns}	0.5	(11.6) ^{ns}
MGH (kein Elternteil) × T15									-10.3	(9.2) ^{ns}	-10.4	(9.1) ^{ns}
MGH (ein Elternteil) × T19									5.8	(13.3) ^{ns}	6.4	(13.3) ^{ns}
MGH (kein Elternteil) × T19									-4.6	(10.0) ^{ns}	-5.0	(10.0) ^{ns}
Geschlecht × Kein SPF × T11									-11.0	(15.9) ^{ns}	-13.1	(14.9) ^{ns}
Geschlecht × Kein SPF × T15									-10.2	(19.8) ^{ns}	-10.7	(19.7) ^{ns}
Geschlecht × Kein SPF × T19									-9.0	(19.9) ^{ns}	-11.1	(19.6) ^{ns}
Familien mit hohem sozioökonomischen Status ^H							5.8	(0.4)**	5.8	(0.4)**	5.4	(0.4)**
Kognitive Fähigkeiten							7.4	(0.3)**	7.3	(0.3)**	7.3	(0.3)**
Schulebene												
Mittelwert des sozioökonomischen Status											3.2	(0.8)**
Anteil der Schülerinnen und Schüler ohne sonderpädagogischen Förderbedarf											19.2	(15.7) ^{ns}
Anteil der Schülerinnen und Schüler mit Migrationshintergrund (ein Elternteil)											-13.3	(9.7) ^{ns}
Anteil der Schülerinnen und Schüler ohne Migrationshintergrund (kein Elternteil)											11.5	(4.6)**
Erklärte Varianzanteile												
Zwischen den Schulen (24.6%)	0.4		0.4		0.4		0.4		0.4		72.7	
Innerhalb der Schulen (75.3%)	0.0		-1.0		9.1		31.9		32.0		32.0	
Gesamt	0.1		-0.6		6.9		24.1		24.2		42.1	

Anmerkungen: MGH = Migrationshintergrund; SPF = sonderpädagogischer Förderbedarf; T11 = TIMSS 2011; T15 = TIMSS 2015; T19 = TIMSS 2019

b = Regressionsgewicht (unstandardisiert)

Signifikanzniveau: ns = nicht signifikant; * = signifikant ($p \leq .05$); ** = signifikant ($p \leq .01$)

A = Testzeitpunkt (0 = TIMSS 2007; 1 = TIMSS 2011; 0 = TIMSS 2015; 0 = TIMSS 2019)

B = Testzeitpunkt (0 = TIMSS 2007; 0 = TIMSS 2011; 1 = TIMSS 2015; 0 = TIMSS 2019)

C = Testzeitpunkt (0 = TIMSS 2007; 0 = TIMSS 2011; 0 = TIMSS 2015; 1 = TIMSS 2019)

D = Geschlecht (0 = Mädchen; 1 = Jungen)

E = Migrationshintergrund nach Geburtsland der Eltern (0 = beide Elternteile im Ausland geboren; 1 = ein Elternteil im Ausland geboren; 0 = kein Elternteil im Ausland geboren)

F = Migrationshintergrund nach Geburtsland der Eltern (0 = beide Elternteile im Ausland geboren; 0 = ein Elternteil im Ausland geboren; 1 = kein Elternteil im Ausland geboren)

G = Sonderpädagogischer Förderbedarf (0 = sonderpädagogischer Förderbedarf; 1 = kein sonderpädagogischer Förderbedarf)

H = Sozioökonomischer Status der Familie: Sozialindex (unter der Annahme der Messinvarianz seit TIMSS 2007 skaliert)

nommen. In Modell 4 wird zusätzlich um den sozioökonomischen Status und die kognitive Fähigkeiten kontrolliert. In Modell 5 werden durch Aufnahme von Interaktionen Veränderungen der Anteile von Kindern mit Migrationshintergrund, dem Geschlecht sowie Kindern mit besonderen Unterstützungsbedarfen über die Zeit berücksichtigt. In Modell 6 wird zusätzlich um Merkmale der Klassenzusammensetzung kontrolliert.

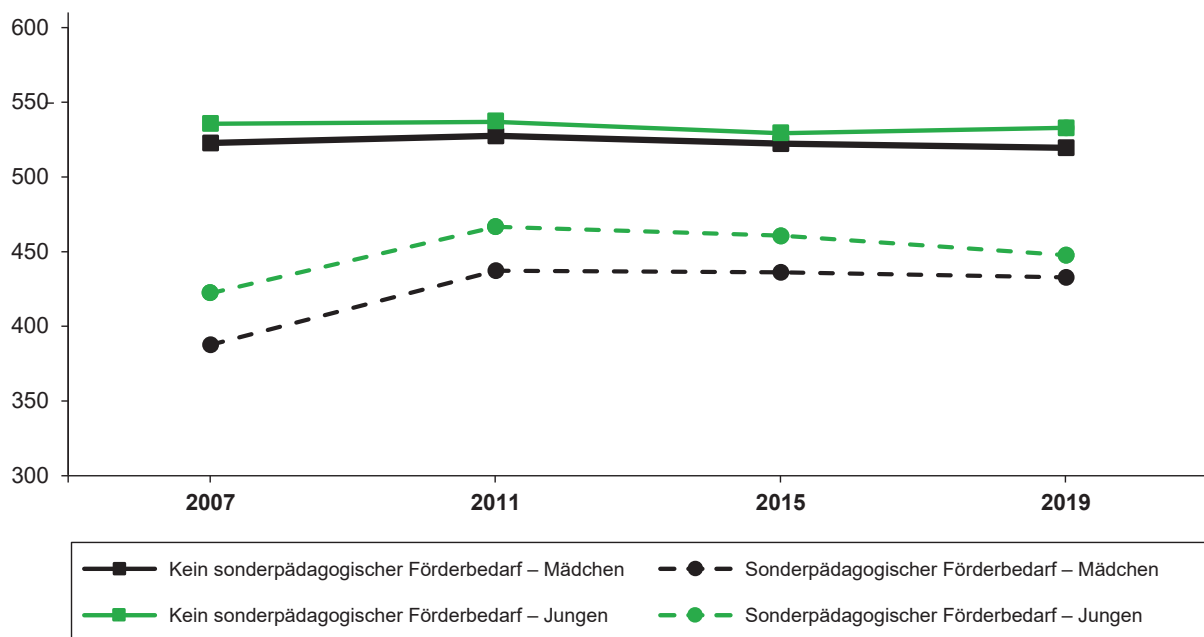
Als Ergebnis der Modelle 5 und 6 ist zunächst festzuhalten, dass mithilfe der hier genutzten Indikatoren 31.1 beziehungsweise 32 Prozent aller Leistungsunterschiede zwischen den Schülerinnen und Schülern innerhalb der Schulen erklärt werden können. In Bezug auf die Unterschiede zwischen den Studienzyklen ist sowohl für Mathematik als auch die Naturwissenschaften festzustellen, dass unter Berücksichtigung der Veränderung relevanter Merkmale der Schülerschaft kein signifikanter Unterschied zwischen TIMSS 2007 und 2019 zu beobachten ist. Die Leistungen bleiben in Mathematik stabil und die allgemein für die Naturwissenschaften festgestellte signifikante Leistungsreduktion von TIMSS 2007 auf 2019 lässt sich auch nach Kontrolle weiterer Schülerinnen- und Schülermerkmale nicht nachweisen. Der in TIMSS 2015 festgestellte positive Leistungstrend in Naturwissenschaften und Mathematik für den Zeitraum TIMSS 2011 zu TIMSS 2015 lässt sich somit für den längeren Zeitraum TIMSS 2007 zu 2019 nicht replizieren.

Um die Interpretation der Ergebnisse aus den Tabellen 12.2 und 12.3 für zentrale Schülerinnen- und Schülergruppen zu erleichtern, werden im Folgenden die Ergebnisse noch einmal getrennt nach Geschlecht, besonderen Unterstützungsbedarf und Migration genauer betrachtet. In Abschnitt 12.4 wird auf die Ergebnisse zur Klassenzusammensetzung aus Modell 6 eingegangen.

12.3.2 Trends in den mittleren Schülerinnen- und Schülerleistungen nach Geschlecht und Unterstützungsbedarfen

Wie in Kapitel 8 in diesem Band beschrieben, lässt sich für Deutschland sowohl für Mathematik als auch für die Naturwissenschaften keine Reduktion der geschlechterspezifischen Leistungsdisparitäten von TIMSS 2007 zu 2019 beobachten. Für TIMSS 2015 konnten Wendt, Steinmayr und Kasper (2016) zeigen, dass sich die von 2007 zu 2015 beobachtete Reduktion der Leistungsdisparitäten für Mathematik auf eine Veränderung der Zusammensetzung der Schülerschaft zurückführen ließ. Für die Naturwissenschaften blieb auch unter Kontrolle von Merkmalen der Schülerzusammensetzung ein signifikanter Leistungsvorsprung der Jungen vor den Mädchen bestehen. Die Ergebnisse wiesen dabei vor allem auf eine Interaktion zwischen dem Geschlecht und dem sonderpädagogischen Förderbedarf hin.

In Abbildung 12.1 sind die Leistungsentwicklungen für Regelschülerinnen und Regelschüler sowie Kinder mit besonderen Unterstützungsbedarfen getrennt nach Geschlecht für die Mathematik abgetragen. Es zeigt sich folgendes Bild: Kinder mit besonderen Unterstützungsbedarfen erzielen in allen Studienzyklen signifikant schlechtere Leistungen in Mathematik. Modell 5 in Tabelle 12.2 ist zu entnehmen, dass dieser besondere Unterstützungsbedarf im Durchschnitt über alle Studienzyklen einen Unterschied von 48.9 Punkten in Mathematik ausmacht. Darüber hinaus ist Abbildung 12.1 zu entnehmen, dass die geschlechterspezifischen Leistungsunterschiede zwischen Kindern mit besonderen Unterstützungsbedarfen insbesondere von 2007 bis 2015 deutlich grö-

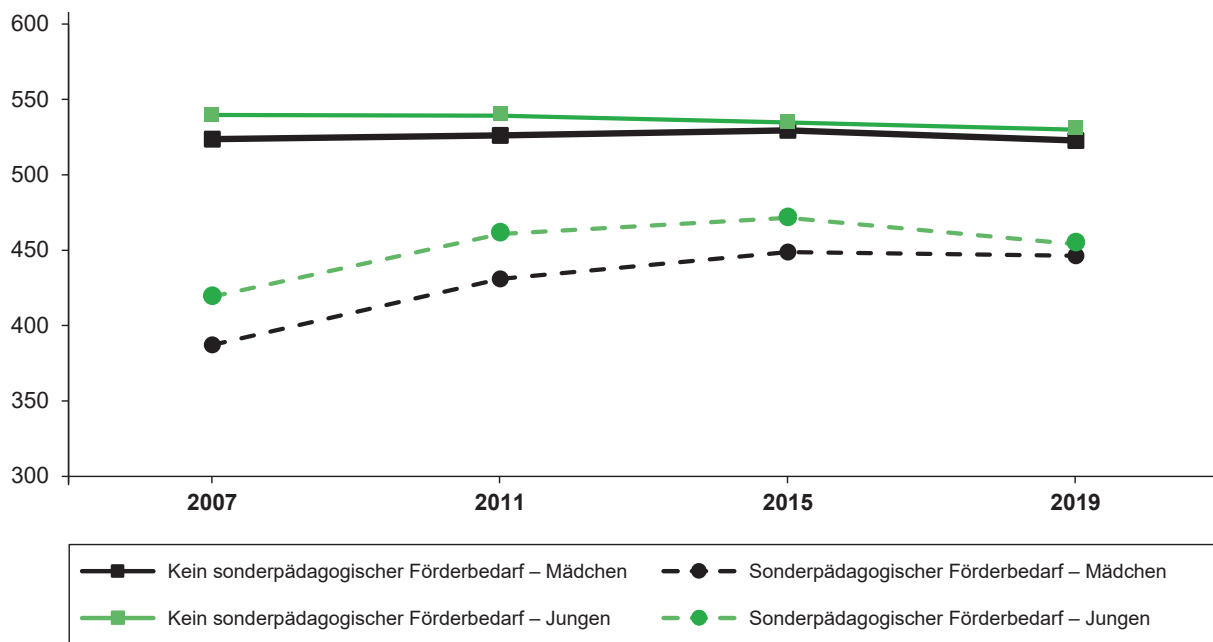
Abbildung 12.1: Veränderungen der Schülerinnen- und Schülerleistungen in Mathematik zwischen TIMSS 2007, 2011, 2015 und 2019 nach Geschlecht und Unterstützungsbedarfen

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

ber ausfallen als zwischen Regelschülerinnen und Regelschülern. Wie den Modellen 5 und 6 in Tabelle 12.2 zu entnehmen ist, sind die entsprechenden Interaktionen zwischen Geschlecht, Kindern mit besonderen Unterstützungsbedarfen und Testzeitpunkt jedoch nicht signifikant, das heißt, dieser differenzielle Verlauf muss dem Zufall zugeschrieben werden. Für die Berechnung der geschlechterbezogenen Leistungsdisparitäten sind hingegen substantielle Leistungsunterschiede zugunsten der Jungen nachzuweisen. Diese betragen im Durchschnitt über alle Studienzyklen rund 19 Leistungspunkte. Ein Trend hinsichtlich der Reduktion der geschlechterspezifischen Leistungsdisparitäten zeigt sich weder für die Regelschülerinnen und Regelschüler noch für die Kinder mit besonderen Unterstützungsbedarfen zwischen 2007 und 2019. Auch unter Berücksichtigung der Veränderungen in der Schülerzusammensetzung über die Zeit (Modelle 5 und 6) sind die geschlechterbezogenen Leistungsdisparitäten weiterhin nachzuweisen.

Die Ergebnisse für die Naturwissenschaften fallen vergleichbar zu denen für die Mathematik aus. Unter Berücksichtigung der Mehrebenenstruktur zeigen sich für alle vier betrachteten Studienzyklen keine signifikanten geschlechterspezifischen Leistungsdisparitäten (Modell 2 in Tabelle 12.3). Unter Berücksichtigung des sonderpädagogischen Förderbedarfs, des Migrationshintergrunds (Modell 3) sowie auch der weiteren Hinzunahme von Indikatoren der sozialen Herkunft, des *Kognitiven Fähigkeitstests* (KFT) sowie veränderten Anteilen in der Schülerzusammensetzung (Modelle 4 bis 6) lassen sich deutliche Leistungsunterschiede zugunsten der Jungen feststellen. Diese betragen im Durchschnitt über alle Studienzyklen hinweg 22 Leistungspunkte. Abbildung 12.2 illustriert die Leistungsentwicklungen im Trend. Die geschlechterspezifischen Leistungsdisparitäten unter den Regelschülerinnen und Regelschülern sind – mit einer kleinen nominellen Reduktion in 2011 – weitestgehend konstant. Für Kinder mit besonderen Unterstützungsbedarfen zeigt sich eine Verbesserung der

Abbildung 12.2: Veränderungen der Schülerinnen- und Schülerleistungen in den Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 nach Geschlecht und Unterstützungsbedarfen

IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

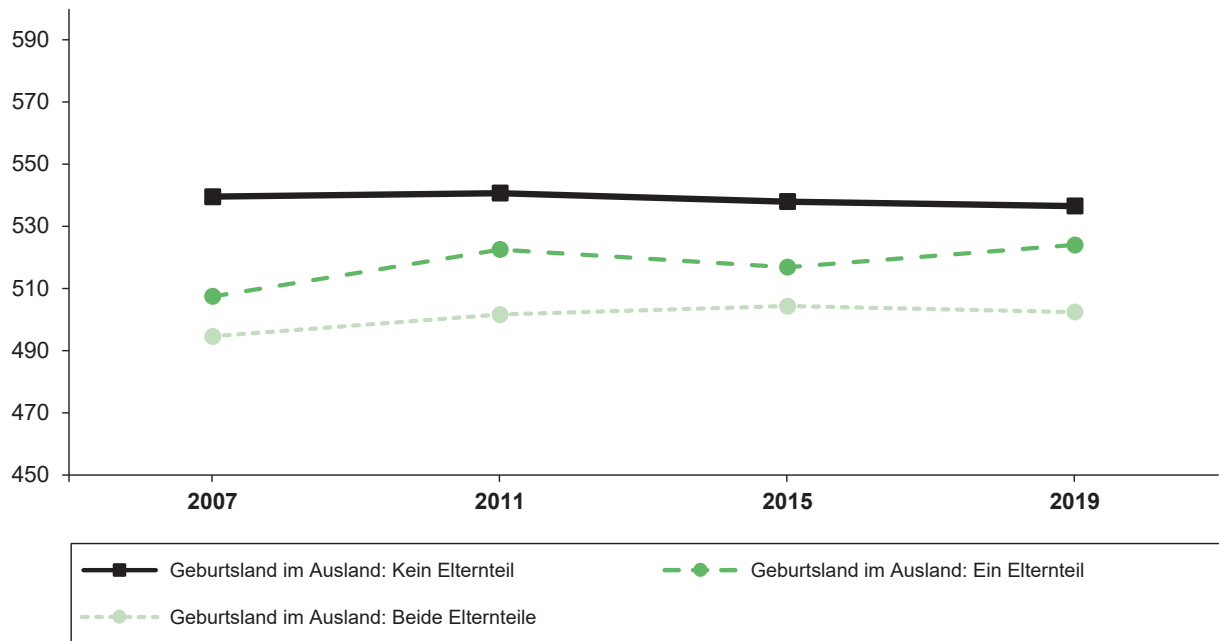
Leistungsmittelwerte, die mit einer Reduktion der Disparitäten einhergeht. Wie die nicht signifikanten Interaktionen zwischen Geschlecht, Kindern mit besonderen Unterstützungsbedarfen und Testzeitpunkt (Modell 5 und 6) zeigen, ist dieser Effekt jedoch nur nominell.

12.3.3 Trends in den mittleren Schülerinnen- und Schülerleistungen nach Migrationshintergrund

Wie in Kapitel 10 in diesem Band beschrieben, lässt sich für Deutschland sowohl für Mathematik als auch für die Naturwissenschaften eine Reduktion der migrationsbezogenen Leistungsdisparitäten von TIMSS 2007 zu 2019 beobachten. In den Abbildungen 12.3 und 12.4 sind diese Ergebnisse noch einmal grafisch illustriert. Für Mathematik (Abbildung 12.3) und Naturwissenschaften (Abbildung 12.4) zeigt sich folgendes Bild: In allen Studienzyklen lassen sich für Kinder mit Migrationshintergrund im Vergleich zu Kindern, deren Eltern beide in Deutschland geboren wurden, signifikant schlechtere Leistungen feststellen. Deutlich wird ein nomineller Trend hinsichtlich der Reduktion migrationspezifischer Leistungsdisparitäten, der sich allerdings bei komplexer Modellierung nicht für alle Studienzyklen statistisch absichern lässt. Kinder mit einem in Deutschland und einem im Ausland geborenen Elternteil verzeichnen eine um 13.1 (Mathematik) beziehungsweise 26.1 (Naturwissenschaften) Punkte höhere Leistung als Kinder mit zwei im Ausland geborenen Elternteilen. Kinder ohne Migrationshintergrund wiederum verzeichnen sogar eine um 28.4 (Mathematik) beziehungsweise 50.7 (Naturwissenschaften) Punkte höhere Leistung gegenüber Kindern mit zwei im Ausland geborenen Elternteilen. Unter weiterer Kontrolle von sozialer Herkunft, KFT sowie Veränderungen in

den Schülerzusammensetzungen (Modell 6) reduzieren sich diese Leistungsdisparitäten. Für Kinder mit einem im Ausland geborenen Elternteil ist der Unterschied zu Kindern mit zwei im Ausland geborenen Elternteilen in Mathematik nicht mehr signifikant. Für die Naturwissenschaften bleiben sie weiterhin bestehen.

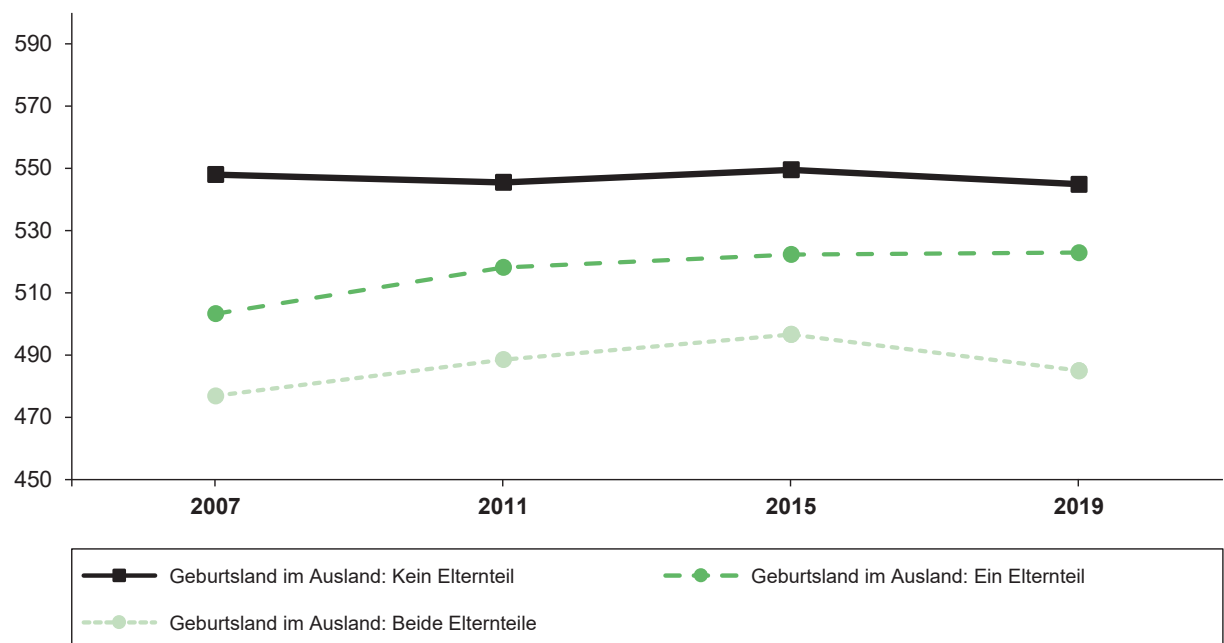
Abbildung 12.3: Veränderungen der Schülerinnen- und Schülerleistungen in Mathematik zwischen TIMSS 2007, 2011, 2015 und 2019 nach Migrationshintergrund



IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

Abbildung 12.4: Veränderungen der Schülerinnen- und Schülerleistungen in den Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 nach Migrationshintergrund



IEA: Trends in International Mathematics and Science Study

© TIMSS 2019

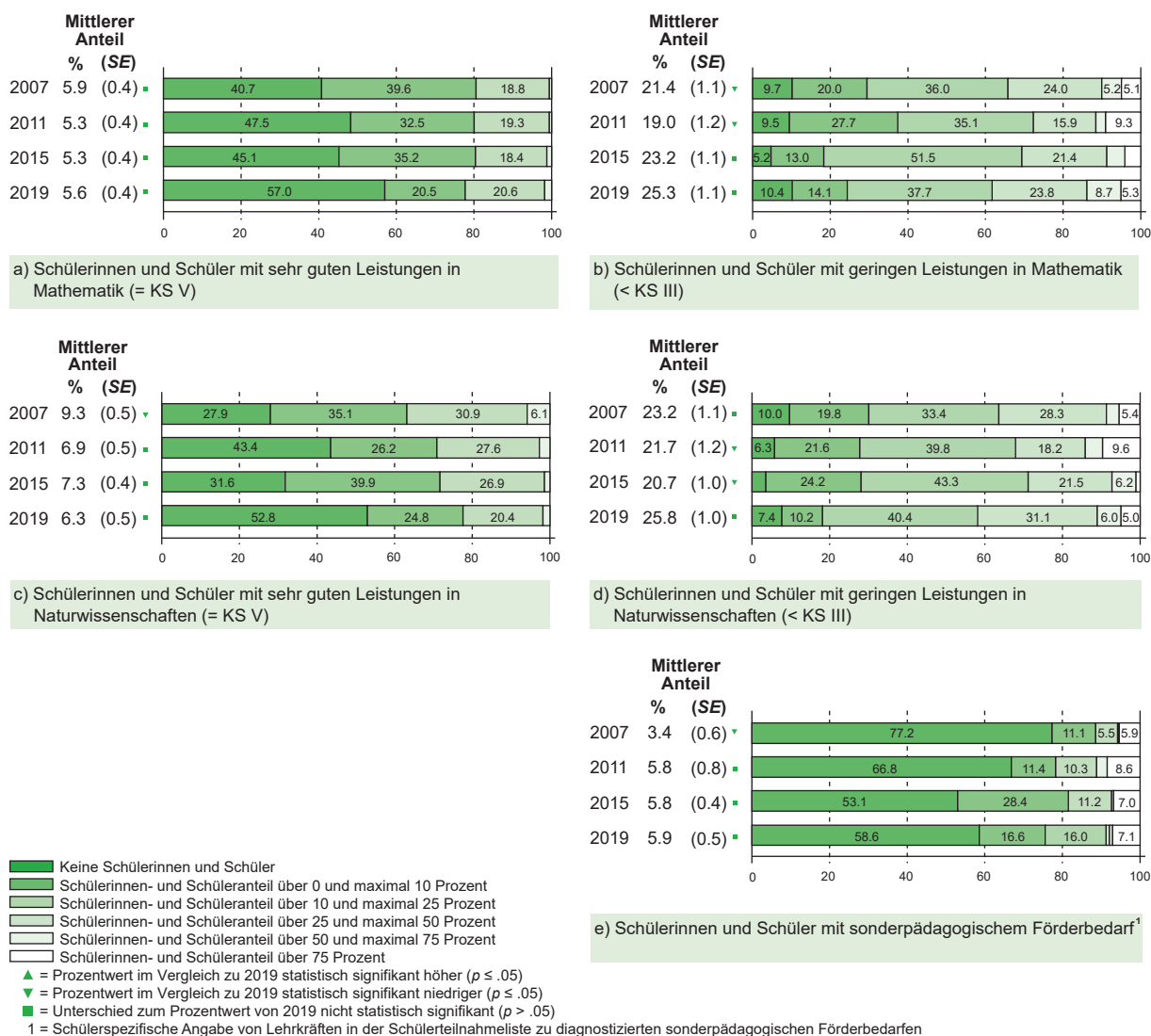
12.4 Veränderungen in den Schülerzusammensetzungen

Wie in den vorangegangenen Abschnitten beschrieben, lassen sich für Grundschulen in Deutschland im Trend sowohl eine veränderte Schülerzusammensetzung mit gestiegenen Anteilen an Kindern mit Migrationshintergrund sowie höheren Anteilen von Kindern mit besonderen Unterstützungsbedarfen als auch eine Zunahme an Leistungsheterogenität beobachten. Anlass zur Sorge bieten die gestiegenen Anteile an Kindern, die Leistungen auf dem Niveau der Kompetenzstufen I und II zeigen. Soll der gewachsenen Heterogenität Rechnung getragen werden, so gilt es, den Unterricht adaptiver beziehungsweise individualisierter zu gestalten. Wer adaptiv lehrt, ist fähig, den „Unterricht so auf die individuellen Voraussetzungen der Schülerinnen und Schüler auszurichten und während des Unterrichts laufend anzupassen, dass für möglichst viele Schülerinnen und Schüler günstige Bedingungen für das Erreichen der Lernziele geschaffen werden“ (Brühwiler, 2014, S. 74). Der damit formulierte Anspruch ist hoch und ob beziehungsweise in welchem Maß er zu erfüllen ist, hängt nicht zuletzt von der Zusammensetzung der jeweiligen Klasse ab. Zudem ist zu fragen, an welchem Anteil an Schulen sich bedeutsame Veränderungen in den Schülerzusammensetzungen ergeben haben.

In diesem Zusammenhang ist es auch von besonderem Interesse, die erhebliche Segregation von Schulen deskriptiv zu analysieren. Die Zuweisung von Schülerinnen und Schülern zu Grundschulen erfolgt nicht zufällig, vielmehr ist nach wie vor die geografische Nähe der Schule zur Wohnung der Familie das entscheidende Kriterium, auf welche Schule ein Kind geht. Insbesondere Großstädte sind heute stark sozial segregiert. Dort verteilt sich die Bevölkerung mit Migrationshintergrund auch nicht zufällig über alle Stadtteile, sondern konzentriert sich regional (Bonsen et al., 2010). Als Folge solcher Segregationsprozesse und der Kopplung von Nachbarschaft und gewählter Schule ist zu erwarten, dass sich die Kompositionen von Klassen beziehungsweise Schulen erheblich unterscheiden, manche Schulen dementsprechend eine überwiegend benachteiligte Schülerschaft aufnehmen, andere eine privilegiert. Bremerich-Vos, Wendt und Hußmann (2017) konnten auf Basis der Daten von IGLU 2016 bereits zeigen, dass die Schülerzusammensetzungen zwischen Grundschulen sich erheblich unterscheiden. In den Abbildungen 12.5 und 12.6 werden die Klassenzusammensetzungen an Grundschulen in Deutschland nach verschiedenen Diversitätsmerkmalen im Vergleich von TIMSS 2007, 2011, 2015 und 2019 als Merkmale der Schulen betrachtet.

Leistungsstarke Schülerinnen und Schüler: Abbildung 12.5 ist zu entnehmen, dass 57 Prozent der Grundschulen in 2019 keine Schülerinnen und Schüler hatten, die eine Leistung auf dem höchsten Niveau (Kompetenzstufe V) in Mathematik erreichten. An weiteren 20,5 Prozent der Schulen liegt der Anteil dieser besonders leistungsstarken Kinder bei weniger als 10 Prozent und an etwa 23 Prozent der Schulen liegt der Anteil dieser Kinder zwischen 10 und 25 Prozent oder darüber. Im Vergleich der Studienzyklen fällt auf, dass es immer weniger Schulen gelingt, leistungsstarke Kinder so zu fördern, dass sie herausragende Leistungen erzielen. Im Jahr 2007 waren es noch rund 59 Prozent aller Grundschulen, an denen auch herausragende Leistungen in Mathematik festgestellt werden konnten, im Jahr 2019 hingegen nur noch 43 Prozent. Gleichzeitig ist Abbildung 12.5 auch zu entnehmen, dass der Anteil von Grundschulen, an denen mehr als 10 Prozent der Schülerinnen und Schüler sehr gute Leistungen zeigen, leicht steigt. Für die

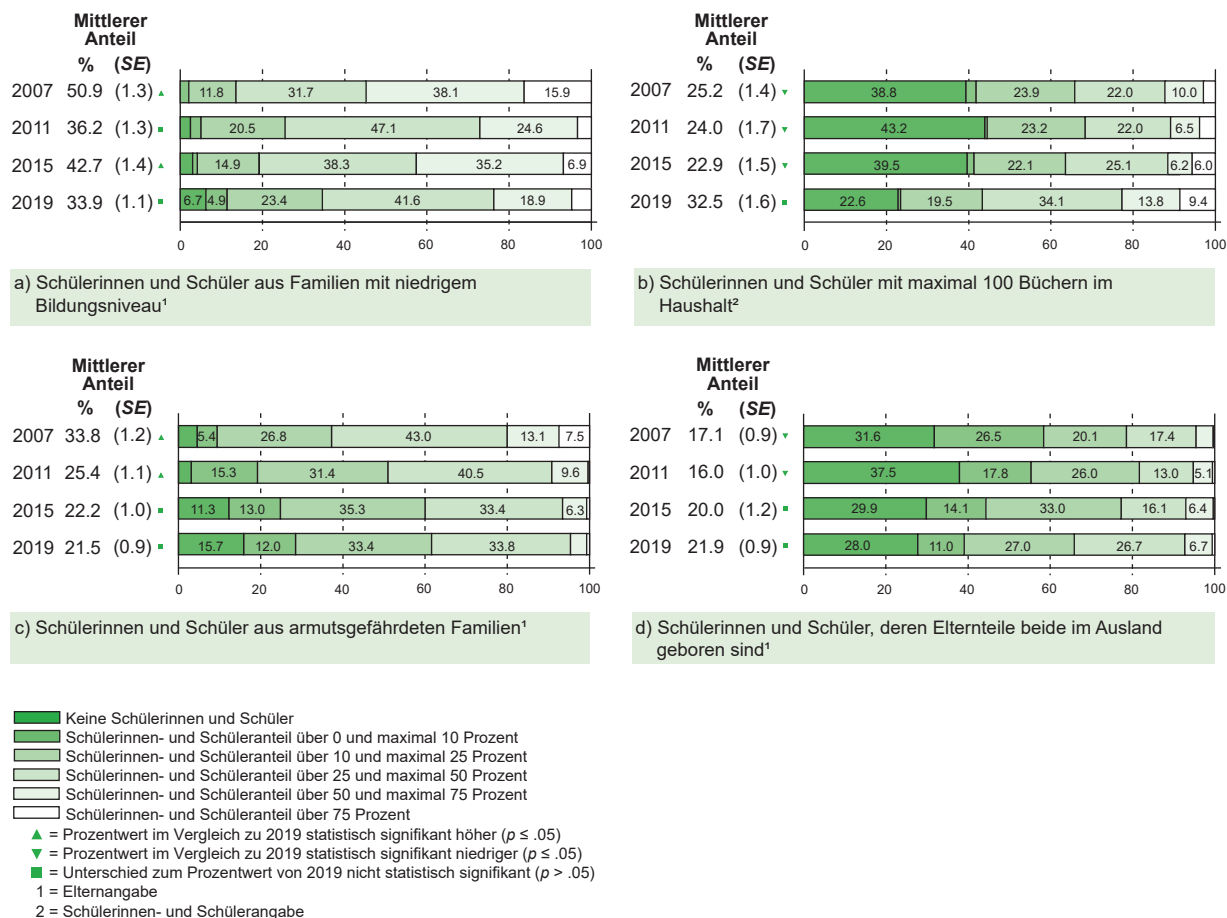
Abbildung 12.5: Mittlere Anteile von Schülerinnen und Schülern nach unterschiedlichen Diversitätsmerkmalen in Klassen der vierten Jahrgangsstufe an Regelschulen in Deutschland im Vergleich von TIMSS 2007, 2011, 2015 und 2019 (Schulen nach Zusammensetzungsmerkmalen in Prozent) I



Naturwissenschaften ist ebenfalls festzustellen, dass es in 2019 im Vergleich zu 2007 weniger Grundschulen gibt, an denen Schülerinnen und Schüler herausragende Leistungen auf der Kompetenzstufe V zeigen. Verglichen mit 2007 ist der Anteil an Schulen, an denen mindestens eine Schülerin oder ein Schüler Leistungen auf der Kompetenzstufe V zeigt, signifikant von rund 72 Prozent aller Grundschulen auf 47 Prozent aller Grundschulen in 2019 gesunken. Der Anteil an Schulen, bei denen mindestens 10 Prozent der Schülerinnen und Schüler Leistungen auf der Kompetenzstufe V zeigen, sinkt im selben Zeitraum von 6.1 Prozent auf rund 2 Prozent.

Leistungsschwache Schülerinnen und Schüler: Für 2019 ist zunächst für die Mathematik festzustellen: In 10.4 Prozent der Schulen finden sich keine Kinder mit Mathematikleistungen unterhalb der Kompetenzstufe III. In weiteren 14 Prozent aller Grundschulen beträgt ihr Anteil maximal 10 Prozent, in 38 Prozent der Schulen liegt der Anteil dieser Kinder zwischen 10 und 25 Prozent

Abbildung 12.6: Mittlere Anteile von Schülerinnen und Schülern nach unterschiedlichen Diversitätsmerkmalen in Klassen der vierten Jahrgangsstufe an Regelschulen in Deutschland im Vergleich von TIMSS 2007, 2011, 2015 und 2019 (Schulen nach Zusammensetzungsmerkmalen in Prozent) II



und in weiteren 24 Prozent der Schulen liegt der Anteil der leistungsschwachen Schülerinnen und Schüler zwischen 25 und 50 Prozent. In den verbleibenden 14 Prozent der Grundschulen kommt mindestens die Hälfte der Viertklässlerinnen und Viertklässler über die beiden unteren Kompetenzstufen nicht hinaus. Im Vergleich zu 2007 ist der Anteil von Schulen ohne oder mit sehr wenigen leistungsschwachen Kindern in Mathematik (weniger als 10%) von knapp 30 Prozent aller Grundschulen auf nunmehr knapp 25 Prozent leicht gesunken. Der Anteil an Klassen mit besonders herausfordernden Schülerzusammensetzungen, mit großen Anteilen leistungsschwacher Schülerinnen und Schüler, hat sich nur geringfügig verändert.

Für die naturwissenschaftlichen Leistungen zeigt sich, dass 2019 in 7.4 Prozent der Schulen kein Kind niedrige naturwissenschaftliche Kompetenzen auf dem Niveau unterhalb der Kompetenzstufe III zeigt. In weiteren 10 Prozent aller Grundschulen beträgt ihr Anteil maximal 10 Prozent, in 40 Prozent der Schulen zwischen 10 und 25 Prozent, in weiteren 31.1 Prozent der Schulen liegt der Anteil der Kinder zwischen 25 und 50 Prozent. In etwas mehr als 11 Prozent aller Grundschulen kommt mindestens die Hälfte der Viertklässlerinnen und Viertklässler über die beiden unteren Kompetenzstufen nicht hinaus. Im Vergleich zu 2007 haben sich die Zusammensetzungen etwas verschoben.

Kinder mit diagnostiziertem sonderpädagogischen Förderbedarf: In 2019 finden sich in mehr als der Hälfte (59%) der Schulen keine Kinder mit diagnostiziertem sonderpädagogischen Förderbedarf. In 2007 lag dieser Anteil noch bei 77 Prozent. Aufgrund der Stichprobeziehung (in der Regel nur eine Klasse pro Grundschule) und besonderen Studienteilnahmebedingungen ist zu vermuten, dass der in Tabelle 12.1 berichtete Anteil etwas unterschätzt wird. An vielen Schulen werden zudem auch Kinder sonderpädagogisch gefördert, ohne dass für sie ein entsprechendes Gutachten vorliegt. Für TIMSS 2015 konnte festgestellt werden, dass etwa 12 Prozent aller Grundschülerinnen und Grundschüler sonderpädagogisch gefördert wurden, obwohl nur für knapp 6 Prozent ein diagnostizierter Förderbedarf vorlag (Wendt et al., 2017).

Differenziert nach Merkmalen der sozialen Herkunft der Schülerinnen und Schüler (siehe Kapitel 9 in diesem Band) zeigt sich, dass Kinder aus Familien mit niedrigem bis mittlerem Bildungsniveau in fast allen Klassen in Deutschland vertreten sind (siehe Abbildung 12.6). In jeder fünften Klasse stammt mehr als die Hälfte der Kinder aus Familien, bei denen es zu Hause maximal 100 Bücher gibt, aber in jeder vierten Schule macht deren Anteil höchstens 10 Prozent der Kinder aus. Im Trend zeigt sich eine gleichmäßigere Verteilung über die Schulen, es ist aber auch ein gestiegener Anteil an Grundschulen, in denen mindestens jedes zweite Kind von maximal 100 Büchern zu Hause berichtet, zu beobachten. Substanzielle Anteile – das heißt Anteile im Umfang von mehr als 10 Prozent – von Kindern aus armutsgefährdeten Familien gibt es in Dreiviertel aller Klassen. Hier zeigt sich im Vergleich zu 2007 – einhergehend mit den sinkenden Anteilen in Deutschland – eine günstigere Verteilung auf die Grundschulen. Was Kinder mit Zuwanderungshintergrund angeht, so ist migrationsbezogene Diversität 2019 in 28 Prozent aller Grundschulklassen gar nicht zu finden. An weiteren 11 Prozent der Grundschulen macht ihr Anteil höchstens 10 Prozent aus. In 27 Prozent aller Grundschulen liegt ihr Anteil unterdurchschnittlich zwischen 10 Prozent und einem Viertel der Schülerschaft. In rund einem weiteren Drittel von Grundschulen in Deutschland (27%) hat etwa jedes zweite bis vierte Kind Eltern, die nicht in Deutschland geboren wurden. Nur in 6 Prozent der Grundschulen hat mehr als jedes zweite Kind einen Migrationshintergrund.

Die Ergebnisse zeigen deutlich die große Varianz der Schülerzusammensetzungen zwischen Klassen beziehungsweise Schulen in Deutschland. Berücksichtigt man die hier genannten Dimensionen von Heterogenität, wird deutlich, dass die Herausforderungen im Hinblick auf adaptives Unterrichten je nach Zusammensetzung der Klasse vor Ort stark variieren können. Forschungsergebnisse sprechen dafür, dass die Komposition der Klasse nicht nur die Leistungen der einzelnen Schülerinnen und Schüler beeinflusst, sondern auch Auswirkungen zum Beispiel auf das Selbstkonzept, die Motivation und die Lernfreude hat (Dumont, Neumann, Maaz & Trautwein, 2013; Lintorf, Guill & Wendt, 2017). Auch für den Grundschulbereich liegen mittlerweile einige Studien zu Zusammenhängen zwischen Merkmalen der Klassenzusammensetzung und den mathematischen und naturwissenschaftlichen Kompetenzen am Ende der vierten Jahrgangsstufe vor (für einen Überblick: Vennemann, 2018). Aufgrund des querschnittlichen Designs besitzen viele Studien allerdings nur begrenzte Aussagekraft, können aber bei mehrbenenanalytischer Modellierung bedeutsame Zusammenhänge zwischen unterschiedlichen Merkmalen der Komposition und Leistung feststellen. Vor diesem Hintergrund lassen sich die Ergebnisse der Modelle 6 aus den Tabellen 12.2 und 12.3 interpretieren. In den Modellen 6 wurden auf Schulebene zusätzlich zu den Merkmalen aus den Modellen 5 Merkmale

der Schülerzusammensetzung mitaufgenommen. Es zeigt sich, dass der Anteil an Kindern mit besonderen Unterstützungsbedarfen sowie der Anteil an Kindern mit einem in Deutschland und einem im Ausland geborenen Elternteil keine Zusammenhänge mit den durchschnittlichen mathematischen oder naturwissenschaftlichen Kompetenzen in den Klassen aufweisen. Hingegen zeigt sich für die mittlere soziale Zusammensetzung ein signifikanter Zusammenhang. Auch zeigt sich die migrationspezifische Zusammensetzung als bedeutsam: Für die Interpretation ist wichtig zu berücksichtigen, dass hier Klassen mit minimalen (also keinen Kindern mit Migrationshintergrund) und maximalen Anteilen (100% der Kinder haben Eltern, die beide nicht in Deutschland geboren wurden) verglichen werden. In Klassen ohne Kinder mit Migrationshintergrund im Vergleich zu denen mit maximaler Zusammensetzung zeigen Kinder im Durchschnitt – unabhängig von ihren Individualmerkmalen – minimal bessere Leistungen in der Größenordnung von 10 beziehungsweise 12 Punkten. Längsschnittuntersuchungen, die für den Grundschulbereich vorliegen und zusätzlich auch für die Betrachtung von Kompositionseffekten Prozessmerkmale (z. B. Unterrichtsmerkmale) berücksichtigen (van Ophuysen & Wendt, 2009; Venne-mann, 2018), zeigen, dass die Leistungsunterschiede zwischen Schülerinnen und Schülern zu deutlich größeren Anteilen durch Prozessmerkmale erklärt werden können. Die Befunde verdeutlichen damit, dass es vor allem auf einen angemessenen Umgang mit Heterogenität ankommt (Morris-Lange, Wendt & Wohlfahrt, 2013).

12.5 Zusammenfassung

Die Schülerschaft an Grundschulen in Deutschland ist im Jahr 2019 etwas mehr von Vielfalt geprägt, als dies im Jahr 2007 der Fall war. In Folge struktureller Reformbemühungen, inklusive Beschulungsangebote zu schaffen, ist ein Anstieg des Anteils von Kindern mit sonderpädagogischen Förderbedarfen an Regelschulen festzustellen. Auch hat sich die Schülerschaft hinsichtlich ihrer soziodemografischen Zusammensetzung, vor allem in Bezug auf die migrationsbezogenen Merkmale der Schülerinnen- und Schülerfamilien, verändert. In diesem Kapitel wurde anhand von Trendanalysen vertiefend untersucht, welchen Zusammenhang diese Entwicklung mit den mittleren Schülerinnen- und Schülerleistungen in Deutschland von TIMSS 2007 zu 2011, 2015 und 2019 aufweist.

Betrachtet man die Leistungsergebnisse im Trend und insbesondere im Vergleich der Resultate zwischen TIMSS 2007 und 2019, so ergibt sich für Deutschland im Bereich Mathematik keine statistisch relevante Veränderung. Im Jahr 2011 gab es zwar ein kleines ‚Zwischenhoch‘, es verschwand aber bereits im Jahr 2015 wieder. Dieser Befund bleibt auch bestehen, wenn man die Veränderungen in der Zusammensetzung der Schülerschaft berücksichtigt. Für die Naturwissenschaften war der Leistungsmittelwert über zehn Jahre konstant bei 528 Leistungspunkten. Für 2019 ließ sich erstmals ein um 10 Punkte signifikant niedriger Leistungsmittelwert feststellen. Dieser Unterschied ist inhaltlich als eher gering zu interpretieren (siehe Kapitel 4 in diesem Band) und die Analysen in diesem Kapitel zeigen, dass sich der Differenzkoeffizient bereits bei komplexerer statistischer Modellierung, die das besondere Erhebungsdesign der Studie berücksichtigt, nicht mehr zufallskritisch absichern lässt. Die Analysen in diesem Kapitel zeigen deutlich, dass Veränderungen in der Schülerzusammensetzung an Grundschulen nicht als hauptsächliches Erklärungsmuster für diese Stagnation der

Schülerinnen- und Schülerleistungen am Ende der Grundschulzeit in Deutschland herangezogen werden können. Auch unter Berücksichtigung gestiegener Anteile von Schülerinnen und Schülern mit besonderen Unterstützungsbedarfen oder Migrationshintergrund bleibt die Befundlage konstant. Zudem zeigt die Analyse veränderter Schülerzusammensetzungen an Grundschulen über die Zeit, dass sich nur maximal an jeder fünften Grundschule in Deutschland Veränderungen in der Schülerzusammensetzung ergeben haben, die von besonderer praktischer Relevanz sind. Um zu verhindern, dass die aktuellen Ergebnisse sich tatsächlich zu einem negativen Trend entwickeln, ist Handlungsbedarf geboten. Insbesondere an Grundschulen, an denen es nicht gelingt, zumindest einer deutlichen Mehrheit der Schülerinnen und Schüler mathematische und naturwissenschaftliche Kompetenzen auf dem Niveau der Kompetenzstufe III zu vermitteln, sollten Lehrkräfte beispielsweise durch zusätzliche Ressourcen wie etwa Personal oder Reduzierung von Mehrbelastungen unterstützt werden. Dies betrifft in etwa jede vierte Grundschule in Deutschland. Auch wenn es geboten ist, Schülerinnen und Schüler mit ungünstigen Voraussetzungen gezielt zu fördern, sind dabei die leistungsstarken Schülerinnen und Schüler nicht zu vernachlässigen. Auffällig ist vor allem der Befund, dass es 2019 im Vergleich zu 2007 in immer weniger Schulen gelingt, gute Schülerinnen und Schüler so zu fördern, dass sie auch herausragende Leistungen auf dem Niveau der Kompetenzstufe V zeigen können. Waren Kinder mit sehr guten Leistungen in 2007 noch an 60 beziehungsweise 70 Prozent aller Grundschulen zu finden, ist es in 2019 nicht mal mehr jede zweite Grundschule. Ein kompetenter Umgang mit Heterogenität ist leicht zu postulieren, aber nur schwer in der Praxis umzusetzen. Die in diesem Kapitel berichteten Ergebnisse zeigen deutlich, dass Herausforderungen im Hinblick auf adaptives Unterrichten je nach Zusammensetzung der Klasse stark variieren.

Einschränkend ist zu erwähnen, dass in diesem Kapitel lediglich einige zentrale Veränderungen in der Zusammensetzung der Schülerschaft betrachtet wurden. Damit bleibt unklar, inwiefern die gefundenen Trendverläufe auch für Schülerinnen- und Schülergruppen gelten, die hier nicht abgebildet oder differenziert erfasst wurden. Beachtenswert ist, dass sich bedeutsame Leistungstrends vor allem für jene Gruppen beobachten lassen, deren Charakteristika immer schon sehr heterogen waren und sich im Zuge gesellschaftlicher Wandlungsprozesse deutlich verändert haben (wie beispielsweise Schülerinnen und Schüler mit einem in Deutschland und einem im Ausland geborenen Elternteil). Die Frage, ob die Nutzung von übergeordneten kategorialen Unterscheidungen von Kindern in bestimmte Teilgruppen (beispielsweise mit und ohne Migrationshintergrund) auch im Trend sinnvoll ist, wäre in vertiefenden Analysen zu untersuchen. Auch muss bedacht werden, dass in den vorgestellten Modellen die Veränderung in der Zusammensetzung der Schülerschaft in der Grundgesamtheit über die Zeit untersucht wurde. Denkbar ist, dass sich diese Veränderung unterschiedlich stark in der Schülerschaft einzelner Schulen dargestellt hat, Schulen also in diesem Sinne differenzielle Wandlungsprozesse erfahren haben (z.B. unterschiedliche Entwicklungen in städtischen oder ländlichen Gebieten). Ob diese Wandlungsprozesse auch zu heterogenen Leistungsgradienten beigetragen haben, wäre ebenfalls in vertiefenden Analysen zu untersuchen.

Literatur

- Autorengruppe Bildungsberichterstattung. (2016). *Bildung in Deutschland 2016. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Verfügbar unter: <http://www.bildungsbericht.de/de/bildungsberichte-seit-2006/bildungsbericht-2016/pdf-bildungsbericht-2016/bildungsbericht-2016>
- Autorengruppe Bildungsberichterstattung. (2020). *Bildung in Deutschland 2020. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt*. Verfügbar unter: https://www.bildungsbericht.de/static_pdfs/bildungsbericht-2020.pdf
- Bonsen, M., Bos, W., Gröhlich, C., Harney, B., Imhäuser, K., Makles, A., Schräpler, J., Terpoorten, T., Weishaupt, H. & Wendt, H. (Hrsg.) (2010). *Zur Konstruktion von Sozialindizes. Ein Beitrag zur Analyse sozialräumlicher Benachteiligung von Schulen als Voraussetzung für qualitative Schulentwicklung* (Bildungsforschung, Bd. 31). Bonn: BMBF.
- Bremerich-Vos, A., Wendt, H. & Bos, W. (2017). Lesekompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 79–143). Münster: Waxmann.
- Bremerich-Vos, A., Wendt, H. & Hußmann, A. (2017). Bausteine adaptiven Leseunterrichts angesichts gewachsener Heterogenität. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Hrsg.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 317–330). Münster: Waxmann.
- Brühwiler, C. (2014). *Adaptive Lehrkompetenz und schulisches Lernen*. Münster: Waxmann.
- Dumont, H., Neumann, N., Maaz, K. & Trautwein, U. (2013). Die Zusammensetzung der Schülerschaft als Einflussfaktor für Schulleistungen: Internationale und nationale Befunde. *Psychologie in Erziehung und Unterricht*, 60(3), 163–183. <https://doi.org/10.2378/peu2013.art14d>
- Kasper, D., Wendt, H., Bos, W. & Köller, O. (2016). Trends in mathematischen und naturwissenschaftlichen Kompetenzen am Ende der Grundschulzeit in Deutschland. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 367–382). Münster: Waxmann.
- Lintorf, K., Guill, K. & Wendt, H. (2017). ‚Was dein Kind kann, kann meins schon lange!‘. Effekte der sozialen Komposition von Grundschulklassen auf den Übergang. *Zeitschrift für Erziehungswissenschaft*, 20(4), 711–727. <https://doi.org/10.1007/s11618-017-0762-2>
- McCulloch, C. E., Searle, S. R. & Neuhaus, J. M. (2008). *Generalized, linear, and mixed models*. Hoboken: Wiley.
- Morris-Lange, S., Wendt, H. & Wohlfarth, C. (2013). *Segregation an deutschen Schulen. Ausmaß, Folgen und Handlungsempfehlungen für bessere Bildungschancen*. Berlin: SVR GmbH.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. London: Sage Publication.
- Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S. & Haag, N. (2017). (Hrsg.). *IQB Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Münster: Waxmann.
- Statistisches Bundesamt. (2017). *Bildung und Kultur. Allgemeinbildende Schulen* (Fachserie 11, Reihe 1, 2016/2017). Verfügbar unter: https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00033325/2110100177004_korr08122017.pdf
- van Ophuysen, S. & Wendt, H. (2009). Zur Veränderung der Mathematikleistung von Klasse 4 bis 6. Welchen Einfluss haben Kompositions- und Unterrichtsmerkmale? *Zeitschrift für Erziehungswissenschaft*, 12, 302–327. https://doi.org/10.1007/978-3-531-92216-4_13
- Vennemann, M. (2018). *Individual- und Kompositionseffekte und der Kompetenzzuwachs in Mathematik und Naturwissenschaft am Ende der Grundschule*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-23506-2>

- Wendt, H., Steinmayr, R. & Kasper, D. (2016). Geschlechterunterschiede in mathematischen und naturwissenschaftlichen Kompetenzen. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert & D. Kasper (Hrsg.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 257–297). Münster: Waxmann.
- Wendt, H., Bos, W., Goy, M. & Jusufi, D. (2017). *TIMSS 2015 – Skalenhandbuch zur Dokumentation der Erhebungsinstrumente und Arbeit mit den Datensätzen*. Münster: Waxmann.

Anhang

Tabelle 1: Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2007

TIMSS 2007		Besonderheiten bzgl. der nationalen Zielpopulation				Besonderheiten bzgl. der nationalen Schul- und Schülerteilnahmequoten				
		Getestete Jahrgangsstufe	Durchschnittsalter**	Aus-schöpfungsgrad***	Aus-schlüsse****	Schulteilnahmequote in %		Schülerteilnahmequote in %	Gesamtteilnahmequote in %	
						ohne Ersatzschulen	mit Ersatzschulen		ohne Ersatzschulen	mit Ersatzschulen
Teilnehmer										
	Algerien	4	10.2	100	2.1	99	99	97	97	97
	Armenien	4	10.6	100	3.4	93	100	96	90	96
	Australien	4	9.9	100	4.0	99	100	95	94	95
³	Dänemark	4	11.0	100	4.1	71	91	94	66	85
	Deutschland	4	10.4	100	1.3	96	100	97	93	96
	El Salvador									
¹ ³	England ^A	5	10.2	100	2.1	83	90	93	77	84
²	Georgien	4	10.1	85	4.8	92	100	98	90	98
² ³	Hongkong	4	10.2	100	5.4	81	84	96	78	81
	Iran	4	10.2	100	3.0	100	100	99	99	99
²	Italien	4	9.8	100	5.3	91	100	97	88	97
	Japan	4	10.5	100	1.1	97	99	97	94	95
	Jemen	4	11.2	100	2.0	99	100	98	97	98
²	Kasachstan	4	10.6	94	5.3	99	100	100	99	100
	Katar	4	9.7	100	1.8	100	100	97	97	97
	Kolumbien									
⁶	Kuwait	4	10.2	100	0.0	100	100	85	85	85
²	Lettland			72	4.6	93	97	95	89	92
²	Litauen	4	10.8	93	5.4	99	100	94	93	94
³	Marokko	4	10.6	100	1.4	81	81	96	77	77
¹ ²	Neuseeland ^A	4.5-5.5	10.0	100	5.4	97	100	96	93	96
³	Niederlande	4	10.2	100	4.8	48	95	97	46	91
²	Norwegen	4	9.8	100	5.1	88	97	95	83	92
	Österreich	4	10.3	100	5.0	98	99	98	96	97
	Russische Föderation	4	10.8	100	3.6	100	100	98	98	98
¹ ³	Schottland ^A	5	9.8	100	4.5	77	94	94	72	88
	Schweden	4	10.8	100	3.1	98	100	97	94	97
	Singapur	4	10.4	100	1.5	100	100	96	96	96
	Slowakei	4	10.4	100	3.3	98	100	97	95	97
	Slowenien	4	9.8	100	2.1	92	99	95	87	93
	Taiwan	4	10.2	100	2.8	100	100	100	100	100
	Tschechische Republik	4	10.3	100	4.9	89	98	94	83	92
	Tunesien	4	10.2	100	2.9	100	100	99	99	99
	Ukraine	4	10.3	100	0.6	96	96	97	93	93
	Ungarn	4	10.7	100	4.4	93	99	97	90	96
² ³	USA	4	10.3	100	9.2	70	89	95	66	84
Benchmark-Teilnehmer										
²	Alberta, Kanada	4	9.8	100	7.6	99	99	96	94	94
²	Britisch-Kolumbien, Kanada	4	9.8	100	9.2	98	100	96	94	96
² ³ ⁶	Dubai, VAE	4	10.0	100	5.4	75	75	91	67	67
²	Massachusetts, USA	4	10.3	100	10.4	92	96	96	88	92
² ³	Minnesota, USA	4	10.3	100	8.3	53	100	97	52	97
²	Ontario, Kanada	4	9.8	100	6.3	95	96	95	91	92
²	Québec, Kanada	4	10.1	100	6.4	97	98	86	83	84

Tabelle 2: Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2011




TIMSS 2011	Teilnehmer	Teilnahme an TIMSS 2007*	Besonderheiten bzgl. der nationalen Zielpopulation				Besonderheiten bzgl. der nationalen Schul- und Schülerteilnahmequoten				Besonderheiten bzgl. der Leistungsmessung		
			Getestete Jahrgangsstufe	Durchschnittsalter**	Ausschöpfungsgrad***	Ausschlüsse****	Schulteilnahmequote in %		Schülerteilnahmequote in %	Gesamteilnahmequote in %		Anteil der Schülerinnen und Schüler ohne gültige Leistungswerte in %	
							ohne Ersatzschulen	mit Ersatzschulen		ohne Ersatzschulen	mit Ersatzschulen	Naturwissenschaften	Mathematik
	Armenien	(✓)	4	10.0	100	2.0	100	100	98	98	98	8	8
^{2 3}	Aserbaidschan		4	10.2	100	7.2	84	100	100	84	100	5	7
	Australien	✓	4	10.0	100	4.4	96	98	95	91	93	2	3
⁶	Bahrain		4	10.4	100	1.1	92	92	98	90	90	5	9
³	Belgien (Fläm. Gem.)		4	10.0	100	5.0	76	95	98	75	92	1	0
	Chile		4	10.1	100	3.7	86	99	96	82	95	3	6
^{2 3}	Dänemark	✓	4	11.0	100	6.3	79	92	95	75	87	1	1
	Deutschland	✓	4	10.4	100	1.9	96	99	96	92	95	1	1
^{1 3}	England ^B	✓	5	10.2	100	2.0	81	83	94	76	78	2	2
	Finnland		4	10.8	100	3.1	97	99	96	93	96	0	1
²	Georgien	✓	4	10.0	92	4.9	97	98	99	95	96	4	9
²	Hongkong	✓	4	10.1	100	8.6	87	88	93	81	82	1	0
	Iran	✓	4	10.2	100	4.5	100	100	99	99	99	5	11
	Irland		4	10.3	100	2.5	97	99	95	93	95	2	2
³	Italien	✓	4	9.7	100	3.7	81	98	97	78	95	1	2
	Japan	✓	4	10.5	100	3.2	96	99	97	93	97	0	0
^{4 5}	Jemen	✓	4	11.2	100	3.7	99	99	97	95	95	39	48
²	Kasachstan	(✓)	4	10.4	100	6.3	99	100	99	98	99	2	3
²	Katar	(✓)	4	10.0	100	6.2	100	100	99	99	99	11	15
²	Kroatien		4	10.7	100	7.9	99	100	95	94	95	1	3
^{2 4 5}	Kuwait	✓	4	9.7	78	0.3	99	99	94	91	91	18	28
²	Litauen	✓	4	10.7	93	5.6	94	100	94	89	94	1	1
¹	Malta ^B		5	9.8	100	3.6	100	100	95	95	95	6	4
^{4 5}	Marokko	(✓)	4	10.5	100	2.0	100	100	97	96	96	28	27
^{1 3 6}	Neuseeland ^B	✓	4.5-5.5	9.9	100	4.9	83	96	94	77	90	2	5
³	Niederlande	✓	4	10.2	100	4.0	49	82	97	47	79	0	0
³	Nordirland		4	10.4	100	3.5	62	85	93	58	79	2	2
³	Norwegen	(✓)	4	9.7	100	4.3	57	82	85	48	70	1	3
⁴	Oman		4	9.9	100	1.5	98	98	98	96	96	13	19
²	Österreich	✓	4	10.3	100	5.1	100	100	98	98	98	0	1
	Polen		4	9.9	100	3.8	100	100	96	96	96	2	4
	Portugal		4	10.0	100	2.5	87	98	94	81	92	1	1
	Republik Korea (Südkorea)		4	10.4	100	2.5	100	100	98	98	98	0	0
	Rumänien		4	10.9	100	4.0	99	100	98	97	97	5	8
²	Russische Föderation	✓	4	10.8	100	5.3	100	100	98	98	98	0	1
	Saudi-Arabien		4	10.0	100	1.6	95	100	99	94	99	6	14
	Schweden	✓	4	10.7	100	4.1	97	99	92	89	91	1	2
²	Serbien	✓	4	10.8	100	9.4	97	100	97	94	97	2	4
²	Singapur	✓	4	10.4	100	6.3	100	100	96	96	96	1	1
	Slowakei	✓	4	10.4	100	4.6	95	99	96	91	96	1	3
	Slowenien	✓	4	9.9	100	2.6	96	97	97	93	94	1	2
²	Spanien	✓	4	9.8	100	5.3	96	99	97	94	97	1	3
	Taiwan	✓	4	10.2	100	1.4	100	100	99	99	99	0	0
	Thailand		4	10.5	100	1.5	85	100	99	84	99	4	6
²	Tschechische Republik	✓	4	10.4	100	5.1	90	99	95	85	94	0	2
^{4 5}	Tunesien	✓	4	10.0	100	2.5	100	100	99	99	99	21	25
	Türkei		4	10.1	100	2.5	97	100	98	95	98	4	6
	Ungarn	✓	4	10.7	100	4.2	98	99	97	94	96	2	4
^{2 3}	USA	✓	4	10.2	100	7.0	79	84	95	76	80	1	1
	Vereinigte Arabische Emirate (VAE)		4	9.8	100	3.3	100	100	97	97	97	7	11
Teilnahme mit Jgst. 6													
¹	Botsuana		6	12.8	100	0.3	100	100	99	99	99	14	11
^{1 4}	Honduras		6	12.7	100	4.5	91	100	97	88	97	6	17
^{1 4}	Jemen		6	13.2	100	4.0	99	99	96	96	96	15	26
Benchmark-Teilnehmer													
	Abu Dhabi, VAE		4	9.7	100	2.7	99	99	98	97	97	8	14
²	Alberta, Kanada	✓	4	9.9	100	7.5	98	99	96	93	95	1	2
²	Dubai, VAE	✓	4	9.8	100	5.1	100	100	96	96	96	5	8
²	Florida, USA		4	10.4	89	12.1	96	96	95	91	91	1	1
²	North Carolina, USA		4	10.2	93	10.1	94	94	95	89	89	1	1
²	Ontario, Kanada	✓	4	9.8	100	5.3	97	98	96	93	94	1	2
	Québec, Kanada	✓	4	10.1	100	3.7	95	96	95	90	91	0	0

Tabelle 3: Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2015

TIMSS 2015		Teilnahme an TIMSS 2011* Teilnahme an TIMSS 2007*		Besonderheiten bzgl. der nationalen Zielpopulation				Besonderheiten bzgl. der nationalen Schul- und Schülerteilnahmequoten ^F						Besonderheiten bzgl. der Leistungs- messung			
				Getestete Jahrgangsstufe	Durchschnittsalter**	Ausschöpfunggrad***	Ausschlüsse****	Schulteilnahme- quote in %		Schülerteilnahmequote in %	Gesamteilnahme- quote in %		Anteil der Schülerinnen und Schüler ohne gültige Leistungswerte in %				
								ohne Ersatz- schulen	mit Ersatz- schulen		ohne Ersatz- schulen	mit Ersatz- schulen	Natur- wissen- schaften	Mathe- matik ^G			
⁶	Armenien ^F	✓	(✓)	4	9.9	--	--	--	--	--	--	--	--	--	--		
	Australien	✓	✓	4	10.0	100	4.2	98	99	95	94	94	2	3			
² ⁷	Bahrain ^{GH}	✓		4	9.9	100	5.6	100	100	100	99	99	99	99	7	5	
³	Belgien (Fläm. Gem.)	✓		4	10.1	100	1.4	74	97	98	73	95	1	1			
	Bulgarien			4	10.8	100	2.9	97	97	96	93	93	2	4			
	Chile	✓		4	10.2	100	3.7	87	94	94	82	88	3	7			
² ³	Dänemark	✓	✓	4	10.9	100	7.5	53	91	95	50	86	2	2			
	Deutschland	✓	✓	4	10.4	100	2.7	97	99	96	93	95	1	2			
¹	England ^C	✓	✓	5	10.1	100	2.3	95	98	98	92	96	1	2			
	Finnland	✓		4	10.8	100	2.0	99	100	97	95	97	1	1			
²	Frankreich			4	9.9	100	5.3	96	99	98	93	97	3	4			
²	Georgien ^E	✓	✓	4	9.7	90	4.9	99	100	98	97	98	5	8			
³	Hongkong	✓	✓	4	10.1	100	2.2	76	82	93	70	76	1	0			
⁷	Indonesien ^{GH}			4	10.4	100	0.2	100	100	100	99	99	99	12	11		
⁷	Iran ^{GH}	✓	✓	4	10.2	100	4.0	100	100	100	99	99	99	11	9		
	Irland	✓		4	10.4	100	2.7	100	100	96	96	96	1	1			
² ³	Italien	✓	✓	4	9.7	100	6.2	80	99	95	75	94	1	3			
	Japan	✓	✓	4	10.5	100	2.9	96	99	98	94	97	1	0			
² ³	Kanada ^E			4	9.9	79	6.1	80	86	94	74	80	2	3			
	Kasachstan	✓	(✓)	4	10.3	100	3.9	97	99	98	95	97	1	1			
	Katar	✓	(✓)	4	10.1	100	3.8	100	100	99	99	99	9	13			
	Kroatien	✓		4	10.6	100	4.4	99	100	95	93	94	1	3			
⁴ ⁵ ⁷	Kuwait ^{GH}	(✓)	(✓)	4	9.7	100	3.0	94	94	94	94	96	90	90	25	20	
²	Litauen	✓	✓	4	10.7	100	6.1	99	100	94	93	94	1	1			
⁵ ⁷	Marokko ^{GH}	✓	(✓)	4	10.3	100	1.5	100	100	100	99	99	99	99	19	12	
¹ ³	Neuseeland ^C	✓	✓	4.5-5.5	10.0	100	4.8	81	96	94	76	90	4	6			
³	Niederlande	✓	✓	4	10.0	100	3.2	48	87	96	46	83	1	1			
³	Nordirland	✓		4	10.4	100	2.7	65	76	93	60	71	1	2			
	Oman	✓		4	9.6	100	0.8	97	98	99	96	97	9	14			
	Polen	(✓)		4	10.7	100	4.0	91	100	92	84	92	1	2			
²	Portugal	✓		4	9.9	100	6.5	89	99	93	83	92	1	1			
	Republik Korea (Südkorea)	✓		4	10.5	100	2.5	100	100	97	97	97	0	0			
	Russische Föderation	✓	✓	4	10.8	100	4.0	100	100	98	98	98	0	1			
⁴	Saudi-Arabien	✓		4	10.0	100	1.9	95	100	93	88	93	14	22			
²	Schweden	✓	✓	4	10.8	100	5.7	100	100	95	95	95	1	2			
²	Serbien	✓		4	10.7	100	11.3	99	100	96	95	96	2	4			
²	Singapur	✓	✓	4	10.4	100	10.1	100	100	96	96	96	1	1			
	Slowakei	✓	✓	4	10.4	100	4.2	98	100	97	95	97	2	5			
	Slowenien	✓	✓	4	9.8	100	4.5	96	99	95	91	93	1	2			
²	Spanien	✓		4	9.9	100	5.6	98	99	96	95	95	1	3			
	Taiwan	✓	✓	4	10.2	100	2.4	99	100	99	98	99	1	0			
	Tschechische Republik	✓	✓	4	10.4	100	4.2	100	100	95	95	95	1	2			
	Türkei	(✓)		4	9.9	100	3.6	100	100	98	98	98	4	7			
	Ungarn	✓	✓	4	10.7	100	4.8	99	99	97	96	96	2	4			
² ³	USA	✓	✓	4	10.2	100	6.8	77	85	96	74	81	1	2			
	Vereinigte Arabische Emirate (VAE)	✓		4	9.8	100	4.7	100	100	97	97	97	8	12			
	Zypern			4	9.8	100	4.6	100	100	98	98	98	4	3			
Teilnahme mit Jgst. 5																	
¹	Norwegen			5	10.7	100	4.7	93	93	95	89	89	1	1			
Benchmark-Teilnehmer																	
² ⁴	Abu Dhabi, VAE	✓		4	9.8	100	5.8	100	100	97	97	97	13	18			
³ ⁷	Buenos Aires, Argentinien ^{GH}			4	9.8	100	1.9	86	86	91	91	93	93	74	75	14	10
²	Dubai, VAE	✓	✓	4	9.8	100	5.3	100	100	97	97	97	3	5			
²	Florida, USA ^{E F}	✓		4	10.4	90	4.7	100	100	95	95	95	1	2			
	Norwegen (4. Jgst.)	✓	✓	4	9.7	100	5.0	94	94	95	89	89	3	4			
	Ontario, Kanada	✓	✓	4	9.8	100	3.4	95	95	95	90	90	1	3			
² ³	Québec, Kanada	✓	✓	4	10.1	100	5.4	48	62	95	46	59	1	1			

Tabelle 4: Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2019

TIMSS 2019		Teilnehmer					Besonderheiten bzgl. der nationalen Zielpopulation				Besonderheiten bzgl. der nationalen Schul- und Schülerteilnahmequoten				Besonderheiten bzgl. der Leistungsmessung		
											Schulerteilnahmequote in %		Schulerteilnahmequote in %	Gesamterteilnahmequote in %			
							ohne Ersatzschulen	mit Ersatzschulen	ohne Ersatzschulen	mit Ersatzschulen	Naturwissenschaften	Mathematik					
7	Albanien ¹		✓				4	10.0	100	4.2	99	99	99	98	98	3	2
8	Armenien	✓	✓	✓		(✓)	4	9.9	100	1.2	100	100	97	97	97	4	4
	Aserbaidschan	✓	✓		✓		4	10.3	100	2.6	94	98	95	89	92	8	4
	Australien	✓	✓	✓	✓	✓	4	10.1	100	4.8	99	100	94	93	94	2	4
	Bahrain	✓	✓	✓	✓		4	9.8	100	0.8	100	100	98	98	98	4	9
3	Belgien (Fläm. Gem.)	✓	✓	✓	✓		4	10.0	100	3.0	66	95	93	62	89	2	1
7	Bosnien und Herzegowina ¹	✓	✓				4	10.1	100	2.0	100	100	96	95	95	5	2
	Bulgarien	✓	✓	✓			4	10.7	100	3.4	97	100	95	92	95	4	5
	Chile ¹	✓		✓	✓		4	10.1	100	3.8	89	99	96	86	95	5	12
3 8	Dänemark	✓		✓	✓	✓	4	10.9	100	3.1	70	95	87	61	83	1	2
	Deutschland	✓		✓	✓	✓	4	10.4	100	3.9	97	100	97	94	97	2	2
1 2 8	England ^D	✓		✓	✓	✓	5	10.2	100	5.8	86	93	96	82	89	1	2
	Finnland	✓		✓	✓		4	10.8	100	3.3	99	100	97	96	97	1	2
	Frankreich	✓		✓			4	9.9	100	4.4	100	100	98	98	98	4	6
2	Georgien ^E	✓		✓	✓	✓	4	10.1	92	4.7	97	99	97	94	96	6	7
3 8	Hongkong	✓		✓	✓	✓	4	10.1	100	3.5	67	88	90	60	79	1	0
8	Iran		✓	✓	✓	✓	4	10.2	100	4.2	100	100	99	99	99	9	14
	Irland		✓	✓	✓		4	10.4	100	3.0	100	100	91	91	91	2	2
8	Italien	✓		✓	✓	✓	4	9.6	100	4.9	96	100	97	92	97	2	2
3 8	Japan		✓	✓	✓	✓	4	10.4	100	2.2	84	98	97	82	95	1	0
2	Kanada ^F	✓		✓			4	9.9	79	7.0	86	90	95	82	86	2	3
2 8	Kasachstan		✓	(✓)	✓	(✓)	4	10.4	100	5.8	100	100	99	99	99	2	2
8	Katar	✓		✓	✓	(✓)	4	9.9	100	2.2	100	100	97	97	97	8	12
2 7	Kosovo ¹		✓				4	9.9	100	8.6	100	100	97	97	97	9	2
	Kroatien	✓		✓	✓		4	10.5	100	4.2	95	97	91	85	87	1	2
7 8	Kuwait ¹		✓	✓	(✓)	(✓)	4	9.7	100	1.7	97	98	96	93	94	14	8
2 8	Lettland		✓			(✓)	4	10.8	100	6.9	92	99	94	87	93	0	1
2 8	Litauen	✓		✓	✓	✓	4	10.7	100	6.7	100	100	94	94	94	1	2
	Malta	✓			✓		4	9.8	100	4.5	100	100	96	96	96	3	4
5 7 8	Marokko ¹		✓	✓	✓	(✓)	4	10.1	100	1.8	100	100	99	99	99	17	6
7	Montenegro ¹		✓				4	9.8	100	4.6	100	100	98	98	98	8	3
1 2 8	Neuseeland ^D		✓	✓	✓	✓	4.5 - 5.5	10.0	100	6.9	87	99	94	81	93	3	6
3 8	Niederlande	✓		✓	✓	✓	4	10.1	100	3.5	46	75	97	45	73	2	1
3 8	Nordirland		✓	✓	✓		4	10.4	100	2.8	60	86	91	55	78	2	2
1 3 8	Norwegen	✓		✓	(✓)	(✓)	5	10.7	100	4.7	70	90	94	66	84	1	2
	Oman ¹		✓	✓			4	9.7	100	2.2	99	100	98	98	98	10	14
2 8	Österreich	✓		✓	✓	✓	4	10.4	100	5.4	99	99	97	97	97	1	1
2 3 4 5 7	Pakistan ¹		✓				4	10.6	100	7.5	77	99	98	75	96	31	16
2 4 5 7	Philippinen ¹		✓				4	10.1	100	7.7	100	100	98	98	98	34	20
8	Polen		✓	✓	(✓)		4	10.3	100	3.1	96	100	93	89	93	1	3
2	Portugal	✓		✓		✓	4	10.0	100	7.8	87	100	94	81	94	2	2
	Republik Korea (Südkorea)	✓		✓	✓		4	10.5	100	2.3	99	99	98	97	97	0	0
7	Republik Nordmazedonien ¹		✓				4	9.8	100	3.8	98	100	95	94	95	11	3
2 8	Russische Föderation	✓		✓	✓	✓	4	10.8	100	6.3	99	99	98	97	97	0	1
2 7	Saudi-Arabien ¹		✓	✓	✓		4	9.9	100	10.5	98	99	99	97	98	12	5
2	Schweden	✓		✓	✓	✓	4	10.8	100	5.4	100	100	95	95	95	1	3
2	Serbien		✓	✓	✓		4	10.6	100	8.2	97	100	97	95	97	3	5
2 8	Singapur	✓		✓	✓	✓	4	10.4	100	12.8	100	100	97	97	97	1	0
2	Slowakei	✓		✓	✓	✓	4	10.4	100	5.5	97	99	97	93	96	2	5
2	Spanien	✓		✓	✓		4	9.9	100	5.4	97	99	95	92	95	1	4
1 5 7	Südafrika ^{D1}		✓	✓			5	11.5	100	1.1	96	99	98	94	97	28	6
	Taiwan	✓		✓	✓	✓	4	10.2	100	2.0	95	99	99	94	98	0	0
	Tschechische Republik	✓		✓	✓	✓	4	10.4	100	4.7	99	100	96	95	96	1	2
1 2	Türkei	✓		(✓)	(✓)		5	10.6	100	7.0	99	100	99	98	99	3	5
	Ungarn	✓		✓	✓	✓	4	10.5	100	4.1	93	99	97	90	96	1	3
2 3	USA	✓		✓	✓	✓	4	10.2	100	7.2	76	88	96	73	84	2	3
	Vereinigte Arabische Emirate (VAE)	✓		✓	✓		4	9.7	100	3.2	100	100	96	96	96	7	8
	Zypern		✓	✓			4	9.8	100	4.6	99	100	97	96	97	2	2
	Benchmark-Teilnehmer																
	Abu Dhabi, VAE	✓		✓	✓		4	9.7	100	3.6	100	100	95	95	95	14	14
2 8	Dubai, VAE	✓		✓	✓	✓	4	9.9	100	5.6	100	100	97	97	97	1	2
	Madrid	✓					4	9.9	100	3.5	100	100	96	96	96	1	2
	Moskau	✓					4	10.8	100	2.1	99	100	98	97	97	0	0
8	Ontario, Kanada	✓		✓	✓	✓	4	9.8	100	4.9	93	95	95	88	90	2	3
3 8	Québec, Kanada	✓		✓	✓	✓	4	10.1	100	4.0	82	86	96	79	83	1	1

- Als Abweichung von der internationalen Vorgabe werden folgende Fälle gekennzeichnet: Getestete Jahrgangsstufe entspricht nicht der vierten Jahrgangsstufe (vgl. zugleich Fußnoten A, B und C); Ausschöpfungsgrad der nationalen Zielpopulation kleiner 100 %; Ausschlüsse von der nationalen Zielpopulation (Gesamtquote) größer 5 %, aber kleiner oder gleich 10 %; Schulteilnahmequote kleiner 85 %, aber größer oder gleich 75 %; Gesamteilnahmequote kleiner 75 %, aber größer oder gleich 60 %; Anteil der Schülerinnen und Schüler ohne gültige Leistungswerte in Mathematik bzw. Naturwissenschaften größer 15 %, aber kleiner oder gleich 25 %.
- Als starke Abweichung von der internationalen Vorgabe werden folgende Fälle gekennzeichnet: Ausschlüsse von der nationalen Zielpopulation (Gesamtquote) größer 10 %; Schulteilnahmequote kleiner 75 %; Gesamteilnahmequote kleiner 60 %; Anteil der Schülerinnen und Schüler ohne gültige Leistungswerte in Mathematik bzw. Naturwissenschaften größer 25 %.
-  = Teilnahme an eTIMSS
-  = Teilnahme an Papier-TIMSS
-  = Teilnahme an Less Difficult Mathematics
- 1 = Die nationale Zielpopulation entspricht nicht oder nicht ausschließlich der vierten Jahrgangsstufe.
- 2 = Der Ausschöpfungsgrad und/oder die Ausschlüsse von der nationalen Zielpopulation erfüllen nicht die internationalen Vorgaben.
- 3 = Die Teilnahmequoten auf Schul- und/oder Schülerebene erreichen nicht die internationalen Vorgaben.
- 4 = Sehr hoher Anteil (mind. 15 %) an Schülerinnen und Schülern mit nicht skalierbaren Leistungswerten in Mathematik
- 5 = Sehr hoher Anteil (mind. 15 %) an Schülerinnen und Schülern mit nicht skalierbaren Leistungswerten in Naturwissenschaften
- 6 = Abweichender Testzeitpunkt
- 7 = Fünf Teilnehmer sowie ein Benchmark-Teilnehmer haben sowohl an TIMSS 2015 als auch an TIMSS Numeracy teilgenommen: Bahrain, Indonesien, Iran, Kuwait, Marokko und Buenos Aires. Die zugehörigen Mathematikresultate basieren in Anlehnung an die internationale Berichterstattung auf einer Zusammenführung der Ergebnisse beider Studienkomponenten (Mullis, Martin, Foy & Hooper, 2016). Jordanien und Südafrika haben sich ausschließlich an TIMSS Numeracy beteiligt (ebd.), sie waren somit nicht Teil der Berichterstattung zu TIMSS 2015. In 2019 gab es mit TIMSS Less Difficult Mathematics eine Weiterentwicklung von TIMSS Numeracy, die bereits Teile des regulären Tests enthält, sodass die Schülerinnen und Schüler nicht an zwei Testungen teilnehmen müssen (Cotter, Centurino & Mullis, 2020). An TIMSS Less Difficult Mathematics haben Albanien, Bosnien und Herzegowina, Kosovo, Kuwait, Marokko, Montenegro, Pakistan, Philippinen, Republik Nordmazedonien, Saudi-Arabien und Südafrika teilgenommen (Mullis, Martin, Foy, Kelly & Fishbein, 2020).
- 8 = Bei Trendvergleichen über die Studienzyklen TIMSS 2007, 2011, 2015 und 2019 werden sechs Teilnehmer aufgrund von Änderungen in den Übersetzungen der Testinstrumente oder in den Erhebungsbedingungen nicht berücksichtigt: Armenien, Kasachstan, Katar, Kuwait, Marokko und Norwegen. Zusätzlich wird Österreich trotz Nicht-Teilnahme in 2015 in die Trendvergleiche aufgenommen, da es sich dabei um den einzigen anderen deutschsprachigen Staat handelt, der in jüngerer Vergangenheit Bildungsstandards eingeführt hat (siehe Kapitel 2 in diesem Band). Darüber hinaus sind bei Trendvergleichen folgende Besonderheiten hinsichtlich Schulstichproben und eingesetzten Testsprachen zu beachten: In Litauen berücksichtigen die Trendergebnisse keine Schülerinnen und Schüler, die in Polnisch oder Russisch unterrichtet wurden. In Italien, Litauen und Norwegen sind die Testsprachen bzw. Kombinationen von Testsprachen nicht über alle vier Studienzyklen identisch. Bei folgenden Teilnehmern und Benchmark-Teilnehmern sollten Änderungen in den Ausschluss- und Teilnahmequoten über die vier Studienzyklen bei Trendvergleichen berücksichtigt werden: Dänemark, England, Hongkong, Italien, Japan, Norwegen und Singapur sowie Dubai (VAE) und Québec (Kanada). Für den Iran ist im Trendvergleich zu beachten, dass die Ergebnisse für 2015 auf zusammengeführten Testleistungen aus TIMSS 2015 und TIMSS Numeracy beruhen (Mullis, Martin, Foy, Kelly & Fishbein, 2020; Martin, Mullis, Foy & Hooper, 2016; Mullis, Martin, Foy & Hooper, 2016; zu den Testsprachen siehe Ebbs, Wagner & Netten, 2020; Ebbs & Korsnakova, 2016; Johansone & Malak, 2008; Martin & Mullis, 2012).
- * = Symbole in Klammern indizieren eine Teilnahme am Studienzyklus unter veränderten Testbedingungen, die eine Vergleichbarkeit der Studienzyklen beeinträchtigen und Trendvergleiche ausschließen (Mullis, Martin, Foy, Kelly & Fishbein, 2020).
- ** = Durchschnittsalter (mit Dezimalstellen) berechnet als: Anzahl der Jahre + Anzahl der Monate/12 (z. B. 9 Jahre, 3 Monate = $9 + 3/12 = 9.25$ Jahre)
- *** = Ausschöpfungsgrad der nationalen Zielpopulation in Prozent bezogen auf die internationale Vorgabe (100 %)
- **** = Ausschlüsse von der nationalen Zielpopulation (Gesamtquote) in Prozent
- A = Da das Durchschnittsalter der Viertklässlerinnen und Viertklässler in England, Neuseeland und Schottland bedeutsam unter der internationalen Vorgabe (9.5 Jahre) lag, wurde gemäß der Kombination der Kriterien ‚formale Beschulungszeit‘ und ‚Durchschnittsalter des Schülerjahrgangs‘ eine höhere Jahrgangsstufe als äquivalente nationale Zielpopulation festgelegt (Mullis, Martin & Foy, 2008, S. 378-381).
- B = Da das Durchschnittsalter der Viertklässlerinnen und Viertklässler in England, Malta und Neuseeland bedeutsam unter der internationalen Vorgabe (9.5 Jahre) lag, wurde gemäß der Kombination der Kriterien ‚formale Beschulungszeit‘ und ‚Durchschnittsalter des Schülerjahrgangs‘ eine höhere Jahrgangsstufe als äquivalente nationale Zielpopulation festgelegt (Mullis, Martin, Foy & Arora, 2012; Martin, Mullis, Foy & Stanco, 2012).
- C = Da das Durchschnittsalter der Viertklässlerinnen und Viertklässler in England und Neuseeland bedeutsam unter der internationalen Vorgabe (9.5 Jahre) lag, wurde gemäß der Kombination der Kriterien ‚formale Beschulungszeit‘ und ‚Durchschnittsalter des Schülerjahrgangs‘ eine höhere Jahrgangsstufe als äquivalente nationale Zielpopulation festgelegt (Martin, Mullis, Foy & Hooper, 2016; Mullis, Martin, Foy & Hooper, 2016).
- D = Da das Durchschnittsalter der Viertklässlerinnen und Viertklässler in England, Neuseeland und Südafrika bedeutsam unter der internationalen Vorgabe (9.5 Jahre) lag, wurde gemäß der Kombination der Kriterien ‚formale Beschulungszeit‘ und ‚Durchschnittsalter des Schülerjahrgangs‘ eine höhere Jahrgangsstufe als äquivalente nationale Zielpopulation festgelegt (Mullis, Martin, Foy, Kelly & Fishbein, 2020).

- E = Hinweise zum geringen Ausschöpfungsgrad: In 2015 wurden in Florida nur Schülerinnen und Schüler getestet, die staatliche Schulen besuchen. In 2015 und 2019 wurden in Georgien nur in Georgisch unterrichtete Schülerinnen und Schüler getestet, in Kanada nur Schülerinnen und Schüler aus den Provinzen Alberta, Manitoba, Neufundland, Ontario und Québec getestet (Mullis, Martin, Foy, Kelly & Fishbein, 2020; Martin, Mullis, Foy & Hooper, 2016; Mullis, Martin, Foy & Hooper, 2016)
- F = Ergebnisse aus TIMSS 2015 zu Armenien und Florida werden nicht berichtet. Armenien ist zum Zeitpunkt dieser Berichtslegung aufgrund verzögerter Datenaufbereitung nicht Teil der internationalen Berichterstattung zu TIMSS 2015; Florida hat den Datensatz nicht für eine Nutzung freigegeben.
- G = Für Bahrain, Indonesien, Iran, Kuwait, Marokko und Buenos Aires werden die Teilnahmequoten für Mathematik (Zusammenführung von TIMSS 2015 und TIMSS Numeracy) und Naturwissenschaften separat ausgewiesen.
- H = Für Bahrain, Indonesien, Iran, Kuwait, Marokko und Buenos Aires werden die Anteile von Schülerinnen und Schülern ohne gültige Leistungswerte in Mathematik für die Zusammenführung von TIMSS 2015 und TIMSS Numeracy berichtet.
- I = Für Albanien, Bosnien und Herzegowina, Kosovo, Kuwait, Marokko, Montenegro, Pakistan, Philippinen, Republik Nordmazedonien, Saudi-Arabien und Südafrika werden die Ergebnisse von TIMSS Less Difficult Mathematics berichtet.
- J = Die in diesem Band berichteten Ergebnisse basieren auf dem Stand vom 1.11.2020. Die Befunde für die Inhaltsbereiche in Mathematik sind für Chile, die kognitiven Anforderungsbereiche in den Naturwissenschaften für die Republik Nordmazedonien und den Oman als vorläufig anzusehen.

Literatur

- Cotter, K. E., Centurino, V. A. S. & Mullis, I. V. S. (2020). Developing the TIMSS 2019 mathematics and science achievement instruments. In M. O. Martin, M. von Davier & I. V. S. Mullis (Hrsg.), *Methods and procedures: TIMSS 2019 technical report* (S. 1.1–1.36). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Verfügbar unter: <https://timssandpirls.bc.edu/timss2019/methods/chapter-1.html>
- Ebbs, D. & Korsnakova, P. (2016). Translation and translation verification. In M. O. Martin, I. V. S. Mullis & M. Hooper (Hrsg.), *Methods and procedures in TIMSS 2015* (S. 7.1–7.16). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Verfügbar unter: <http://timssandpirls.bc.edu/publications/timss/2015-methods/chapter-7.html>
- Johansone, I. & Malak, B. (2008). Translation and national adaptations of the TIMSS 2007 assessment and questionnaires. In J. F. Olson, M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS 2007 technical report* (S. 63–75). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O. & Mullis, I. V. S. (Hrsg.). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Foy, P. & Hooper, M. (2016). *TIMSS 2015 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Foy, P. & Stanco, G. M. (2012). *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., von Davier, M. & Mullis, I. V. S. (Hrsg.). *Methods and procedures: TIMSS 2019 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Verfügbar unter: <https://timssandpirls.bc.edu/timss2019/methods/>
- Mullis, I. V. S., Martin, M. O. & Foy, P. (2008). *TIMSS 2007 international mathematics report. Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P. & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P. & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Verfügbar unter: <http://timssandpirls.bc.edu/timss2019/international-results/>

Zum Zeitpunkt der Veröffentlichung des vorliegenden Bandes lagen noch nicht alle Kapitel des *Methods and procedures: TIMSS 2019 technical report* vor. Interessierte Lesende finden weiterführende Informationen unter: <https://timssandpirls.bc.edu/timss2019/methods/>

Abbildungsverzeichnis

Abbildung 2.1:	Staaten und Regionen, die in 2019 an eTIMSS und Papier-TIMSS teilgenommen haben	32
Abbildung 2.2:	Teilnehmer an TIMSS 2007, 2011, 2015 und 2019	32
Abbildung 2.3:	Normalverteilung mit Perzentilband, <i>Benchmarks</i> und Kompetenzstufen	53
Abbildung 3.1:	Beispielaufgaben zu Inhaltsbereichen und kognitiven Anforderungsbereichen	69
Abbildung 3.2:	Beispiele curricular nicht valider Testaufgaben in Mathematik	74
Abbildung 3.3:	Kompetenzstufen und Beispielaufgaben (Gesamtskala Mathematik) I	77
Abbildung 3.4:	Kompetenzstufen und Beispielaufgaben (Gesamtskala Mathematik) II	78
Abbildung 3.5:	Testleistung der Schülerinnen und Schüler im internationalen Vergleich – Gesamtskala Mathematik	83
Abbildung 3.6:	Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 – Gesamtskala Mathematik	85
Abbildung 3.7:	Prozentuale Verteilung der Schülerinnen und Schüler auf die fünf Kompetenzstufen (Mathematik) im internationalen Vergleich	89
Abbildung 3.8:	Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den Inhaltsbereichen <i>Arithmetik, Messen und Geometrie</i> sowie <i>Daten</i> im Vergleich zur Gesamtskala	93
Abbildung 3.9:	Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 in den Inhaltsbereichen <i>Arithmetik, Messen und Geometrie</i> und <i>Daten</i> II	96
Abbildung 3.10:	Testleistung der Schülerinnen und Schüler im internationalen Vergleich in den kognitiven Anforderungsbereichen <i>Reproduzieren, Anwenden</i> und <i>Problemlösen</i> im Vergleich zur Gesamtskala	98
Abbildung 3.11:	Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 in den kognitiven Anforderungsbereichen <i>Reproduzieren, Anwenden</i> und <i>Problemlösen</i> II	101
Abbildung 3.12:	Mittlere positive Einstellung von Schülerinnen und Schülern zur Mathematik sowie prozentuale Verteilungen nach niedriger, mittlerer und hoher Einstellung – TIMSS 2007, 2011, 2015 und 2019 im Vergleich	103
Abbildung 3.13:	Prozentuale Verteilung von Schülerinnen und Schülern auf den fünf Kompetenzstufen nach Einstellung zur Mathematik	104
Abbildung 3.14:	Mittleres mathematikbezogenes Selbstkonzept von Schülerinnen und Schülern sowie prozentuale Verteilungen nach niedrigem, mittlerem und hohem Selbstkonzept – TIMSS 2007, 2011, 2015 und 2019 im Vergleich	105
Abbildung 3.15:	Prozentuale Verteilung von Schülerinnen und Schülern in Mathematik auf den fünf Kompetenzstufen nach mathematikbezogenem Selbstkonzept	106
Abbildung 4.1:	Beispielaufgaben für die Inhaltsbereiche und kognitiven Anforderungsbereiche	126
Abbildung 4.2:	Beispiel für eine curricular nicht valide Aufgabe	131
Abbildung 4.3:	Kompetenzstufen und Beispielaufgaben (Gesamtskala Naturwissenschaften) I	134
Abbildung 4.4:	Kompetenzstufen und Beispielaufgaben (Gesamtskala Naturwissenschaften) II	135
Abbildung 4.5:	Testleistung der Schülerinnen und Schüler im internationalen Vergleich – Gesamtskala Naturwissenschaften	139
Abbildung 4.6:	Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 – Gesamtskala Naturwissenschaften	144
Abbildung 4.7:	Prozentuale Verteilung der Schülerinnen und Schüler auf die fünf Kompetenzstufen (Naturwissenschaften) im internationalen Vergleich	148
Abbildung 4.8:	Testleistungen der Schülerinnen und Schüler in Naturwissenschaften im internationalen Vergleich in den Inhaltsbereichen <i>Biologie, Physik/Chemie</i> und <i>Geografie</i>	150

Abbildung 4.9:	Vergleich der Testleistungen in Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 in den Inhaltsbereichen <i>Biologie, Physik/Chemie</i> und <i>Geografie II</i>	152
Abbildung 4.10:	Testleistungen der Schülerinnen und Schüler in Naturwissenschaften im internationalen Vergleich in den kognitiven Anforderungsbereichen <i>Reproduzieren, Anwenden</i> und <i>Problemlösen</i>	154
Abbildung 4.11:	Vergleich der Testleistungen in Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 in den kognitiven Anforderungsbereichen <i>Reproduzieren, Anwenden</i> und <i>Problemlösen II</i>	157
Abbildung 4.12:	Mittlere positive Einstellung von Schülerinnen und Schülern zum Sachunterricht sowie prozentuale Verteilungen nach niedriger, mittlerer und hoher positiver Einstellung – TIMSS 2007, 2011, 2015 und 2019 im Vergleich	159
Abbildung 4.13:	Mittleres sachunterrichtbezogenes Selbstkonzept von Schülerinnen und Schülern sowie prozentuale Verteilungen nach hohem, mittlerem und niedrigem Selbstkonzept – TIMSS 2007, 2011, 2015 und 2019 im Vergleich	161
Abbildung 4.14:	Prozentuale Verteilung von Schülerinnen und Schülern auf den fünf Kompetenzstufen nach Einstellungen zum Fach Sachunterricht.....	162
Abbildung 4.15:	Prozentuale Verteilung von Schülerinnen und Schülern auf den fünf Kompetenzstufen nach sachunterrichtsbezogenem Selbstkonzept.....	162
Abbildung 5.1:	Testdesign für deutsche Stichproben in TIMSS 2015 und TIMSS 2019.....	172
Abbildung 6.1:	Wahrgenommene Unterrichtsqualität im Mathematikunterricht (Anteile der Schülerinnen und Schüler in Prozent)	190
Abbildung 6.2:	Wahrgenommene Unterrichtsqualität im naturwissenschaftsbezogenen Sachunterricht (Anteile der Schülerinnen und Schüler in Prozent).....	191
Abbildung 6.3:	Wahrgenommene Störungen im Mathematikunterricht nach hoher, mittlerer und geringer Ausprägung (Anteile der Schülerinnen und Schüler in Prozent)	193
Abbildung 6.4:	Wahrgenommene Motivierungsqualität im Mathematik- und naturwissenschaftsbezogenen Sachunterricht (Anteile der Schülerinnen und Schüler in Prozent)	196
Abbildung 6.5:	Leistungsbezogene Differenzierung im Mathematikunterricht aus Sicht der Lehrkräfte	201
Abbildung 6.6:	Leistungsbezogene Differenzierung im naturwissenschaftsbezogenen Sachunterricht aus Sicht der Lehrkräfte	202
Abbildung 7.1:	Privater Nachhilfeunterricht in Mathematik im internationalen Vergleich (Anteil der Schülerinnen und Schüler nach Elternangaben in Prozent)	211
Abbildung 7.2:	Privater Nachhilfeunterricht in Sachunterricht (Anteil der Schülerinnen und Schüler nach Elternangaben in Prozent)	212
Abbildung 7.3:	Außerschulische Förderung im Vergleich von IGLU 2006 und TIMSS 2011 bis TIMSS 2019 (Anteil der Schülerinnen und Schüler nach Elternangaben in Prozent)	213
Abbildung 7.4:	Privater Nachhilfeunterricht in Mathematik differenziert nach Halbjahresnoten und Armutsrisiko der Schülerinnen und Schüler (Anteile der Schülerinnen und Schüler in Prozent)	214
Abbildung 7.5:	Außerschulische Unterstützung von Schülerinnen und Schülern mit Mathematikleistungen unterhalb der Kompetenzstufe III (Anteil der Schülerinnen und Schüler in Prozent und deren Standardfehler in Klammern, bezogen auf alle Schülerinnen und Schüler mit entsprechenden Leistungen)	216
Abbildung 8.1:	Testleistung nach Geschlecht – Gesamtskala Mathematik	231
Abbildung 8.2:	Mittlere Testleistungen von Mädchen und Jungen und prozentuale Verteilungen auf die Kompetenzstufen – Mathematik	232

Abbildung 8.3:	Testleistungen in Inhaltsbereichen nach Geschlecht im internationalen Vergleich – Mathematik	234
Abbildung 8.4:	Testleistungen in kognitiven Anforderungsbereichen nach Geschlecht im internationalen Vergleich – Mathematik	236
Abbildung 8.5:	Vergleich der Leistungsentwicklungen von Mädchen und Jungen in Mathematik in TIMSS 2007, 2011, 2015 und 2019 (internationaler Vergleich).....	238
Abbildung 8.6:	Vergleich der Leistungsunterschiede in Mathematik von Mädchen und Jungen in TIMSS 2007, 2011, 2015 und 2019 (internationaler Vergleich)	240
Abbildung 8.7:	Testleistung nach Geschlecht – Gesamtskala Naturwissenschaften	242
Abbildung 8.8:	Mittlere Testleistungen von Mädchen und Jungen und prozentuale Verteilungen auf die Kompetenzstufen – Naturwissenschaften	243
Abbildung 8.9:	Testleistungen in Inhaltsbereichen nach Geschlecht im internationalen Vergleich – Naturwissenschaften	244
Abbildung 8.10:	Testleistungen in kognitiven Anforderungsbereichen nach Geschlecht im internationalen Vergleich – Naturwissenschaften	246
Abbildung 8.11:	Vergleich der Leistungsentwicklungen von Mädchen und Jungen in Naturwissenschaften in TIMSS 2007, 2011, 2015 und 2019 (internationaler Vergleich).....	248
Abbildung 8.12:	Vergleich der Leistungsunterschiede in den Naturwissenschaften von Mädchen und Jungen in TIMSS 2007, 2011, 2015 und 2019 (internationaler Vergleich)	250
Abbildung 8.13:	Prozentuale Verteilung auf der Skala <i>positive Einstellung zur Mathematik</i> in TIMSS 2007, 2011, 2015 und 2019 – Mädchen und Jungen im Vergleich	252
Abbildung 8.14:	Prozentuale Verteilung auf der Skala <i>mathematikbezogenes Selbstkonzept</i> in TIMSS 2007, 2011, 2015 und 2019 – Mädchen und Jungen im Vergleich	253
Abbildung 8.15:	Prozentuale Verteilung auf der Skala <i>positive Einstellung zum Fach Sachunterricht</i> in TIMSS 2007, 2011, 2015 und 2019 – Mädchen und Jungen im Vergleich.....	254
Abbildung 8.16:	Prozentuale Verteilung auf der Skala <i>sachunterrichtsbezogenes Selbstkonzept</i> in TIMSS 2007, 2011, 2015 und 2019 – Mädchen und Jungen im Vergleich	255
Abbildung 9.1:	Leistungsvorsprung in Mathematik von Kindern aus Familien, in denen mindestens ein Elternteil einen tertiären Bildungsabschluss erreicht hat (ISCED-Level 6, 7 und 8), vor Kindern aus Familien, in denen dies nicht der Fall ist	270
Abbildung 9.2:	Leistungsvorsprung in Naturwissenschaften von Kindern aus Familien, in denen mindestens ein Elternteil einen tertiären Bildungsabschluss erreicht hat (ISCED-Level 6, 7 und 8), vor Kindern aus Familien, in denen dies nicht der Fall ist	271
Abbildung 9.3:	Leistungsvorsprung in Mathematik von Kindern aus Familien mit mehr als 100 Büchern vor Kindern aus Familien mit maximal 100 Büchern.....	273
Abbildung 9.4:	Leistungsvorsprung in Naturwissenschaften von Kindern aus Familien mit mehr als 100 Büchern vor Kindern aus Familien mit maximal 100 Büchern	274
Abbildung 9.5:	Unterschiede im Leistungsvorsprung in Mathematik von Kindern aus Familien mit mehr als 100 Büchern vor Kindern aus Familien mit maximal 100 Büchern im Vergleich von TIMSS 2007, 2011, 2015 und 2019	275
Abbildung 9.6:	Unterschiede im Leistungsvorsprung in Naturwissenschaften von Kindern aus Familien mit mehr als 100 Büchern vor Kindern aus Familien mit maximal 100 Büchern im Vergleich von TIMSS 2007, 2011, 2015 und 2019	276

Abbildung 9.7:	Zusammenhang zwischen der erklärten Varianz der Mathematikkompetenz auf Individualebene durch die Anzahl der Bücher im Haushalt und den mittleren Mathematikkompetenzen in den Teilnehmerstaaten.....	277
Abbildung 9.8:	Zusammenhang zwischen der erklärten Varianz der Naturwissenschaftskompetenz auf Individualebene durch die Anzahl der Bücher im Haushalt und den mittleren Naturwissenschaftskompetenzen in den Teilnehmerstaaten.....	278
Abbildung 9.9:	Mittlere mathematische Kompetenzen pro EGP-Klasse in Deutschland bei TIMSS 2007, 2011, 2015 und 2019.....	281
Abbildung 9.10:	Mittlere naturwissenschaftliche Kompetenzen pro EGP-Klasse in Deutschland bei TIMSS 2007, 2011, 2015 und 2019	281
Abbildung 9.11:	Armutsgefährdungsquoten nach Migrationshintergrund, höchstem Bildungsabschluss und EGP-Klasse in Deutschland (in Prozent).....	282
Abbildung 9.12:	Mittlere mathematische Kompetenzen nach <i>Armutsgefährdung</i> und EGP-Klasse in Deutschland bei TIMSS 2007, 2011, 2015 und 2019	283
Abbildung 9.13:	Mittlere naturwissenschaftliche Kompetenzen nach <i>Armutsgefährdung</i> und EGP-Klasse in Deutschland bei TIMSS 2007, 2011, 2015 und 2019	284
Abbildung 10.1:	Testleistungen der Schülerinnen und Schüler in Mathematik nach Migrationshintergrund der Eltern im internationalen Vergleich	297
Abbildung 10.2:	Testleistungen der Schülerinnen und Schüler in Naturwissenschaften nach Migrationshintergrund der Eltern im internationalen Vergleich	299
Abbildung 10.3:	Testleistungen der Schülerinnen und Schüler in Mathematik in Deutschland nach Migrationshintergrund der Eltern – TIMSS 2007, 2011, 2015 und 2019 im Vergleich	302
Abbildung 10.4:	Unterschiede in den mittleren Mathematikleistungen im Trend in Deutschland nach Migrationshintergrund der Eltern – TIMSS 2007, 2011, 2015 und 2019 im Vergleich	303
Abbildung 10.5:	Testleistungen der Schülerinnen und Schüler in Naturwissenschaften in Deutschland nach Migrationshintergrund der Eltern – TIMSS 2007, 2011, 2015 und 2019 im Vergleich	303
Abbildung 10.6:	Unterschiede in den mittleren Testleistungen in Naturwissenschaften in Deutschland im Trend nach Migrationshintergrund der Eltern – TIMSS 2007, 2011, 2015 und 2019 im Vergleich.....	304
Abbildung 10.7:	Prozentuale Verteilung der Schülerinnen und Schüler in Deutschland auf die fünf Kompetenzstufen (Mathematik) nach Migrationshintergrund der Eltern – 2007, 2011, 2015 und 2019 im Vergleich	305
Abbildung 10.8:	Prozentuale Verteilung der Schülerinnen und Schüler in Deutschland auf die fünf Kompetenzstufen (Naturwissenschaften) nach Migrationshintergrund der Eltern – TIMSS 2007, 2011, 2015 und 2019 im Vergleich	306
Abbildung 10.9:	Mittelwerte und Differenzen von mathematik- und naturwissenschaftsbezogenen Selbstkonzepten, Einstellungen zu Mathematik und Naturwissenschaften sowie sozialer Eingebundenheit nach Migrationshintergrund.....	309
Abbildung 11.1:	Testleistungen der Schülerinnen und Schüler differenziert nach Mathematiknoten – Gesamtskala Mathematik	321
Abbildung 11.2:	Testleistungen der Schülerinnen und Schüler differenziert nach Schullaufbahnpräferenz der Lehrkräfte – Gesamtskala Mathematik	322
Abbildung 11.3:	Schwellenwerte der Mathematikkompetenz für eine fünfzigprozentige Wahrscheinlichkeit für eine Gymnasialpräferenz der Lehrkräfte beziehungsweise der Erziehungsberechtigten bei TIMSS 2011, 2015 und 2019	327
Abbildung 12.1:	Veränderungen der Schülerinnen- und Schülerleistungen in Mathematik zwischen TIMSS 2007, 2011, 2015 und 2019 nach Geschlecht und Unterstützungsbedarfen.....	339

Abbildung 12.2:	Veränderungen der Schülerinnen- und Schülerleistungen in den Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 nach Geschlecht und Unterstützungsbedarfen	340
Abbildung 12.3:	Veränderungen der Schülerinnen- und Schülerleistungen in Mathematik zwischen TIMSS 2007, 2011, 2015 und 2019 nach Migrationshintergrund	341
Abbildung 12.4:	Veränderungen der Schülerinnen- und Schülerleistungen in den Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 nach Migrationshintergrund.....	341
Abbildung 12.5:	Mittlere Anteile von Schülerinnen und Schülern nach unterschiedlichen Diversitätsmerkmalen in Klassen der vierten Jahrgangsstufe an Regelschulen in Deutschland im Vergleich von TIMSS 2007, 2011, 2015 und 2019 (Schulen nach Zusammensetzungsmerkmalen in Prozent) I.....	343
Abbildung 12.6:	Mittlere Anteile von Schülerinnen und Schülern nach unterschiedlichen Diversitätsmerkmalen in Klassen der vierten Jahrgangsstufe an Regelschulen in Deutschland im Vergleich von TIMSS 2007, 2011, 2015 und 2019 (Schulen nach Zusammensetzungsmerkmalen in Prozent) II.....	344

Tabellenverzeichnis

Tabelle 2.1:	Vergleichsgruppen in TIMSS 2019	31
Tabelle 2.2:	Länderspezifische Modalitäten der Teilnahme an TIMSS 2007, 2011, 2015 und 2019	37
Tabelle 2.3:	Beteiligungszahlen und zentrale Stichprobencharakteristika in TIMSS 2007, 2011, 2015 und 2019	38
Tabelle 2.4:	Veränderung des Mittelwertes von Viertklässlerinnen und Viertklässlern in Deutschland beim Vergleich der Leistung auf der Gesamtskala Mathematik in Abhängigkeit von der Ausschlussquote	41
Tabelle 2.5:	Veränderung des Mittelwertes von Viertklässlerinnen und Viertklässlern in Deutschland beim Vergleich der Leistung auf der Gesamtskala Naturwissenschaften in Abhängigkeit von der Ausschlussquote	42
Tabelle 2.6:	Anzahl von Testaufgaben für Mathematik und Naturwissenschaften in TIMSS 2007, 2011, 2015 und 2019	43
Tabelle 2.7:	Hintergrundfragebögen und Charakterisierung der erfragten Informationen in TIMSS 2019	46
Tabelle 2.8:	Stichprobenumfänge und Durchführungszeitpunkte der TIMSS-2019-Vorbereitungsstudien und der Hauptstudie	47
Tabelle 2.9:	Untersuchungsablauf der TIMSS-Brücken- und TIMSS-Hauptstudie 2019	50
Tabelle 3.1:	Veröffentlichungszeitpunkte der Lehrpläne Grundschule Mathematik für die vierte Jahrgangsstufe, Stand Mai 2019	61
Tabelle 3.2:	Inhaltsbezogene Kompetenzen der KMK-Bildungsstandards in den Lehrplänen der Länder der Bundesrepublik Deutschland, Stand Mai 2019	62
Tabelle 3.3:	Prozessbezogene Kompetenzen der KMK-Bildungsstandards in den Lehrplänen der Länder der Bundesrepublik Deutschland, Stand Mai 2019	63
Tabelle 3.4:	Inhaltsbereiche und darauf bezogene Kompetenzerwartungen	66
Tabelle 3.5:	Kognitive Aktivitäten im Anforderungsbereich <i>Reproduzieren</i>	67
Tabelle 3.6:	Kognitive Aktivitäten im Anforderungsbereich <i>Anwenden</i>	68
Tabelle 3.7:	Kognitive Aktivitäten im Anforderungsbereich <i>Problemlösen</i>	68
Tabelle 3.8:	Verteilung der Testaufgaben auf die mathematischen Inhaltsbereiche	71
Tabelle 3.9:	Verteilung der Testaufgaben auf die kognitiven Anforderungsbereiche	71
Tabelle 3.10:	Merkmale von curricular nicht validen Testaufgaben in Mathematik	73
Tabelle 3.11:	Anteile curricular nicht valider Testaufgaben nach mathematischen Inhaltsbereichen	74
Tabelle 3.12:	Anteile curricular nicht valider Testaufgaben nach kognitiven Anforderungsbereichen	75
Tabelle 3.13:	Beschreibung der fünf Kompetenzstufen für die Gesamtskala Mathematik	76
Tabelle 3.14.1:	Mathematikleistungen nach Perzentilen im Vergleich: TIMSS 2007, 2011, 2015 und 2019 – Gesamtskala Mathematik	86
Tabelle 3.14.2:	Mathematikleistungen nach Perzentilen im Vergleich: 2007, 2011, 2015 und 2019 – Gesamtskala Mathematik	87
Tabelle 3.15.1:	Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007, 2011, 2015 und 2019 (Angaben in Prozent)	90
Tabelle 3.15.2:	Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007, 2011, 2015 und 2019 (Angaben in Prozent)	91
Tabelle 3.16:	Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 in den Inhaltsbereichen <i>Arithmetik, Messen und Geometrie und Daten I.</i>	95
Tabelle 3.17:	Vergleich der Testleistungen zwischen TIMSS 2007, 2011, 2015 und 2019 in den kognitiven Anforderungsbereichen <i>Reproduzieren, Anwenden und Problemlösen I.</i>	100

Tabelle 4.1:	Einschätzung der Lehrkräfte zum prozentualen Anteil der Unterrichtsstunden in Klasse drei und vier in den fünf Perspektiven des Perspektivrahmens	118
Tabelle 4.2:	Einschätzung der Lehrkräfte zur prozentualen Aufteilung der naturwissenschaftlichen Perspektive auf Themen der belebten und unbelebten Natur	119
Tabelle 4.3:	Teilgebiete der einzelnen naturwissenschaftlichen Inhaltsbereiche	123
Tabelle 4.4:	Denkprozesse der kognitiven Anforderungsbereiche	124
Tabelle 4.5:	Verteilung der Testaufgaben auf die naturwissenschaftlichen Inhaltsbereiche.....	128
Tabelle 4.6:	Verteilung der Testaufgaben auf die kognitiven Anforderungsbereiche.....	128
Tabelle 4.7:	Anteile curricular nicht valider Testaufgaben in Naturwissenschaften nach Inhaltsbereichen	130
Tabelle 4.8:	Beschreibung der fünf Kompetenzstufen für die Gesamtskala Naturwissenschaften.....	133
Tabelle 4.9.1:	Testleistungen nach Perzentilen im Vergleich: TIMSS 2007, 2011, 2015 und 2019 – Gesamtskala Naturwissenschaften.....	142
Tabelle 4.9.2:	Testleistungen nach Perzentilen im Vergleich: TIMSS 2007, 2011, 2015 und 2019 – Gesamtskala Naturwissenschaften.....	143
Tabelle 4.10.1:	Prozentuale Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007, 2011, 2015 und 2019	146
Tabelle 4.10.2:	Prozentuale Verteilung der Schülerinnen und Schüler auf die Kompetenzstufen in TIMSS 2007, 2011, 2015 und 2019	147
Tabelle 4.11:	Vergleich der Testleistungen in Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 in den Inhaltsbereichen <i>Biologie, Physik/Chemie und Geografie I</i>	151
Tabelle 4.12:	Vergleich der Testleistungen in Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 in den kognitiven Anforderungsbereichen <i>Reproduzieren, Anwenden und Problemlösen I</i>	156
Tabelle 5.1:	Moduseffekte im Hinblick auf Itemschwierigkeiten (<i>d</i>) und Itemdiskriminationen (<i>e</i>) für TIMSS 2019.....	179
Tabelle 5.2:	Mittelwerte und Standardabweichungen in TIMSS 2015 (PBA) und TIMSS 2019 (PBA und CBA).....	180
Tabelle 5.3:	Trendschätzungen für Mathematik und Naturwissenschaften für TIMSS 2015 nach TIMSS 2019	181
Tabelle 6.1:	Regression von Leistung in Mathematik und Naturwissenschaften auf verschiedene Basisdimensionen der Unterrichtsqualität, erlebte Motivierungsqualität und individuelle Merkmale von Lernenden (Mehrebenenmodell)	197
Tabelle 7.1:	Übereinstimmung der Angaben von Erziehungsberechtigten und Lehrkräften zur Nutzung von privatem Nachhilfeunterricht in Mathematik und Deutsch (Lesen und Rechtschreibung)	217
Tabelle 7.2:	Fortbildungsteilnahme der Mathematiklehrkräfte (international) nach Inhalten (Anteile der Schülerinnen und Schüler mit entsprechenden Fortbildungen der Mathematiklehrkräfte in Prozent).....	218
Tabelle 7.3:	Fortbildungsteilnahme der Sachunterrichtslehrkräfte (international) nach Inhalten (Anteile der Schülerinnen und Schüler mit entsprechenden Fortbildungen der Sachunterrichtslehrkräfte in Prozent).....	219
Tabelle 8.1:	Mittelwerte und Verteilungen der Noten in Mathematik nach Geschlecht (Anteile der Schülerinnen und Schüler in Prozent)	241
Tabelle 8.2:	Mittelwerte und Verteilungen der Noten im Sachunterricht nach Geschlecht (Anteile der Schülerinnen und Schüler in Prozent).....	251
Tabelle 9.1:	Soziale Gradienten für die mathematischen und naturwissenschaftlichen Kompetenzen in Deutschland bei TIMSS 2007, 2011, 2015 und 2019	279
Tabelle 9.2:	Mittlere mathematische und naturwissenschaftliche Kompetenzen pro EGP-Klasse in Deutschland.....	280

Tabelle 9.3:	Mittlere mathematische und naturwissenschaftliche Kompetenzen nach <i>Armutsgefährdung</i> und EGP-Klasse in Deutschland.....	282
Tabelle 10.1:	Angaben zur Verteilung von Viertklässlerinnen und Viertklässlern im Hinblick auf den Gebrauch der deutschen Sprache, den Erwerbsstatus von Mutter und Vater sowie den Kindergartenbesuch und die <i>Armutsgefährdung</i> der Familien nach Migrationshintergrund (Anteile in Prozent)	301
Tabelle 10.2:	Regressionsmodell zur Erklärung von Unterschieden in mathematischen und naturwissenschaftlichen Kompetenzen zwischen Kindern mit und ohne Migrationshintergrund für TIMSS 2019.....	307
Tabelle 11.1:	Übersicht über die Schulformen der Sekundarstufe I in den Ländern der Bundesrepublik Deutschland (Schuljahr 2018/19)	317
Tabelle 11.2:	Schullaufbahnpräferenzen von Lehrkräften und Erziehungsberechtigten von IGLU 2001 bis TIMSS 2019 (Angaben in Zeilenprozent)	319
Tabelle 11.3:	Übereinstimmung der Schullaufbahnpräferenzen von Lehrkräften und Erziehungsberechtigten (Angaben in Zeilenprozent)	320
Tabelle 11.4:	Zusammenhang zwischen der Schullaufbahnpräferenz der Lehrkräfte und leistungsrelevanten Merkmalen der Schülerinnen und Schüler (ohne und mit Berücksichtigung der Kompetenzen und Schulnoten).....	323
Tabelle 11.5:	Relative Chancen (<i>odds ratios</i>) für eine Gymnasialpräferenz der Lehrkräfte bzw. der Erziehungsberechtigten nach sozialer Lage (EGP) der Schülerinnen- und Schülerfamilien.....	324
Tabelle 11.6:	Relative Chancen (<i>odds ratios</i>) für eine Gymnasialpräferenz der Lehrkräfte bzw. der Erziehungsberechtigten für Kinder aus Familien der <i>service class</i> (EGP I und II) im Vergleich mit Kindern aus Familien der <i>working class</i> (EGP V, VI und VII) bei TIMSS 2011, 2015 und 2019	325
Tabelle 11.7:	Schwellenwerte der Mathematik- und Naturwissenschaftskompetenz für eine fünfzigprozentige Wahrscheinlichkeit für eine Gymnasialpräferenz der Lehrkräfte beziehungsweise der Erziehungsberechtigten	326
Tabelle 12.1:	Charakteristika von Viertklässlerinnen und Viertklässlern in TIMSS 2007, 2011, 2015 und 2019	333
Tabelle 12.2:	Veränderungen der Schülerinnen- und Schülerleistungen in Mathematik zwischen TIMSS 2007, 2011, 2015 und 2019 unter Kontrolle von Schülerinnen- und Schülermerkmalen (Mehrebenenmodell).....	336
Tabelle 12.3:	Veränderungen der Schülerinnen- und Schülerleistungen in Naturwissenschaften zwischen TIMSS 2007, 2011, 2015 und 2019 unter Kontrolle von Schülerinnen- und Schülermerkmalen (Mehrebenenmodell).....	337
Tabelle 1:	Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2007.....	351
Tabelle 2:	Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2011	352
Tabelle 3:	Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2015.....	353
Tabelle 4:	Weiterführende Informationen zu den Teilnehmerstaaten – TIMSS 2019.....	354

Heike Wendt, Wilfried Bos,
Christoph Selter, Olaf Köller,
Knut Schwippert, Daniel Kasper
(Hrsg.)

TIMSS 2015

Mathematische und
naturwissenschaftliche
Kompetenzen von
Grundschulkindern in
Deutschland im
internationalen Vergleich

2016, 408 Seiten, br., 39,90 €,
ISBN 978-3-8309-3566-7

E-Book: Open Access



Heike Wendt
Wilfried Bos
Christoph Selter
Olaf Köller
Knut Schwippert
Daniel Kasper
(Hrsg.)



TIMSS 2015

Mathematische und
naturwissenschaftliche
Kompetenzen von
Grundschulkindern
in Deutschland im
internationalen
Vergleich

WAXMANN

Mit Beiträgen von

Wilfried Bos, Kristina A. Frey,
Martin Goy, Karin Guill, Donieta Jusufi,
Daniel Kasper, Thilo Kleickmann,
Olaf Köller, Jennifer Lorenz,
Raphaella Porsch, Svenja Rieser,
Knut Schwippert, Christoph Selter,
Ruven Stahns, Ricarda Steinmayr,
Mirjam Steffensky, Tobias C. Stubbe,
Renate Valtin, Daniel Walter,
Gerd Walther, Anke Walzebug,
Heike Wendt

Im Jahr 2015 beteiligte sich Deutschland zum dritten Mal an der Grundschuluntersuchung Trends in International Mathematics and Science Study (TIMSS 2015). Mit TIMSS werden alle vier Jahre die Fachleistungen von Schülerinnen und Schülern der vierten Jahrgangsstufe in den Bereichen Mathematik und Naturwissenschaften im internationalen Vergleich untersucht.

An TIMSS 2015 waren – neben Deutschland – weltweit 47 Staaten und Regionen als reguläre Teilnehmer mit Schülerinnen und Schülern der vierten Jahrgangsstufe beteiligt. In diesem Band werden die Ergebnisse von TIMSS 2015 für die Bildungsdiskussion in Deutschland erschlossen. Die Ergebnisse des internationalen Vergleichs werden vor dem Hintergrund von Themen dargestellt, die das Lehren und Lernen an Grundschulen in Deutschland verändert und den Bildungsdiskurs der letzten Jahre besonders geprägt haben. Mit vertiefenden Analysen und der Einordnung der Ergebnisse in den aktuellen Forschungsstand stellt der Band eine differenzierte und anschlussfähige Bestandsaufnahme zur Qualität mathematischer und naturwissenschaftlicher Bildung in der Grundschule dar und nimmt Entwicklungen seit 2007 in den Blick.

Dieser Bericht wendet sich somit an eine Leserschaft, die an bildungspolitischen, pädagogischen und fachdidaktischen Fragestellungen interessiert ist.



WAXMANN

www.waxmann.com
info@waxmann.com



Volker Frederking, Silke Ladel (Hrsg.)

Grundschule digital

Innovative Konzepte für die Fächer
Deutsch und Mathematik

WAXMANN

Volker Frederking,
Silke Ladel (Hrsg.)

Grundschule digital

Innovative Konzepte für die Fächer
Deutsch und Mathematik

2020, ca. 220 Seiten, br., 32,90 €,
ISBN 978-3-8309-4134-7

E-Book: 29,99 €,
ISBN 978-3-8309-9134-2

Mit Beiträgen von

Christine Bescherer, Jan M. Boelmann,
Andreas Fest, Volker Frederking,
Heiko Etzold, Uta Hauck-Thum,
Andres Jurk, Lisa König,
Ulrich Kortenkamp, Günter Krauthausen,
Silke Ladel, Peter Mahns, Svenja Noack,
Guido Nottbusch, Alexandra Pilgrim,
Jennifer Reiske, Klaus Ripper,
Sabine Röttig, Franziska Stutz,
Claudia Vorst

Der Prozess der Digitalisierung eröffnet für das Lehren und Lernen in der Grundschule neue Chancen und Möglichkeiten. Wie diese konkret aussehen können, wird in diesem Band am Beispiel der Fächer Deutsch und Mathematik veranschaulicht. Vorgestellt werden insgesamt acht innovative Konzepte für die unterrichtliche Praxis, die im Rahmen von Forschungsprojekten entwickelt und in der Praxis erfolgreich erprobt wurden.

Für den Deutschunterricht

- Digital Storytelling und intermediales Geschichtenverstehen,
- Bildliteralität und Bilderbuchanalyse unter besonderer Berücksichtigung der Bild-Text-Korrespondenz,
- Förderung der Leseflüssigkeit durch das Lesen mit Hörbuch unter Anpassung der Sprechergeschwindigkeit,
- Adaptable Books und LES-O-MAT – Buchempfehlungen von Kindern für Kinder
- Trickfilm- und Hörspielerstellung mit digitalen Medien

Für den Mathematikunterricht

- Integrative Erarbeitung von Würfelkonfigurationen mit Hilfe von physischen und digitalen Arbeitsmitteln
- Algorithmen im Alltag
- Entdeckendes Lernen mathematischer Zusammenhänge durch altersgerechte Programmierumgebungen



WAXMANN

www.waxmann.com
info@waxmann.com