



Anhang A

PISA 2015 - TECHNISCHE HINWEISE

Alle Tabellen in Anhang A sind online verfügbar

Anhang A1: Indizes zu den Kontextfragebogen für Schüler und Schulen

Anhang A2: PISA-Zielpopulation, PISA-Stichproben und Definition der Schulen

<http://dx.doi.org/10.1787/888933433129>

Anhang A3: Technische Hinweise zu den in diesem Band enthaltenen Analysen

Anhang A4: Qualitätssicherung

Anhang A5: Änderungen bei der Durchführung und Skalierung von PISA 2015 und Auswirkungen auf die Trendanalysen

<http://dx.doi.org/10.1787/888933433162>

Anhang A6: The PISA 2015 field Trial mode-effect study (nicht auf Deutsch verfügbar)

Anmerkung zu P-S-J-G (China)

P-S-J-G (China) bezieht sich auf die vier an PISA teilnehmenden chinesischen Provinzen Peking, Shanghai, Jiangsu und Guangdong.

Anmerkung zu CABA (Argentinien)

CABA (Argentinien) bezieht sich auf die Region der Ciudad Autonoma de Bueno Aires, Argentinien.

Anmerkung zu ejR Mazedonien

ejR Mazedonien bezieht sich auf die ehemalige jugoslawische Republik Mazedonien.

Anmerkungen zu Zypern

Anmerkung der Türkei: Die Informationen zu „Zypern“ in diesem Dokument beziehen sich auf den südlichen Teil der Insel. Es existiert keine Instanz, die sowohl die türkische als auch die griechische Bevölkerung der Insel vertritt. Die Türkei erkennt die Türkische Republik Nordzypern (TRNZ) an. Bis im Rahmen der Vereinten Nationen eine dauerhafte und gerechte Lösung gefunden ist, wird sich die Türkei ihren Standpunkt in der „Zypernfrage“ vorbehalten.

Anmerkung aller in der OECD vertretenen EU-Mitgliedstaaten und der Europäischen Union: Die Republik Zypern wird von allen Mitgliedern der Vereinten Nationen mit Ausnahme der Türkei anerkannt. Die Informationen in diesem Dokument beziehen sich auf das Gebiet, das sich unter der tatsächlichen Kontrolle der Regierung der Republik Zypern befindet.

Anmerkung zu Israel

Die statistischen Daten für Israel wurden von den zuständigen israelischen Stellen bereitgestellt, die für sie verantwortlich zeichnen. Die Verwendung dieser Daten durch die OECD erfolgt unbeschadet des Status der Golanhöhen, von Ost-Jerusalem und der israelischen Siedlungen im Westjordanland gemäß internationalem Recht.

ANHANG A1.

INDIZES ZU DEN KONTEXTFRAGEBOGEN FÜR SCHÜLER UND SCHULEN

Erläuterung der Indizes

Dieser Abschnitt erklärt die auf den PISA-2015-Kontextfragebogen für Schülerinnen und Schüler sowie Schulen beruhenden Indizes, die in diesem Band verwendet werden.

Bei mehreren PISA-Messgrößen handelt es sich um Indizes, die Antworten von Schülern, ihren Eltern, Lehrkräften oder Schulvertretern (in der Regel Schulleiterinnen und Schulleiter) auf eine Reihe miteinander zusammenhängender Fragen zusammenfassen. Die Fragen wurden auf der Grundlage theoretischer Überlegungen und früherer Forschungsarbeiten aus einem größeren Fragenkatalog ausgewählt. Dieser konzeptionelle Rahmen wird in *PISA 2015 Assessment and Analytical Framework* (OECD, 2016) eingehend beschrieben. Zur Bestätigung des theoretisch erwarteten Verhaltens der Indizes und zur Validierung ihrer Vergleichbarkeit zwischen den Ländern wurden Strukturgleichungsmodelle verwendet. Zu diesem Zweck wurde für jedes Land separat und für alle OECD-Länder zusammen eine Modellrechnung durchgeführt. Wegen einer ausführlichen Beschreibung anderer PISA-Indizes und Einzelheiten zu den gewählten Methoden vgl. *PISA 2015 Technical Report* (OECD, erscheint demnächst).

Es gibt drei Arten von Indizes: einfache Indizes und Skalenindizes in Form von neuen Indizes bzw. Trendindizes.

Einfache Indizes sind Variablen, die durch arithmetische Transformation oder Umkodierung eines oder mehrerer Items in den einzelnen Erhebungen jeweils auf genau dieselbe Art und Weise konstruiert werden. Hier werden die Antworten je Item (item response) zur Berechnung aussagekräftiger Variablen verwendet, etwa bei der Umkodierung der vierstelligen ISCO-08-Codes zur Erstellung des Index der höchsten beruflichen Stellung der Eltern (HISEI) oder bei der Berechnung der Schüler/Lehrer-Quote auf der Grundlage der Daten aus dem Schulleiterfragebogen.

Skalenindizes – sowohl in Form neuer Indizes als auch in Form von Trendindizes – sind Variablen, die durch Skalierung mehrerer Items konstruiert werden. Sofern nicht anders vermerkt, wurden die fraglichen Indizes mit Hilfe eines zweiparametrischen Item-Response-Modells skaliert (im Fall von Items mit mehr als zwei Antwortkategorien wurde ein generalisiertes Partial-Credit-Modell verwendet), und die Indexwerte entsprechen Warm-Likelihood-Estimates (WLE) (Warm, 1985). Wegen Einzelheiten zur Konstruktion der einzelnen Skalenindizes vgl. *PISA 2015 Technical Report* (OECD, erscheint demnächst). Die Skalierung erfolgte generell in drei Stufen:

1. Die Itemparameter wurden auf der Basis gleich gewichteter Stichproben von Schülerinnen und Schülern aus allen Ländern und Volkswirtschaften geschätzt. Berücksichtigt wurden lediglich Fälle mit mindestens drei gültigen Antworten zu Items, die Teil des Index sind. Bei **Trendindizes** wurde ein gängiges Verfahren zur Verknüpfung der Kalibrierungen verwendet: Im Fall von Ländern bzw. Volkswirtschaften, die sowohl an PISA 2006 als auch an PISA 2015 teilnahmen, wurden bei der Kalibrierung der Item-Parameter beide Stichproben berücksichtigt; für die Schätzung wurde jede Erhebungsrunde und bei jeder Erhebungsrunde jedes Land bzw. jede Volkswirtschaft gleichermaßen berücksichtigt.
2. Die Schätzwerte wurden für alle Schüler und alle Schulen durch Verankerung der im vorangegangenen Schritt ermittelten Itemparameter berechnet.
3. Bei **neuen Skalenindizes** wurden die Warm-Likelihood-Estimates dann so standardisiert, dass der mittlere Indexwert für die OECD-Schülerpopulation 0 entspricht und die Standardabweichung 1 beträgt (wobei beim Standardisierungsprozess alle Länder gleich gewichtet wurden). **Trendindizes** wurden so angeglichen, dass für die OECD-Länder der Mittelwert und die Standardabweichung der umskalierten Schätzungen von PISA 2006 und der ursprünglichen Schätzungen der PISA-2006-Datenbank übereinstimmen. Die Trendindizes werden daher auf der ursprünglich in PISA 2006 verwendeten Skala dargestellt, damit die Indexwerte direkt mit jenen in der PISA-2006-Datenbank verglichen werden können.

Den verschiedenen Antwortkategorien der Fragen wurden in der Reihenfolge, in der letztere in den Schüler-, Schulleiter- und Elternfragebogen erschienen, sequenzielle Codes zugewiesen. Diese Codes wurden zum Zweck der Konstruktion der Indizes bzw. Skalen z.T. umgepolt; wenn dies der Fall war, wird in diesem Abschnitt jeweils darauf hingewiesen. Negative Werte bei einem Index lassen nicht zwangsläufig auf negative Antworten der Schülerinnen und Schüler auf die gestellten Fragen schließen. Ein negativer Wert weist lediglich darauf hin, dass die betreffenden Befragten weniger positiv antworteten, als der Durchschnitt der Befragten in den OECD-Ländern. Dementsprechend bedeutet ein positiver Wert bei einem Index, dass die jeweiligen Befragten positivere Antworten gaben als der Durchschnitt der Befragten in den OECD-Ländern. Die bei den folgenden Beschreibungen in Klammern < > gesetzten Begriffe wurden in den nationalen Fassungen der Schüler-, Schulleiter- und Elternfragebogen durch den entsprechenden nationalen Ausdruck ersetzt. So wurde z.B. der Begriff <Abschluss entsprechend ISCED-Stufe 5A> in den Vereinigten Staaten übersetzt in „Bachelor’s degree, post-graduate certificate program, Master’s degree program



or first professional degree program“. Desgleichen wurde der Ausdruck <Testsprachenunterricht> in Luxemburg übersetzt in „Deutschunterricht“ oder „Französischunterricht“, je nachdem ob die Schülerinnen und Schüler die deutsche oder die französische Fassung der Erhebungsinstrumente erhielten.

Neben den in diesem Anhang beschriebenen einfachen Indizes und Skalenindizes wurde in diesem Band eine Reihe von Variablen aus den Fragebogen verwendet, die Einzelitems entsprechen und bei der Konstruktion der Indizes nicht berücksichtigt wurden. Diese nicht umkodierte Variablen weisen das Präfix „ST“ für die Items im Schülerfragebogen und „SC“ für die Items im Schulleiterfragebogen auf. Alle Kontextfragebogen und die internationale PISA-Datenbank mit sämtlichen Variablen sind auf www.oecd.org/pisa verfügbar.

Einfache Indizes auf Schülerebene

Alter der Schülerinnen und Schüler

Das Alter einer Schülerin bzw. eines Schülers (AGE) wurde als Differenz zwischen dem Monat und dem Jahr der Durchführung der Tests und dem Geburtsmonat und -jahr der Schülerinnen und Schüler berechnet. Die Daten zum Alter der Schülerinnen und Schüler wurden sowohl dem Schülerfragebogen (ST003) als auch den Unterlagen über den bisherigen Bildungsweg der Schülerinnen und Schüler entnommen. Wenn bei einer Schülerin bzw. einem Schüler nicht bekannt war, in welchem Monat der Test durchgeführt wurde, wurde für die Berechnung der für dieses Land ermittelte Median herangezogen.

Bildungsniveau der Eltern

Die Antworten der Schülerinnen und Schüler auf die Fragen ST005, ST006, ST007 und ST008 zum Bildungsstand der Eltern wurden nach ISCED 1997 klassifiziert (OECD, 1999). Für die Konstruktion der Indizes zum Bildungsniveau der Eltern wurden die Bildungsabschlüsse in folgende Kategorien umkodiert: (0) Kein Abschluss, (1) <ISCED-Stufe 1> (Primarbereich), (2) <ISCED-Stufe 2> (Sekundarbereich I), (3) <ISCED-Stufe 3B oder 3C> (berufsbildender/vorberuflicher Sekundarbereich II), (4) <ISCED-Stufe 3A> (allgemeinbildender Sekundarbereich II) und/oder <ISCED-Stufe 4> (nichttertiärer postsekundärer Bereich), (5) <ISCED-Stufe 5B> (berufsbildender Tertiärbereich) und (6) <ISCED-Stufe 5A> und/oder <ISCED-Stufe 6> (theoretisch orientierter Tertiär- und Postgraduiertenbereich). Indizes mit diesen Kategorien wurden für die Mutter (MISCED) und den Vater (FISCED) der Schülerinnen und Schüler erstellt. Der Index des höchsten Bildungsabschlusses der Eltern (HISCED) entspricht der ISCED-Stufe des Elternteils mit dem jeweils höheren Bildungsabschluss. Der Index des höchsten Bildungsabschlusses der Eltern wurde zudem in die geschätzte Anzahl der Ausbildungsjahre (PARED) umkodiert. Die Entsprechungen zwischen dem Bildungsniveau und der Zahl der Ausbildungsjahre sind dem *PISA 2015 Technical Report* (OECD, erscheint demnächst) zu entnehmen.

Höchste berufliche Stellung der Eltern

Die berufsspezifischen Daten über die Väter und Mütter der Schülerinnen und Schüler wurden mithilfe von offenen Fragen erhoben. Die Antworten wurden in vierstelligen ISCO-Codes (ILO, 2007) kodiert und dann auf dem internationalen sozioökonomischen Index der beruflichen Stellung (ISEI) (Ganzeboom & Treiman, 2003) abgebildet. Wie in PISA 2012 wurde in PISA 2015 auf die neuen ISCO- bzw. ISEI-Fassungen aus dem Jahr 2008 zurückgegriffen und nicht auf jene aus dem Jahr 1988, die in den vorangegangenen Erhebungsrunden verwendet worden waren (Ganzeboom, 2010). Auf der Grundlage dieser Daten wurden drei Indizes erstellt: der Index der beruflichen Stellung des Vaters (BFMJ2), der Index der beruflichen Stellung der Mutter (BMMJ1) und der Index der höchsten beruflichen Stellung der Eltern (HISEI), der dem ISEI-Wert des Elternteils mit der jeweils höheren beruflichen Stellung bzw. dem einzigen verfügbaren ISEI-Wert eines Elternteils, entspricht. Bei allen drei Indizes zeigen höhere ISEI-Werte eine höhere berufliche Stellung an.

Migrationshintergrund

Die PISA-Datenbank enthält drei länderspezifische Variablen, die sich auf das Geburtsland der Schülerin bzw. des Schülers, der Mutter und des Vaters beziehen (COBN_S, COBN_M, und COBN_F). Die Items ST019Q01TA, ST019Q01TB und ST019Q01TC wurden in folgende Kategorien umkodiert: (1) Geburtsland entspricht dem Erhebungsland und (2) einem anderen Land. Anhand dieser Variablen wurde der Index des Migrationshintergrunds (IMMIG) mit folgenden Kategorien konstruiert: (1) Schülerinnen und Schüler ohne Migrationshintergrund (Schülerinnen und Schüler mit mindestens einem im Erhebungsland geborenen Elternteil), (2) Schülerinnen und Schüler mit Migrationshintergrund der zweiten Generation (im Erhebungsland geborene Schülerinnen und Schüler mit in einem anderen Land geborenen Eltern) und (3) Schülerinnen und Schüler mit Migrationshintergrund der ersten Generation (nicht im Erhebungsland geborene Schülerinnen und Schüler, deren Eltern ebenfalls in einem anderen Land geboren sind). Schülerinnen und Schülern, die zu ihrem eigenen Geburtsland, dem ihrer Mutter oder ihres Vaters keine Angaben machten, wurden für diese Variable fehlende Werte zugewiesen.

Die zu Hause gesprochene Sprache

Die Schülerinnen und Schüler gaben an, welche Sprache sie für gewöhnlich zu Hause sprechen (ST022), und die Datenbank enthält eine abgeleitete Variable (LANGN) mit einem länderspezifischen Code für die jeweilige Sprache. Von dieser Information wurde zudem eine international vergleichbare Variable (ST022Q01TA) mit den folgenden Kategorien abgeleitet: (1) bei diesem Schüler ist die zu Hause gesprochene Sprache mit der Testsprache identisch und (2) die zu Hause gesprochene Sprache ist nicht mit der Testsprache identisch.

Klassenwiederholung

Die Variable zur Klassenwiederholung (REPEAT) wurde durch eine Umkodierung der Variablen ST127Q01TA, ST127Q02TA und ST127Q03TA berechnet. REPEAT erhielt den Wert „1“, wenn die Schülerin bzw. der Schüler in mindestens einer ISCED-Stufe eine Klasse wiederholt hatte, und den Wert „0“, wenn mindestens einmal die Antwort „nein, nie“ und keine der Antwortkategorien für Klassenwiederholung gewählt wurde. Wenn für eine der Stufen keine der drei Antwortkategorien ausgewählt wurde, wurde für den Index ein fehlender Wert zugewiesen.

Bildungsgang

Im Rahmen von PISA werden Daten zu den in den einzelnen Ländern für 15-jährige Schülerinnen und Schüler angebotenen Bildungsgängen erhoben. Diese Daten werden anhand der Unterlagen über den bisherigen Bildungsweg der Schülerinnen und Schüler und anhand des Schülerfragebogens ermittelt. In der endgültigen Datenbank sind alle nationalen Bildungsgänge in einer separaten abgeleiteten Variablen (PROGN) enthalten, wobei die ersten sechs Ziffern dem Code des nationalen Zentrums und die letzten beiden Ziffern dem nationalen Code für den jeweiligen Bildungsgang entsprechen. Alle Bildungsgänge wurden nach der Internationalen Standardklassifikation des Bildungswesens (ISCED) klassifiziert (OECD, 1999). Aus den Daten über die Bildungsgänge wurden folgende Indizes abgeleitet:

- Der Index der Stufe des Bildungsgangs (ISCEDL) gibt an, ob sich die Schülerinnen und Schüler in Sekundarstufe I oder II befinden (ISCED 2 oder ISCED 3).
- Der Index der Bezeichnung des Bildungsgangs (ISCEDD) gibt die Bezeichnung des Bildungsgangs an (A = allgemeinbildender Zweig, der den Zugang zur nächsthöheren Bildungsstufe ermöglicht, B = Bildungsgang, der den Zugang zur nächsthöheren berufsbildenden Stufe ermöglicht, C = Bildungsgang, der auf den direkten Arbeitsmarktzugang ausgerichtet ist, M = modularer Bildungsgang, der mehrere dieser Merkmale kombiniert).
- Der Index der Lehrplanausrichtung (ISCEDO) gibt an, ob die Lehrpläne des betreffenden Bildungsgangs allgemeinbildend, berufsvorbereitend oder berufsbildend sind.

Naturwissenschaftlich orientierte Berufsvorstellungen

In PISA 2015 wurden die Schülerinnen und Schüler aufgefordert, die Frage (ST114) zu beantworten, „welchen Beruf [sic ihrer Ansicht nach] mit ca. 30 Jahren haben [werden]“. Die Antworten auf diese offene Frage wurden in der Variablen OCOD3 in vierstelligen ISCO-Codes kodiert (ILO, 2007). Von dieser Variablen wurde der Index naturwissenschaftlich orientierter Berufsvorstellungen abgeleitet.

Als naturwissenschaftlich orientierte Berufsvorstellungen gelten solche, deren Verwirklichung eine über die Pflichtschulzeit hinausgehende naturwissenschaftliche Ausbildung, in der Regel im Rahmen der formalen Hochschulbildung, erfordert. Die Klassifizierung der Berufe in Gruppen mit und ohne Naturwissenschaftsbezug basiert auf dem vierstelligen Code der Internationalen Standardklassifikation der Berufe aus dem Jahr 2008 (ISCO-08).

Nur akademische Berufe (ISCO-Berufshauptgruppe 2) sowie Techniker und gleichrangige nichttechnische Berufe (ISCO-Berufshauptgruppe 3) wurden als Berufe eingestuft, die der Definition naturwissenschaftlich orientierter Berufsvorstellungen entsprechen. Mehrere Berufe der Berufshauptgruppe „Führungskräfte“ (ISCO-Hauptgruppe 1) sind im weiteren Sinn eindeutig naturwissenschaftlich orientiert. Dazu zählen u.a. Führungskräfte in Forschung und Entwicklung, Führungskräfte in Krankenhäusern, Führungskräfte in der Produktion im Bau und andere in der Gruppe „Führungskräfte in der Produktion und bei speziellen Dienstleistungen“ klassifizierte Berufe (Berufsgruppe 13). Dabei wurde jedoch unterstellt, dass es sich bei Führungspositionen, für die sowohl eine naturwissenschaftliche Berufserfahrung als auch Ausbildung wichtige Grundvoraussetzung ist, nicht um Positionen für Berufsanfänger handelt, und 15-jährige Schülerinnen und Schüler mit naturwissenschaftlich orientierten Berufsvorstellungen daher nicht annehmen würden, mit 30 Jahren eine solche Tätigkeit auszuüben.

Auch bei mehreren Arten von Fachkräften in der Land- und Forstwirtschaft und Fischerei (ISCO-Berufshauptgruppe 6) könnte angenommen werden, dass sie Tätigkeiten mit naturwissenschaftlichem Bezug ausüben. In O*NET OnLine (2016), einer US-amerikanischen Klassifizierung von Berufen mit Bezug zu Mathematik, Informatik, Naturwissenschaften und Technik (MINT) werden diese Berufe tatsächlich berücksichtigt. Für diese Tätigkeiten ist jedoch in der Regel keine formale naturwissenschaftsbezogene Berufsausbildung nach der Pflichtschulzeit bzw. kein naturwissenschaftliches Hochschulstudium erforderlich. Bei den naturwissenschaftlich orientierten Berufsvorstellungen wurden indes lediglich Berufshauptgruppen berücksichtigt, die das ISCO-Anforderungsniveau 3 und 4 voraussetzen.

Bei akademischen Berufen und Technikern bzw. gleichrangigen nichttechnischen Berufen ist die Trennlinie zwischen Berufen mit und ohne Naturwissenschaftsbezug zuweilen unscharf, und in den verschiedenen Klassifikationen werden unterschiedliche Differenzierungen vorgenommen.

Die in diesem Bericht verwendete Klassifizierung umfasst vier Berufsgruppen¹:

1. **Naturwissenschaftler, Mathematiker und Ingenieure:** Alle Naturwissenschaftler, Mathematiker und Ingenieure (Berufsgruppe 21) außer Produkt- und Textildesigner (2163) sowie Grafik- und Multimediadesigner (2166).



2. **Akademische und verwandte Gesundheitsberufe:** Alle Gesundheitsberufe der Berufsgruppe 22 (z.B. Ärzte, Krankenpflegekräfte, Tierärzte) außer „akademische und vergleichbare Fachkräfte in der traditionellen und komplementären Medizin“ (Berufsuntergruppe 223).
3. **IKT-Fachkräfte:** Alle akademischen und vergleichbaren Fachkräfte in der Informations- und Kommunikationstechnologie (Berufsgruppe 25).
4. **Techniker und gleichrangige nichttechnische Berufe mit Naturwissenschaftsbezug,** darunter:
 - Material- und ingenieurtechnische Fachkräfte (Berufsuntergruppe 311)
 - Biotechniker und verwandte technische Berufe (Berufsuntergruppe 314)
 - Flugsicherungstechniker (3155)
 - Medizinische und pharmazeutische Fachberufe (Berufsuntergruppe 321), außer medizinische und zahntechnische Prothetiktechniker (3214)
 - Telekommunikationstechniker (3522).

Unterschiede zwischen der vorliegenden und anderen Klassifizierungen

Vergleicht man drei bestehende Klassifizierungen von Berufsvorstellungen 15-Jähriger, die auf der Internationalen Standardklassifikation der Berufe aus dem Jahr 1988 basieren (ISCO-88), mit der vorliegenden ISCO-08-basierten Klassifizierung, lassen sich einige Unterschiede erkennen. Einige davon resultieren (wie im folgenden Abschnitt erörtert) aus der aktualisierten Berufskodierung. Die übrigen sind in Tabelle A1.1 zusammengefasst.

Erstellung einer vergleichbaren Klassifizierung für ISCO-88

Auch der Fragebogen von PISA 2006 enthielt die vorstehend genannte offene Frage (ID im Jahr 2006: ST30), in der Datenbank von PISA 2006 wurden die Schülerantworten jedoch nach ISCO-88 kodiert. Dass die Klassifizierung vollständig vergleichbar ist, kann nicht gewährleistet werden. Damit die Veränderungen im Zeitverlauf erfasst werden können, wurden die in Tabelle A1.2 beschriebenen Entsprechungen zugrunde gelegt, um eine ähnliche Klassifizierung auf Basis der Daten von PISA 2006 zu ermöglichen.

Tabelle A1.1 ▪ **Unterschiede bei der Definition naturwissenschaftlich orientierter Berufsvorstellungen**

	Vorliegende Klassifizierung	OECD (2007)	Sikora und Pokropek (2012)	Kjærnsli und Lie (2011)
Führungspositionen mit Naturwissenschaftsbezug	nein	ja	ja	nein
Psychologen	nein	ja	ja	nein
Soziologen und Sozialarbeiter	nein	ja	nein	nein
Fotografen, Bediener von Bild- und Tonaufzeichnungsanlagen, Fernseh-, Rundfunk- und Fernmeldeanlagenbediener	nein	ja	ja	nein
Nicht akademische statistische, mathematische und verwandte Fachkräfte	nein	nein	ja	nein
Flugzeugführer und verwandte Berufe (z.B. Piloten, Flugverkehrslotsen)	nein	ja	ja	nein
Schiffsführer und verwandte Berufe (nautische Schiffsoffiziere usw.)	nein	nein	ja	nein
Medizinische Assistenten, zahnmedizinische Assistenten, veterinärmedizinische Assistenten, nicht akademische Krankenpflege- und Geburtshilfefachkräfte	nein	ja	ja	nein
Datenverarbeitungsassistenten, EDV-Operateure, Steuerer von Industrierobotern	nein	nein	nein	ja
Flugsicherungstechniker	ja	ja	ja	nein
Pharmazeutisch-technische Assistenten	ja	ja	ja	nein
Diätologen und Ernährungsberater	ja	ja	ja	nein

Tabelle A1.2 ■ **Entsprechungen für naturwissenschaftlich orientierte Berufsvorstellungen in ISCO-08 und ISCO-88**

Gruppe	ISCO-08	ISCO-88
<i>Naturwissenschaftler, Mathematiker und Ingenieure</i>	21xx (außer 2163 und 2166)	21xx (außer 213x), 221x
<i>Gesundheitsberufe</i>	22xx (außer 223x)	22xx (außer 221x), 3223, 3226
<i>IKT-Fachkräfte</i>	25xx	213x
<i>Techniker und gleichrangige nichttechnische Berufe mit Naturwissenschaftsbezug</i>	311x, 314x, 3155, 321x (außer 3214), 3522	311x, 3133, 3145, 3151, 321x, 3228

Im Hinblick auf die Konstruktion des Index naturwissenschaftlich orientierter Berufsvorstellungen sind zwischen ISCO-88 und ISCO-08 folgende Hauptunterschiede festzustellen:

- „Bediener medizinischer Geräte“ (ISCO-88: 3133) entspricht in ISCO-08 der Berufsuntergruppe „Medizintechniker im Bereich bildgebende Verfahren und Therapiegeräte“; „Bau-, Brandschutz-, Brandinspektoren“ (ISCO-88: 3151) entspricht in ISCO-08 weitestgehend der Berufsuntergruppe „Bautechniker“.
- „Diätassistenten und Ernährungsberater“ (ISCO-88: 3223) werden in ISCO-08 als akademische Berufe eingestuft. Aus Gründen der Konsistenz wurde diese ISCO-88-Berufsuntergruppe den Gesundheitsberufen zugerechnet.
- Die Untergruppe „Physiotherapeuten und verwandte Berufe“ (ISCO-88: 3226) verteilt sich in ISCO-08 auf zwei verschiedene Kategorien, wobei Physiotherapeut den akademischen Berufen zugeordnet wurde. Angesichts der Tatsache, dass sich Schülerinnen und Schüler wesentlich häufiger als künftige Physiotherapeuten sehen als in verwandten Berufen, wurde diese ISCO-88-Berufsuntergruppe den Gesundheitsberufen zugerechnet.
- Mehrere Gesundheitsberufe, die in ISCO-88 als „moderne medizinische Fachberufe“ eingestuft wurden, werden in ISCO-08 den „akademischen und verwandten Gesundheitsberufen“ zugerechnet (z.B. Sprachtherapeuten, Orthoptisten). Während „akademische und verwandte Gesundheitsberufe“ im Allgemeinen als Berufe mit Naturwissenschaftsbezug eingestuft werden, ist dies bei „Assistenzberufen im Gesundheitswesen“ („Gesundheitsfachkräfte“ bzw. „medizinische Fachberufe“ in ISCO-88) nicht der Fall. Bei der Anwendung der Klassifizierung auf ISCO-88 wurden die unter diesem Code erfassten Berufe nicht den Berufen mit Naturwissenschaftsbezug zugerechnet.
- „Telekommunikationstechniker“ (ISCO-08: 3522) bilden in ISCO-88 keine eigene Untergruppe. Dort sind sie der Untergruppe „Elektronik- und Fernmeldetechniker“ (ISCO-88: 3114) zugeordnet.
- „Akademische und vergleichbare Fachkräfte in der Informations- und Kommunikationstechnologie“ bilden in ISCO-08 eine eigenständige Berufsgruppe (25), während sie in ISCO-88 der Berufsgruppe „Physiker, Mathematiker und Ingenieurwissenschaftler“ zugeordnet wurden.

Skalenindizes auf Schülerebene

Neue Skalenindizes

Interesse an Naturwissenschaften

Der Index des Interesses an naturwissenschaftlichen Themen (INTBRSCI) wurde anhand der Antworten der Schülerinnen und Schüler auf eine für PISA 2015 neu entwickelte Frage (ST095) konstruiert. Die Schülerinnen und Schüler gaben mittels einer fünfstufigen Likert-Skala mit den Antwortkategorien „überhaupt nicht interessiert“, „kaum interessiert“, „eher interessiert“, „sehr interessiert“ und „Ich weiß nicht, was das ist“ Auskunft über ihr Interesse an den folgenden Themen: „Lebensräume (z.B. Ökosysteme, Nachhaltigkeit)“, „Bewegung und Kräfte (z.B. Geschwindigkeit, Reibung, Magnetismus, Schwerkraft)“, „Energie und ihre Umwandlung (z.B. Konservierung, chemische Reaktionen)“, „Das Universum und seine Geschichte“ und „Wie Naturwissenschaften uns helfen können, Krankheiten zu verhindern“. Die letzte Antwortkategorie („Ich weiß nicht, was das ist“) wurde für die Zwecke der Konstruktion des INTBRSCI-Index als fehlend umkodiert. Höhere Indexwerte signalisieren ein größeres Interesse an diesen Themen.

Epistemische Überzeugungen in Bezug auf Naturwissenschaften

Der Index der epistemischen Überzeugungen in Bezug auf Naturwissenschaften (EPIST) wurde anhand der Antworten der Schülerinnen und Schüler auf eine für PISA 2015 neu entwickelte Frage nach ihren Ansichten zu naturwissenschaftlichen Forschungsansätzen (ST131) konstruiert. Die Schülerinnen und Schüler gaben auf einer vierstufigen Likert-Skala mit den Antwortkategorien „stimme überhaupt nicht zu“, „stimme eher nicht zu“, „stimme eher zu“ und „stimme völlig zu“ an, inwieweit sie den folgenden Aussagen zustimmen: „Eine gute Möglichkeit herauszufinden, ob etwas wahr ist, ist ein Experiment durchzuführen“, „Naturwissenschaftliche Erkenntnisse ändern sich manchmal“, „Gute Antworten basieren auf Befunden



aus vielen unterschiedlichen Experimenten“, „Es ist gut, Experimente mehrmals durchzuführen, um sicherzustellen, dass die Ergebnisse stimmen“, „Manchmal ändern Wissenschaftlerinnen und Wissenschaftler ihre Meinung über das, was in den Naturwissenschaften wahr ist“ und „Erkenntnisse in naturwissenschaftlichen Fachbüchern ändern sich manchmal“. Höhere Indexwerte signalisieren eine stärkere Zustimmung zu diesen Aussagen.

Trendbezogene Skalenindizes

Freude an Naturwissenschaften

Der Index der Freude an Naturwissenschaften (JOYSCIE) wurde auf Basis einer Trendfrage (ST094) aus PISA 2006 (ID bei PISA 2006: ST16) konstruiert, bei der die Schülerinnen und Schüler auf einer vierstufigen Likert-Skala mit den Kategorien „stimme völlig zu“, „stimme eher zu“, „stimme eher nicht zu“ oder „stimme überhaupt nicht zu“ angeben sollten, inwieweit sie den folgenden Aussagen zustimmen: „Im Allgemeinen macht es mir Spaß, mich mit naturwissenschaftlichen Themen zu befassen“, „Ich lese gerne etwas über Naturwissenschaften“, „Ich beschäftige mich gerne mit naturwissenschaftlichen Problemen“, „Ich eigne mir gerne neues Wissen in den Naturwissenschaften an“ und „Ich bin interessiert, Neues in den Naturwissenschaften zu lernen“. Die abgeleitete Variable JOYSCIE wurde auf die entsprechende Skala in der PISA-2006-Datenbank abgestimmt, wodurch ein Trendvergleich zwischen PISA 2006 und PISA 2015 ermöglicht wurde. Höhere Indexwerte signalisieren eine stärkere Zustimmung zu diesen Aussagen.

Selbstwirksamkeit in Naturwissenschaften

Der Index der Selbstwirksamkeit in Naturwissenschaften (SCIEEFF) wurde auf Basis einer Trendfrage (ST129) konstruiert, die aus PISA 2006 (ID in PISA 2006: ST17) übernommen wurde. Die Schülerinnen und Schüler wurden gefragt, wie sie auf einer vierstufigen Antwortskala mit den Kategorien „Das wäre einfach für mich“, „Ich könnte das mit ein bisschen Mühe schaffen“, „Es würde mir schwerfallen, das allein zu schaffen“ und „Das könnte ich nicht“ ihre Erfolgchancen bei den folgenden naturwissenschaftlichen Aufgabenstellungen einschätzen: „Die naturwissenschaftliche Fragestellung erkennen, die einem Zeitungsbericht über ein Gesundheitsthema zugrunde liegt“, „Erklären, warum Erdbeben in manchen Gegenden häufiger vorkommen als in anderen“, „Die Rolle der Antibiotika bei der Behandlung von Krankheiten beschreiben“, „Wissenschaftliche Fragestellungen herausfinden, die mit der Müllentsorgung zusammenhängen“, „Vorhersagen, wie Änderungen in der Natur das Überleben bestimmter Tierarten beeinflussen können“, „Die wissenschaftlichen Informationen auf einem Lebensmitteletikett interpretieren“, „Zeigen, wie neue Erkenntnisse zu einem neuen Verständnis über die Möglichkeit von Leben auf dem Mars führen können“ und „Die bessere von zwei Erklärungen über die Bildung von saurem Regen erkennen“. Die Antworten wurden umgepolt, so dass höhere Indexwerte einer höheren Selbstwirksamkeit in Naturwissenschaften entsprechen. Die abgeleitete Variable SCIEEFF wurde auf die entsprechende Skala in der PISA-2006-Datenbank abgestimmt, wodurch ein Trendvergleich zwischen PISA 2006 und PISA 2015 ermöglicht wurde.

Naturwissenschaftliche Aktivitäten

Der Index naturwissenschaftlicher Aktivitäten (SCIEACT) wurde auf Basis einer Trendfrage (ST146) aus PISA 2006 (ID in PISA 2006: ST19) konstruiert. Die Schülerinnen und Schüler wurden gebeten, auf einer vierstufigen Skala mit den Antwortkategorien „sehr oft“, „regelmäßig“, „manchmal“, „nie oder fast nie“ anzugeben, wie oft sie sich mit den folgenden naturwissenschaftsbezogenen Aktivitäten beschäftigen: „Fernsehsendungen über Naturwissenschaften anschauen“, „Bücher über naturwissenschaftliche Themen ausborgen oder kaufen“, „Internetseiten zu naturwissenschaftlichen Themen besuchen“, „Naturwissenschaftliche Zeitschriften oder Artikel in Zeitungen lesen“, „Eine Naturwissenschafts-AG besuchen“, „Naturphänomene mithilfe von Computerprogrammen/virtuellen Labors simulieren“, „Technische Prozesse mithilfe von Computerprogrammen/virtuellen Labors simulieren“, „Internetseiten von Umweltorganisationen besuchen“ und „Nachrichten von Naturwissenschafts- oder Umweltorganisationen über Blogs und Mikroblogging verfolgen“. Die Antworten wurden umgepolt, so dass höhere Indexwerte einer stärkeren Beschäftigung mit naturwissenschaftlichen Aktivitäten entsprechen. Die abgeleitete Variable SCIEACT wurde auf die entsprechende Skala in der PISA-2006-Datenbank abgestimmt, wodurch ein Trendvergleich zwischen PISA 2006 und PISA 2015 ermöglicht wurde.

Instrumentelle Lernmotivation in Naturwissenschaften

Der Index der instrumentellen Lernmotivation in Naturwissenschaften (INSTSCIE) wurde auf Basis einer Trendfrage (ST113) aus PISA 2006 (ID in PISA 2006: ST35) konstruiert. Die Schülerinnen und Schüler gaben auf einer vierstufigen Likert-Skala mit den Kategorien „stimme völlig zu“, „stimme eher zu“, „stimme eher nicht zu“ und „stimme überhaupt nicht zu“ an, inwieweit sie den folgenden Aussagen zustimmen: „Sich im Unterricht in naturwissenschaftlichen Fächern anzustrengen, zahlt sich aus, weil mir das bei der Arbeit, die ich später machen möchte, helfen wird“, „Was ich in den naturwissenschaftlichen Fächern lerne, ist wichtig für mich, weil ich es für das brauche, was ich später machen möchte“, „Für naturwissenschaftliche Unterrichtsfächer zu lernen, lohnt sich für mich, weil das Gelernte meine beruflichen Aussichten verbessern wird“ und „Viele Dinge, die ich in den naturwissenschaftlichen Unterrichtsfächern lerne, werden mir dabei helfen, einen Job zu bekommen“. Die Antworten wurden umgepolt, so dass höhere Indexwerte einer höheren instrumentellen Lernmotivation in Naturwissenschaften entsprechen. Die abgeleitete Variable INSTSCIE wurde auf die entsprechende Skala in der PISA-2006-Datenbank abgestimmt, wodurch ein Trendvergleich zwischen PISA 2006 und PISA 2015 ermöglicht wurde.

Skalierung von Indizes im Zusammenhang mit dem PISA-Index des wirtschaftlichen, sozialen und kulturellen Status

Der PISA-Index des wirtschaftlichen, sozialen und kulturellen Status (ESCS) wurde wie in früheren Erhebungsrounden aus drei Variablen des familiären Hintergrunds abgeleitet: dem höchsten Bildungsabschluss der Eltern (PARED), der höchsten beruflichen Stellung der Eltern (HISEI) und der Ausstattung des Elternhauses (HOMEPOS), einschließlich der Anzahl der Bücher im Haushalt. Bei PARED und HISEI handelt es sich um einfache Indizes, wie vorstehend beschrieben. HOMEPOS ist ein Hilfsindikator für den Wohlstand der Familie.

Ausstattung des Haushalts

Bei PISA 2015 gaben die Schülerinnen und Schüler Auskunft über das Vorhandensein von 16 Haushaltsitems in ihrem Zuhause (ST011), darunter drei länderspezifische Items, die als geeignete Messgrößen für den Wohlstand der Familie im Kontext des jeweiligen Landes angesehen wurden. Zusätzlich gaben die Schülerinnen und Schüler an, in welchem Umfang bestimmte Items vorhanden sind und über wie viele Bücher der Haushalt verfügt (ST012, ST013).

HOMEPOS ist ein Summenindex aller Haushalts- und Ausstattungsgüter (ST011, ST012 und ST013). Die Skala der Ausstattung des Elternhauses wurde für PISA 2015 anders berechnet als in früheren Erhebungsrounden, um das IRT-Modell auf das Modell abzustimmen, das für alle kognitiven und nicht kognitiven Skalen verwendet wurde. Die Kategorien für die Anzahl der Bücher im Haushalt blieben in PISA 2015 unverändert. Die ST011-Items (1 = „ja“, 2 = „nein“) wurden umgepolt, so dass ein höherer Wert die Präsenz des Indikators signalisiert.

Berechnung des ESCS

Zur Berechnung des PISA-Index des wirtschaftlichen, sozialen und kulturellen Status (ESCS) wurden die Werte für Schülerinnen und Schüler mit fehlendem PARED, HISEI oder HOMEPOS durch Imputation mit geschätzten Werten sowie einer Zufallskomponente auf Basis einer Regression auf die anderen beiden Variablen ermittelt. Wenn bei mehr als einer der drei Variablen Daten fehlten, wurde der ESCS nicht berechnet; stattdessen wurde für den ESCS ein fehlender Wert ausgewiesen.

Der PISA-Index des wirtschaftlichen, sozialen und kulturellen Status wurde aus einer Hauptkomponentenanalyse standardisierter Variablen (in der jede Variable einen OECD-Mittelwert von 0 und eine Standardabweichung von 1 hat) abgeleitet, wobei die Faktorwerte für die erste Hauptkomponente als Messgrößen des PISA-Index des wirtschaftlichen, sozialen und kulturellen Status dienten. Alle Länder und Volkswirtschaften (sowohl OECD- als auch Partnerländer und -volkswirtschaften) waren gleichermaßen an der Hauptkomponentenanalyse beteiligt, während in früheren Erhebungsrounden die Hauptkomponentenanalyse lediglich auf OECD-Ländern beruhte. Für die Darstellung der Ergebnisse wurde die ESCS-Skala jedoch so umgearbeitet, dass 0 der Wert für einen durchschnittlichen Schüler im OECD-Raum und 1 die Standardabweichung in den gleich gewichteten OECD-Ländern ist.

Die Hauptkomponentenanalyse wurde zudem für jedes Teilnehmerland bzw. jede Teilnehmervolkswirtschaft separat durchgeführt, um festzustellen, inwieweit die Indexkomponenten in den verschiedenen Ländern bzw. Volkswirtschaften auf ähnliche Art und Weise operieren.

Berechnung eines ESCS-Trendindex

Obwohl alle früheren PISA-Datenbanken einen Index des wirtschaftlichen, sozialen und kulturellen Status (ESCS) enthielten, veränderten sich die Komponenten des ESCS und das Skalierungsmodell zwischen den Erhebungsrounden, so dass ESCS-Werte verschiedener Erhebungsrounden nicht direkt miteinander verglichen werden können. Um eine Trendanalyse zu ermöglichen, wurde der ESCS in PISA 2015 für die aktuelle Erhebungsrounde berechnet und zudem für die früheren Erhebungsrounden unter Anwendung einer ähnlichen Methodik neu berechnet².

Vor einer Schätzung der Trendwerte mussten geringfügige Anpassungen an den drei Komponenten vorgenommen werden:

- Da bei PISA 2012 das Kodierungssystem für Berufe, auf dem HISEI beruht, von ISCO-88 auf ISCO-08 umgestellt wurde, wurden die Berufsklassifikationen früherer Erhebungsrounden aus dem früheren in das aktuelle System übertragen (vgl. auch *PISA 2012 Technical Report*, Kapitel 3).
- Um die Vergleichbarkeit der PARED-Komponente zwischen den verschiedenen Erhebungsrounden sicherzustellen, wurde für alle Runden das gleiche System zur Abbildung der ISCED-Stufen in PARED angewendet.
- Um die Vergleichbarkeit der HOMEPOS-Komponente zwischen den verschiedenen Erhebungsrounden zu verbessern, wurde die Variable *Anzahl der Bücher im Haushalt* (ST013Q01TA) in eine vierstufige kategoriale Variable umkodiert (0-25 Bücher, 26-100 Bücher, 101-500 Bücher, mehr als 500 Bücher). Die Konstruktion der HOMEPOS-Trendskala erfolgte in drei Schritten. Im ersten Schritt wurden durch eine simultane Kalibrierung der PISA-2015-Daten internationale Itemparameter für alle in PISA 2015 verwendeten Items (mit Ausnahme länderspezifischer Items, d.h. ST011Q17NA, ST011Q18NA und ST011Q19NA) generiert. Abgesehen von der Umkodierung der Variable ST013Q01TA entspricht dieser Schritt der regulären Skalierung von HOMEPOS in PISA 2015 (siehe oben). Im zweiten Schritt wurden einmalige Items aus allen vorherigen Erhebungsrounden (d.h. 2000-2012) skaliert, wobei die meisten 2015 verwendeten Items auf ihre PISA-2015-Parameter fixiert wurden, während ein begrenzter Satz von Itemparametern frei geschätzt werden konnte, mit der Einschränkung, dass sie für alle Länder innerhalb eines Erhebungszyklus gleich sein mussten. Nationale Items (d.h. ST011Q17NA, ST011Q18NA und ST011Q19NA) erhielten



durchweg einmalige (länder- und erhebungsspezifische) Parameter. Im dritten und letzten Schritt wurden Indexwerte (WLE) für alle Schülerinnen und Schüler aus früheren Erhebungsrunden (2000-2012) generiert. Da 17 von 27 bei der Berechnung der HOMEPOS-Trendkomponente berücksichtigten Items in den verschiedenen Erhebungsrunden dieselben Itemparameter haben, kann für die HOMEPOS-Trendwerte eine gemeinsame Skala unterstellt werden, was einen Ländervergleich über verschiedene Erhebungszyklen und somit eine Verwendung bei der Berechnung des Trend-ESCS ermöglicht.

Im Anschluss daran wurde die Hauptkomponentenanalyse zur Generierung der ESCS-Trendwerte wie oben beschrieben durchgeführt, wobei jedoch die Berechnung für alle Erhebungsrunden mit diesen drei vergleichbaren Komponenten erfolgte (Trend-HISEI, Trend-PARED und Trend-HOMEPOS).

Skalenindizes auf Schulebene

Schulische Ressourcen

PISA 2015 enthielt eine Frage mit acht Items zur Ressourcenausstattung der Schulen, mit der die Einschätzung der Schulleitungen zu einer potenziellen Beeinträchtigung des Unterrichts durch verschiedene Faktoren untersucht wurde („Wird der Unterricht an Ihrer Schule durch die folgenden Faktoren beeinträchtigt?“). Die vier Antwortkategorien lauteten „überhaupt nicht“, „kaum“, „bis zu einem gewissen Grad“ und „sehr“. Eine ähnliche Frage wurde auch in früheren Erhebungsrunden gestellt, doch die Items wurden für PISA 2015 verringert und umformuliert und auf zwei abgeleitete Variablen fokussiert. Der Index des Mangels an Bildungspersonal (STAFFSHORT) wurde aus den folgenden vier Items abgeleitet: „ein Mangel an Lehrkräften“, „ungenügend oder schlecht ausgebildete Lehrkräfte“, „ein Mangel an Hilfspersonal“, „ungenügend oder schlecht ausgebildetes Hilfspersonal“. Der Index des Mangels an Bildungsmaterialien (EDUSHORT) wurde auf Basis der folgenden vier Items skaliert: „fehlendes Unterrichtsmaterial (z.B. Schulbücher, IT-Ausstattung, Bibliotheks- oder Laborausstattung)“, „unzulängliches oder schlechtes Unterrichtsmaterial (z.B. Schulbücher, IT-Ausstattung, Bibliotheks- oder Laborausstattung)“, „fehlende materielle Infrastruktur (z.B. Gebäude, Außenanlagen, Heizung/Kühlung, Licht- und Akustikanlagen)“, „unzulängliche oder schlechte materielle Infrastruktur (z.B. Gebäude, Außenanlagen, Heizung/Kühlung, Licht- und Akustikanlagen)“. Positive Werte auf diesen Indizes bedeuten, dass die Schulleitung den Umfang und/oder die Qualität der Ressourcen in ihrer Schule stärker als im OECD-Durchschnitt als Hindernis für den Unterricht betrachtet; negative Werte bedeuten, dass die Schule nach Ansicht der Schulleitung weniger als im OECD-Durchschnitt durch fehlende oder unzulängliche Ressourcen beeinträchtigt wird.

Anteil fehlender Beobachtungen für die in diesem Band verwendeten Variablen

Sofern nicht anders angegeben, wird bei den in diesem Band enthaltenen Analysen keine Anpassung für fehlende Antworten auf Fragebogen vorgenommen. Die angegebenen Prozentsätze und Schätzungen auf Basis von Indizes beziehen sich auf den Anteil der Stichprobe mit gültigen Antworten auf die entsprechenden Fragebogenitems. Tabelle A1.3 (online verfügbar) gibt an, welcher Anteil der Stichprobe von Analysen auf Basis von Schüler- oder Schulleiterfragebogenvariablen erfasst wurde. In Fällen, in denen dieser Anteil zwischen den einzelnen Ländern bzw. Volkswirtschaften oder im Zeitverlauf sehr unterschiedlich ausfällt, ist beim Vergleich von Ergebnissen anhand dieser Dimensionen Vorsicht geboten.

Online verfügbare Tabellen

Table A1.3. Weighted share of responding students covered by analyses based on questionnaires
<http://dx.doi.org/10.1787/888933433112>

Anmerkungen

1. Im Vereinigten Königreich (ohne Schottland) wurden Berufsvorstellungen lediglich anhand des dreistelligen ISCO-Codes kodiert. Daher wurden die Berufe „Produkt- und Textildesigner“ (ISCO-08: 2163) sowie „Grafik- und Multimediadesigner“ (2166) der Berufsgruppe „Naturwissenschaftler, Mathematiker und Ingenieure“ zugerechnet und „medizinische und zahnmedizinische Prothetiktechniker“ (3214) der Berufsgruppe „Techniker und gleichrangige nichttechnische Berufe mit Naturwissenschaftsbezug“, während „Telekommunikationstechniker“ (3522) nicht berücksichtigt wurde. In diesen Berufen sieht sich nur ein kleiner Anteil der Schülerinnen und Schüler mit naturwissenschaftlich orientierten Berufsvorstellungen, so dass dies keine starken Auswirkungen auf die Ergebnisse hat.

2. Aufgrund dieses Verfahrens existieren für 2015 zwei Indizes (ESCS und Trend-ESCS). Die Pearson-Korrelation zwischen den beiden Indizes ist $r = 0,989$ für alle an PISA 2015 teilnehmenden Länder und Volkswirtschaften. Dies umfasst 22 Länder bzw. Volkswirtschaften, in denen die Korrelation $r > 0,990$ war, 50 andere Länder bzw. Volkswirtschaften mit $r = [0,960, 0,990]$ sowie ein weiteres Land (Georgien) mit $r = 0,946$. In Kapitel 6 und 7 wird bei den Ergebnissen für 2015, die sich auf Trends beim ESCS beziehen, anstatt des ESCS-Trendindex für 2015 der ESCS-Index für 2015 verwendet, um die Konsistenz zwischen den Tabellen zu bewahren.

Literaturverzeichnis

Ganzeboom, H.B.G. (2010), „A new international socio-economic index [ISEI] of occupational status for the International Standard Classification of Occupation 2008 [ISCO-08] constructed with data from the ISSP 2002-2007; with an analysis of quality of occupational measurement in ISSP“ Paper für das Jahrestreffen des International Social Survey Programme, Lissabon, 1. Mai 2010.

Ganzeboom, H. B.G. und D.J. Treiman (2003), „Three Internationally Standardised Measures for Comparative Research on Occupational Status“, S. 159-193 in J.H.P. Hoffmeyer-Zlotnik und C. Wolf (Hrsg.), *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, Kluwer Academic Press, New York.

Kjærnsli, M. und S. Lie (2011), „Students' Preference for Science Careers: International Comparisons Based on PISA 2006“, *International Journal of Science Education*, Vol. 33/1, S. 121-144, <http://dx.doi.org/10.1080/09500693.2010.518642>.

OECD (erscheint demnächst), *PISA 2015 Technical Report*, PISA, OECD Publishing, Paris.

OECD (2016), *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*, PISA, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264255425-en>.

OECD (2007), *PISA 2006: Naturwissenschaftliche Kompetenzen für die Welt von morgen*, W. Bertelsmann Verlag, Bielefeld, <http://dx.doi.org/10.1787/9789264041257-de>.

OECD (1999), *Classifying Educational Programmes: Manual for ISCED-97 Implementation in OECD Countries*, OECD Publishing, Paris.

Sikora, J. und A. Pokropek (2012), „Gender Segregation of Adolescent Science Career Plans in 50 Countries“, *Science Education*, Vol. 96/2, S. 234-264, <http://dx.doi.org/10.1002/sce.20479>.

Warm, T.A. (1985), „Weighted Maximum Likelihood Estimation of Ability in Item Response Theory with Tests of Finite Length“, Technical Report CGI-TR-85-08, U.S. Coast Guard Institute, Oklahoma City.

O*NET OnLine (o.J.), „All STEM disciplines“, www.onetonline.org/find/stem?t=0, (Zugriff am 4. Oktober 2016).



ANHANG A2

PISA-ZIELPOPULATION, PISA-STICHPROBEN UND DEFINITION DER SCHULEN

Definition der PISA-Zielpopulation

PISA 2015 untersucht die kumulativen Bildungserträge und Lernergebnisse zu einem Zeitpunkt, an dem sich die meisten Jugendlichen noch in der Phase der Erstausbildung befinden.

Eine große Herausforderung bei einer internationalen Erhebung besteht darin sicherzustellen, dass die internationale Vergleichbarkeit der Zielpopulationen in den verschiedenen Ländern gewährleistet ist.

Aufgrund länderspezifischer Unterschiede im Hinblick auf die Art und die Verbreitung von Einrichtungen des Elementarbereichs, das reguläre Einschulungsalter sowie die institutionelle Struktur des Bildungssystems sind Definitionen der Zielpopulation, die sich auf bestimmte Klassenstufen beziehen, für internationale Vergleiche ungeeignet. Daher werden bei internationalen Vergleichen von Schulleistungen die Populationen in der Regel in Bezug auf ein Zielalter definiert. In einigen früheren internationalen Erhebungen wurden die Zielpopulationen auch auf der Basis der Klassenstufe definiert, die jeweils die breiteste Erfassung einer bestimmten Alterskohorte gewährleistet. Dieser Ansatz hat den Nachteil, dass leichte Abweichungen bei der altersmäßigen Verteilung der Schülerinnen und Schüler auf die einzelnen Klassenstufen häufig dazu führen, dass in verschiedenen Ländern oder in verschiedenen Bildungssystemen innerhalb einzelner Länder unterschiedliche Zielklassenstufen ausgewählt werden, was die Vergleichbarkeit der Ergebnisse zwischen den einzelnen Ländern und zuweilen auch innerhalb der Länder ernsthaft in Frage stellt. Da gewöhnlich nicht alle Schülerinnen und Schüler der gewünschten Altersgruppe in klassenstufenbasierten Stichproben repräsentiert sind, kann es darüber hinaus zu potenziell gravierenderen Verzerrungen der Ergebnisse kommen, wenn die unterrepräsentierten Schülerinnen und Schüler in einigen Ländern üblicherweise die nächsthöhere Klassenstufe, in anderen Ländern wiederum die nächstuntere Klassenstufe besuchen. Dadurch blieben in der einen Ländergruppe Schülerinnen und Schüler mit einem potenziell höheren Leistungsniveau unberücksichtigt, während in der anderen Ländergruppe Schüler mit einem potenziell niedrigeren Leistungsniveau nicht erfasst würden.

Um dieses Problem zu vermeiden, wird in PISA eine altersbezogene Definition der Zielpopulation verwendet, d.h. eine Definition, die von der institutionellen Struktur der Bildungssysteme der jeweiligen Länder unabhängig ist. PISA erfasst Schülerinnen und Schüler, die zu Beginn der Testperiode zwischen 15 Jahren und 3 (vollen) Monaten und 16 Jahren und 2 (vollen) Monaten alt waren – wobei eine Abweichung von plus/minus einem Monat akzeptiert wird – und die eine Bildungseinrichtung mit Klassenstufe 7 oder darüber besuchen, unabhängig davon, welche Klassenstufe oder Art von Bildungseinrichtung sie besuchen und ob es sich um eine Ganztags- oder Halbtagschule handelt. Bildungseinrichtungen werden in dieser Veröffentlichung generell als „Schulen“ bezeichnet, obwohl einige (insbesondere manche Formen berufsbildender Einrichtungen) im landesüblichen Sprachgebrauch u.U. nicht Schulen genannt werden. Wie aufgrund dieser Definition zu erwarten, betrug das Durchschnittsalter der teilnehmenden Schülerinnen und Schüler im OECD-Raum 15 Jahre und 9 Monate. Die Varianz des Durchschnittsalters in den verschiedenen Ländern belief sich auf 2 Monate und 18 Tage (0,20 Jahre), wobei das niedrigste Durchschnittsalter 15 Jahre und 8 Monate und das höchste 15 Jahre und 10 Monate war.

Aufgrund dieser Populationsdefinition trifft PISA Aussagen über die Kenntnisse und Fähigkeiten einer Gruppe von Schülerinnen und Schülern, die in einer vergleichbaren Referenzperiode geboren sind, aber über unterschiedliche inner- wie außerschulische Bildungserfahrungen verfügen können. In PISA gelten diese Kenntnisse und Fähigkeiten als die Erträge der Bildung in einem für alle Länder gleich angesetzten Alter. Je nach der von den einzelnen Ländern verfolgten Politik in Bezug auf Einschulung, Auswahl und Versetzung können sich die betreffenden Schülerinnen und Schüler auf ein engeres oder breiteres Spektrum von Klassenstufen in verschiedenen Bildungssystemen, Bildungsgängen oder Bildungszweigen verteilen. Diese Abweichungen sollten beim Vergleich der PISA-Ergebnisse zwischen den verschiedenen Ländern berücksichtigt werden, da die beobachteten Unterschiede zwischen Schülern im Alter von 15 Jahren aufgrund bzw. im Falle mit der Zeit konvergierender Bildungserfahrungen später möglicherweise nicht mehr zu erkennen sind.

Wenn ein Land in Naturwissenschaften, Lesekompetenz oder Mathematik eine wesentlich höhere Punktzahl erreicht als ein anderes, ist daraus nicht automatisch zu folgern, dass die Schulen oder bestimmte Teile des Bildungssystems in diesem Land effektiver sind als in dem anderen. Man kann daraus aber durchaus schließen, dass der kumulative Effekt der Lernerfahrungen in dem Land mit der höheren Punktzahl von der frühen Kindheit bis zum Alter von 15 Jahren, sowohl im schulischen als auch im außerschulischen Umfeld, zu positiveren Ergebnissen in den von PISA getesteten Grundbildungsbereichen geführt hat.

In der PISA-Zielpopulation nicht enthalten sind Schülerinnen und Schüler des jeweiligen Landes, die eine Schule im Ausland besuchen. Erfasst sind hingegen ausländische Schülerinnen und Schüler, die eine Schule im Erhebungsland besuchen.

Um den Ländern entgegenzukommen, die zum Zweck nationaler Analysen nach Klassenstufen aufgeschlüsselte Ergebnisse wünschten, wurde in PISA 2015 eine Stichprobenoption zur Ergänzung der altersbezogenen Stichprobe durch eine klassenstufenbezogene Stichprobe angeboten.

Erfassung der PISA-Schülerpopulation

Alle Länder und Volkswirtschaften waren um eine möglichst breite Erfassung der 15-jährigen Schülerinnen und Schüler in ihren nationalen Stichproben bemüht, auch solcher in Sondereinrichtungen. Daher erreichte PISA 2015 einen für internationale Erhebungen dieser Art beispiellosen Grad der Populationserfassung.

Die PISA-Stichprobenstandards erlaubten es den Ländern, bis zu insgesamt 5% der relevanten Population auszuschließen, sei es durch Ausschluss von Schulen oder durch Ausschluss von Schülern innerhalb der Schulen. Mit Ausnahme von 12 Ländern – dem Vereinigten Königreich (8,22%), Luxemburg (8,16%), Kanada (7,49%), Norwegen (6,75%), Neuseeland (6,54%), Schweden (5,71%), Estland (5,52%), Australien (5,31%), Montenegro (5,17%), Litauen (5,12%), Lettland (5,07%) und Dänemark (5,04%) – konnten alle Länder diese Regel einhalten, und in 29 Ländern und Volkswirtschaften lag die Gesamtausschlussrate bei weniger als 2%. Wenn sprachlich bedingte Ausschlüsse berücksichtigt, d.h. aus der Gesamtausschlussrate herausgerechnet werden, liegen die Ausschlussraten in Dänemark, Lettland, Neuseeland und Schweden nicht mehr über 5%. Wegen Einzelheiten vgl. www.pisa.oecd.org.

Zu den Ausschlüssen innerhalb der oben genannten Grenzen zählen:

- Auf Schulebene: Schulen, die in schwer erreichbaren Gegenden liegen oder in denen die Durchführung der PISA-Erhebung als nicht praktikabel angesehen wurde; und Schulen, deren Unterrichtsangebot sich auf die Kategorien beschränkt, die unter der Rubrik „Ausschlüsse innerhalb der Schulen“ definiert sind, wie z.B. Blindenschulen. Der Prozentsatz der 15-Jährigen, die solche Schulen besuchten, musste weniger als 2,5% der angestrebten Grundgesamtheit auf Länderebene ausmachen (maximal 0,5% für die erstgenannte Gruppe und maximal 2% für die zweitgenannte Gruppe). Der *PISA 2015 Technical Report* (OECD, erscheint demnächst) enthält Angaben über die Größenordnung, Art und Begründung der Ausschlüsse auf Schulebene.
- Auf Schülerebene: Schülerinnen und Schüler mit kognitiver Behinderung; Schülerinnen und Schüler mit funktionaler Behinderung; Schülerinnen und Schüler mit unzureichender Kenntnis der Testsprache; sonstige Schülerinnen und Schüler (eine Kategorie, die von den nationalen Zentren definiert und vom Internationalen Konsortium genehmigt werden muss); und Schülerinnen und Schüler, die im Haupttestbereich in einer Unterrichtssprache unterrichtet werden, für die keine Unterlagen verfügbar waren. Es war nicht möglich, Schülerinnen und Schüler allein aufgrund eines niedrigen Leistungsniveaus oder normaler Disziplinprobleme auszuschließen. Der Prozentsatz der innerhalb der Schulen ausgeschlossenen 15-Jährigen musste niedriger sein als 2,5% der auf Länderebene angestrebten Grundgesamtheit.

Tabelle A2.1 beschreibt die Zielpopulation der an PISA 2015 teilnehmenden Länder. Nähere Einzelheiten über die Zielpopulation und die Anwendung der PISA-Stichprobenstandards enthält der *PISA 2015 Technical Report* (OECD, erscheint demnächst).

- **Spalte 1** zeigt die Gesamtzahl der 15-Jährigen gemäß den aktuellsten vorliegenden Informationen, d.h. für die meisten Länder aus dem Vorjahr der Erhebung, also 2014.
- **Spalte 2** zeigt die Zahl der 15-Jährigen in Klassenstufe 7 oder darüber (gemäß obiger Definition), also die „in Betracht kommende Population“.
- **Spalte 3** zeigt die angestrebte Grundgesamtheit auf Länderebene. Die Länder konnten von vornherein – hauptsächlich aus praktischen Gründen – bis zu 0,5% der Schülerinnen und Schüler aus der in Betracht kommenden Population ausschließen. In folgenden Fällen wurde diese Höchstgrenze überschritten, was aber im Einvernehmen mit dem PISA-Konsortium geschah: Belgien hat 0,21% seiner Schülerpopulation ausgeschlossen, weil es sich um Schüler handelte, die gleichzeitig einer Erwerbstätigkeit nachgingen; Kanada hat Schüler in Indigenengebieten (Territories and Aboriginal Reserves) ausgeschlossen, insgesamt 1,22% seiner Zielpopulation; Chile hat Schüler ausgeschlossen, die auf den Osterinseln, auf den Juan-Fernández-Inseln und in der Antarktis leben und 0,04% der Zielpopulation ausmachen; und die Vereinigten Arabischen Emirate haben 0,04% ihrer Schülerinnen und Schüler ausgeschlossen, über die keine Informationen vorlagen. Die Region Massachusetts in den Vereinigten Staaten, deren Stichproben international überprüft wurden, hat 13,11% ihrer Schülerpopulation ausgeschlossen, und in North Carolina waren es 5,64% der Zielpopulation. Für diese beiden Regionen erfasst die angestrebte Grundgesamtheit nur 15-jährige Schülerinnen und Schüler, die Klassenstufe 7 oder darüber in öffentlichen Schulen besuchen. Die von der angestrebten Grundgesamtheit ausgeschlossenen Schülerinnen und Schüler gehen auf Privatschulen.
- **Spalte 4** zeigt die Zahl der Schülerinnen und Schüler in Schulen, die von der angestrebten Grundgesamtheit auf Länderebene bei der Stichprobenziehung oder zu einem späteren Zeitpunkt während der Datensammlung ausgeschlossen wurden.
- **Spalte 5** zeigt die Größe der angestrebten Grundgesamtheit auf Länderebene nach Abzug der Schülerinnen und Schüler in den ausgeschlossenen Schulen. Diese wird ermittelt, indem Spalte 4 von Spalte 3 subtrahiert wird.
- **Spalte 6** zeigt den Prozentsatz der Schülerinnen und Schüler in den ausgeschlossenen Schulen. Dieser wird ermittelt, indem Spalte 4 durch Spalte 3 dividiert und mit 100 multipliziert wird.



[Teil 1/1]

Tabelle A2.1 PISA-Zielpopulationen und -Stichproben

	Informationen über die Grundgesamtheit und die Stichproben											Erfassungszindizes			
	Gesamtpopulation der 15-Jährigen	Gesamtzahl der 15-Jährigen in Klassenstufe 7 oder darüber	Angestrebte Grundgesamtheit auf Länderebene	Ausschlüsse auf Schulebene, insg.	Angestrebte Grundgesamtheit auf Länderebene nach Ausschluss auf Schul- und vor Ausschluss auf Schülerebene	Ausschlussrate auf Schulebene (in %)	Zahl der teilnehmenden Schüler	Gewichtete Zahl der teilnehmenden Schüler	Zahl der ausgeschloss. Schüler	Gewichtete Zahl der ausgeschloss. Schüler	Ausschlussrate innerhalb der Schulen (in %)	Ausschlussrate insg. (in %)	Erfassungszindex 1: Erfassung der angestrebten Grundgesamtheit auf Länderebene	Erfassungszindex 2: Erfassung der Schülerpopulation auf Länderebene	Erfassungszindex 3: Erfassung der Population
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
OECD-Länder															
Australien	282 888	282 547	282 547	6 940	275 607	2,46	14 530	256 329	681	7 736	2,93	5,31	0,947	0,947	0,906
Österreich	88 013	82 683	82 683	790	81 893	0,96	7 007	73 379	84	866	1,17	2,11	0,979	0,979	0,834
Belgien	123 630	121 954	121 954	1 597	120 097	1,31	9 651	114 902	39	410	0,36	1,66	0,983	0,981	0,929
Kanada	396 966	381 660	376 994	1 590	375 404	0,42	20 058	331 546	1 830	25 340	7,10	7,49	0,925	0,914	0,835
Chile	255 440	245 947	245 852	2 641	243 211	1,07	7 053	203 782	37	1 393	0,68	1,75	0,983	0,982	0,798
Tschech. Rep.	90 391	90 076	90 076	1 814	88 262	2,01	6 894	84 519	25	368	0,43	2,44	0,976	0,976	0,935
Dänemark	68 174	67 466	67 466	605	66 861	0,90	7 161	60 655	514	2 644	4,18	5,04	0,950	0,950	0,890
Estland	11 676	11 491	11 491	416	11 075	3,62	5 587	10 834	116	218	1,97	5,52	0,945	0,945	0,928
Finnland	58 526	58 955	58 955	472	58 483	0,80	5 882	56 934	124	1 157	1,99	2,78	0,972	0,972	0,973
Frankreich	807 867	778 679	778 679	28 742	749 937	3,69	6 108	734 944	35	3 620	0,49	4,16	0,958	0,958	0,910
Deutschland	774 149	774 149	774 149	11 150	762 999	1,44	6 522	743 969	54	5 342	0,71	2,14	0,979	0,979	0,961
Griechenland	105 530	105 253	105 253	953	104 300	0,91	5 532	96 157	58	965	0,99	1,89	0,981	0,981	0,911
Ungarn	94 515	90 065	90 065	1 945	88 120	2,16	5 658	84 644	55	1 009	1,18	3,31	0,967	0,967	0,896
Island	4 250	4 195	4 195	17	4 178	0,41	3 374	3 966	131	132	3,23	3,62	0,964	0,964	0,933
Irland	61 234	59 811	59 811	72	59 739	0,12	5 741	59 082	197	1 825	3,00	3,11	0,969	0,969	0,965
Israel	124 852	118 997	118 997	2 310	116 687	1,94	6 598	117 031	115	1 803	1,52	3,43	0,966	0,966	0,937
Italien	616 761	567 268	567 268	11 190	556 078	1,97	11 583	495 093	246	9 395	1,86	3,80	0,962	0,962	0,803
Japan	1 201 615	1 175 907	1 175 907	27 323	1 148 584	2,32	6 647	1 138 349	2	318	0,03	2,35	0,976	0,976	0,947
Korea	620 687	619 950	619 950	3 555	616 395	0,57	5 581	569 106	20	1 806	0,32	0,89	0,991	0,991	0,917
Lettland	17 255	16 955	16 955	677	16 278	3,99	4 869	15 320	70	174	1,12	5,07	0,949	0,949	0,888
Luxemburg	6 327	6 053	6 053	162	5 891	2,68	5 299	5 540	331	331	5,64	8,16	0,918	0,918	0,876
Mexiko	2 257 399	1 401 247	1 401 247	5 905	1 395 342	0,42	7 568	1 392 995	30	6 810	0,49	0,91	0,991	0,991	0,617
Niederlande	201 670	200 976	200 976	6 866	194 110	3,42	5 385	191 817	14	502	0,26	3,67	0,963	0,963	0,951
Neuseeland	60 162	57 448	57 448	681	56 767	1,19	4 520	54 274	333	3 112	5,42	6,54	0,935	0,935	0,902
Norwegen	63 642	63 491	63 491	854	62 637	1,35	5 456	58 083	345	3 366	5,48	6,75	0,933	0,933	0,913
Polen	380 366	361 600	361 600	6 122	355 478	1,69	4 478	345 709	34	2 418	0,69	2,38	0,976	0,976	0,909
Portugal	110 939	101 107	101 107	424	100 683	0,42	7 325	97 214	105	860	0,88	1,29	0,987	0,987	0,876
Slowak. Rep.	55 674	55 203	55 203	1 376	53 827	2,49	6 350	49 654	114	912	1,80	4,25	0,957	0,957	0,892
Slowenien	18 078	17 689	17 689	290	17 399	1,64	6 406	16 773	114	247	1,45	3,07	0,969	0,969	0,928
Spanien	440 084	414 276	414 276	2 175	412 101	0,53	6 736	399 935	200	10 893	2,65	3,16	0,968	0,968	0,909
Schweden	97 749	97 210	97 210	1 214	95 996	1,25	5 458	91 491	275	4 324	4,51	5,71	0,943	0,943	0,936
Schweiz	85 495	83 655	83 655	2 320	81 335	2,77	5 860	82 223	107	1 357	1,62	4,35	0,956	0,956	0,962
Türkei	1 324 089	1 100 074	1 100 074	5 746	1 094 328	0,52	5 895	925 366	31	5 359	0,58	1,10	0,989	0,989	0,699
Ver. Königreich	747 593	746 328	746 328	23 412	722 916	3,14	14 157	627 703	870	34 747	5,25	8,22	0,918	0,918	0,840
Ver. Staaten	4 220 325	3 992 053	3 992 053	12 001	3 980 052	0,30	5 712	3 524 497	193	109 580	3,02	3,31	0,967	0,967	0,835
Partnerländer/-volkswirtschaften															
Albanien	48 610	45 163	45 163	10	45 153	0,02	5 215	40 896	0	0	0,00	0,02	1,000	1,000	0,841
Algerien	389 315	354 936	354 936	0	354 936	0,00	5 519	306 647	0	0	0,00	0,00	1,000	1,000	0,788
Argentinien	718 635	578 308	578 308	2 617	575 691	0,45	6 349	394 917	21	1 367	0,34	0,80	0,992	0,992	0,550
Brasilien	3 803 681	2 853 388	2 853 388	64 392	2 788 996	2,26	23 141	2 425 961	119	13 543	0,56	2,80	0,972	0,972	0,638
P-S-J-G (China)	2 084 958	1 507 518	1 507 518	58 639	1 448 879	3,89	9 841	1 331 794	33	3 609	0,27	4,15	0,959	0,959	0,639
Bulgarien	66 601	59 397	59 397	1 124	58 273	1,89	5 928	53 685	49	433	0,80	2,68	0,973	0,973	0,806
Kolumbien	760 919	674 079	674 079	37	674 042	0,01	11 795	567 848	9	507	0,09	0,09	0,999	0,999	0,746
Costa Rica	81 773	66 524	66 524	0	66 524	0,00	6 866	51 897	13	98	0,19	0,19	0,998	0,998	0,635
Kroatien	45 031	35 920	35 920	805	35 115	2,24	5 809	40 899	86	589	1,42	3,63	0,964	0,964	0,908
Zypern*	9 255	9 255	9 253	109	9 144	1,18	5 571	8 785	228	292	3,22	4,36	0,956	0,956	0,949
Dominik. Rep.	193 153	139 555	139 555	2 382	137 173	1,71	4 740	132 300	4	106	0,08	1,79	0,982	0,982	0,685
eJR Mazedonien	16 719	16 717	16 717	275	16 458	1,55	5 324	15 847	8	19	0,12	1,67	0,983	0,983	0,948
Georgien	48 695	43 197	43 197	1 659	41 522	3,88	5 316	38 334	35	230	0,60	4,45	0,955	0,955	0,787
Hongkong (China)	65 100	61 630	61 630	708	60 922	1,15	5 359	57 662	36	374	0,65	1,79	0,982	0,982	0,886
Indonesien	4 534 216	3 182 816	3 182 816	4 046	3 178 770	0,13	6 513	3 092 773	0	0	0,00	0,13	0,999	0,999	0,682
Jordanien	126 399	121 729	121 729	71	121 658	0,06	7 267	108 669	70	1 006	0,92	0,97	0,990	0,990	0,860
Kasachstan	211 407	209 555	209 555	7 475	202 080	3,57	7 841	192 909	0	0	0,00	3,57	0,964	0,964	0,912
Kosovo	31 546	28 229	28 229	1 156	27 073	4,10	4 826	22 333	50	174	0,77	4,84	0,952	0,952	0,708
Libanon	64 044	62 281	62 281	1 300	60 981	2,09	4 546	42 331	0	0	0,00	2,09	0,979	0,979	0,661
Litauen	33 163	32 097	32 097	573	31 524	1,79	6 525	29 915	227	1 050	3,39	5,12	0,949	0,949	0,902
Macau (China)	5 100	4 417	4 417	3	4 414	0,07	4 476	4 507	0	0	0,00	0,07	0,999	0,999	0,884
Malaysia	540 000	448 838	448 838	2 418	446 420	0,54	8 861	412 524	41	2 344	0,56	1,10	0,989	0,989	0,764
Malta	4 397	4 406	4 406	63	4 343	1,43	3 634	4 296	41	41	0,95	2,36	0,976	0,976	0,977
Moldau	31 576	30 601	30 601	182	30 419	0,59	5 325	29 341	21	118	0,40	0,99	0,990	0,990	0,929
Montenegro	7 524	7 506	7 506	40	7 466	0,53	5 665	6 777	300	332	4,66	5,17	0,948	0,948	0,901
Peru	580 371	478 229	478 229	6 355	471 874	1,33	6 971	431 738	13	745	0,17	1,50	0,985	0,985	0,744
Katar	13 871	13 850	13 850	380	13 470	2,74	12 083	12 951	193	193	1,47	4,17	0,958	0,958	0,934
Rumänien	176 334	176 334	176 334	1 823	174 511	1,03	4 876	164 216	3	120	0,07	1,11	0,989	0,989	0,931
Russ. Föderation	1 176 473	1 172 943	1 172 943	24 217	1 148 726	2,									

[Teil 1/2]

Tabelle A2.2 Ausschlüsse

	Zahl der ausgeschlossenen Schüler (ungewichtet)					
	Schüler mit funktionaler Behinderung	Schüler mit kognitiver Behinderung	Schüler mit Sprachproblemen	Aus anderen Gründen ausgeschlossene Schüler	Mangels verfügbarer Unterlagen in der Unterrichtssprache ausgeschloss. Schüler	Ausgeschlossene Schüler insgesamt
	(Code 1)	(Code 2)	(Code 3)	(Code 4)	(Code 5)	(%)
	(1)	(2)	(3)	(4)	(5)	(6)
OECD-Länder						
Australien	85	528	68	0	0	681
Österreich	8	15	61	0	0	84
Belgien	4	18	17	0	0	39
Kanada	156	1 308	366	0	0	1 830
Chile	6	30	1	0	0	37
Tschech. Rep.	2	9	14	0	0	25
Dänemark	18	269	156	70	1	514
Estland	17	93	6	0	0	116
Finnland	2	90	17	8	7	124
Frankreich	5	21	9	0	0	35
Deutschland	4	25	25	0	0	54
Griechenland	3	44	11	0	0	58
Ungarn	3	13	9	30	0	55
Island	9	66	47	9	0	131
Irland	25	57	55	60	0	197
Israel	22	68	25	0	0	115
Italien	78	147	21	0	0	246
Japan	0	2	0	0	0	2
Korea	3	17	0	0	0	20
Lettland	7	47	16	0	0	70
Luxemburg	4	254	73	0	0	331
Mexiko	4	23	3	0	0	30
Niederlande	1	13	0	0	0	14
Neuseeland	23	140	167	0	3	333
Norwegen	11	253	81	0	0	345
Polen	11	20	0	3	0	34
Portugal	4	99	2	0	0	105
Slowak. Rep.	7	71	2	34	0	114
Slowenien	33	36	45	0	0	114
Spanien	9	144	47	0	0	200
Schweden	154	0	121	0	0	275
Schweiz	8	42	57	0	0	107
Türkei	1	23	7	0	0	31
Ver. Königreich	77	690	102	0	1	870
Ver. Staaten	16	120	44	13	0	193
Partnerländer/-volkswirtschaften						
Albanien	0	0	0	0	0	0
Algerien	0	0	0	0	0	0
Argentinien	10	10	1	0	0	21
Brasilien	20	99	0	0	0	119
P-S-J-G (China)	6	25	2	0	0	33
Bulgarien	39	6	4	0	0	49
Kolumbien	3	4	2	0	0	9
Costa Rica	3	1	0	9	0	13
Kroatien	2	75	9	0	0	86
Zypern*	12	164	52	0	0	228
Dominik. Rep.	1	3	0	0	0	4
ejR Mazedonien	7	1	0	0	0	8
Georgien	3	25	7	0	0	35
Hongkong (China)	0	35	1	0	0	36
Indonesien	0	0	0	0	0	0
Jordanien	43	17	10	0	0	70
Kasachstan	0	0	0	0	0	0
Kosovo	9	13	27	0	0	50
Libanon	0	0	0	0	0	0
Litauen	12	213	2	0	0	227
Macau (China)	0	0	0	0	0	0
Malaysia	10	22	9	0	0	41
Malta	8	27	6	0	0	41
Moldau	12	8	1	0	0	21
Montenegro	14	23	5	0	258	300
Peru	4	9	0	0	0	13
Katar	76	110	7	0	0	193
Rumänien	1	1	1	0	0	3
Russ. Föderation	3	10	0	0	0	13
Singapur	3	15	7	0	0	25
Chinesisch Taipeh	3	19	0	0	0	22
Thailand	1	19	2	0	0	22
Trinidad und Tobago	0	0	0	0	0	0
Tunesien	0	0	3	0	0	3
Ver. Arab. Emirate	16	24	23	0	0	63
Uruguay	2	4	0	0	0	6
Vietnam	0	0	0	0	0	0

Ausschlusscodes:

Code 1: Funktionale Behinderung – die Schülerin/der Schüler hat eine mittelschwere bis schwere dauerhafte körperliche Behinderung.

Code 2: Kognitive Behinderung – die Schülerin/der Schüler hat eine mentale oder emotionale Behinderung und wurde nach entsprechenden Tests bzw. nach der professionellen Meinung qualifizierter Kräfte als kognitiv retardiert eingestuft.

Code 3: Unzureichende Kenntnis der Testsprache – keine der Testsprachen des jeweiligen Landes ist die Muttersprache der Schülerin/des Schülers und die Schülerin/der Schüler lebt seit weniger als einem Jahr im betreffenden Land.

Code 4: Sonstige – von den nationalen Zentren definierte und dem internationalen Konsortium genehmigte Ausschlussgründe.

Code 5: Keine Unterlagen in der Unterrichtssprache verfügbar.

Anmerkung: Wegen einer ausführlicheren Erklärung der in dieser Tabelle enthaltenen Einzelheiten vgl. *PISA 2015 Technical Report* (OECD, erscheint demnächst).

* Vgl. Anmerkungen am Anfang dieses Anhangs.

StatLink  <http://dx.doi.org/10.1787/888933433129>



[Teil 2/2]

Tabelle A2.2 Ausschlüsse

	Zahl der ausgeschlossenen Schüler (gewichtet)					
	Schüler mit funktionaler Behinderung (Code 1)	Schüler mit kognitiver Behinderung (Code 2)	Schüler mit Sprachproblemen (Code 3)	Aus anderen Gründen ausgeschlossene Schüler (Code 4)	Mangels verfügbarer Unterlagen in der Unterrichtssprache ausgeschlossen. Schüler (Code 5)	Ausgeschlossene Schüler insgesamt (12)
	(7)	(8)	(9)	(10)	(11)	(12)
OECD-Länder						
Australien	932	6 011	793	0	0	7 736
Österreich	74	117	675	0	0	866
Belgien	33	192	185	0	0	410
Kanada	1 901	18 018	5 421	0	0	25 340
Chile	194	1 190	9	0	0	1 393
Tschech. Rep.	40	140	188	0	0	368
Dänemark	122	1 539	551	421	11	2 644
Estland	29	176	13	0	0	218
Finnland	18	858	156	67	58	1 157
Frankreich	562	2 144	914	0	0	3 620
Deutschland	423	2 562	2 357	0	0	5 342
Griechenland	43	729	193	0	0	965
Ungarn	57	284	114	554	0	1 009
Island	9	67	47	9	0	132
Irland	213	526	516	570	0	1 825
Israel	349	1 070	384	0	0	1 803
Italien	3 316	5 199	880	0	0	9 395
Japan	0	318	0	0	0	318
Korea	291	1 515	0	0	0	1 806
Lettland	21	115	38	0	0	174
Luxemburg	4	254	73	0	0	331
Mexiko	842	4 802	1 165	0	0	6 810
Niederlande	33	469	0	0	0	502
Neuseeland	233	1 287	1 568	0	24	3 112
Norwegen	105	2 471	790	0	0	3 366
Polen	876	1 339	0	203	0	2 418
Portugal	29	818	13	0	0	860
Slowak. Rep.	44	567	12	288	0	912
Slowenien	84	71	92	0	0	247
Spanien	511	7 662	2 720	0	0	10 893
Schweden	2 380	0	1 944	0	0	4 324
Schweiz	91	540	726	0	0	1 357
Türkei	43	4 094	1 222	0	0	5 359
Ver. Königreich	2 724	27 808	4 001	0	214	34 747
Ver. Staaten	7 873	67 816	26 525	7 366	0	109 580
Partnerländer/-volkswirtschaften						
Albanien	0	0	0	0	0	0
Algerien	0	0	0	0	0	0
Argentinien	579	770	18	0	0	1 367
Brasilien	1 743	11 800	0	0	0	13 543
P.-S.-J.-G. (China)	438	2 970	201	0	0	3 609
Bulgarien	347	51	35	0	0	433
Kolumbien	181	309	17	0	0	507
Costa Rica	22	5	0	71	0	98
Kroatien	13	501	75	0	0	589
Zypern*	16	212	65	0	0	292
Dominik. Rep.	24	82	0	0	0	106
ejR Mazedonien	15	4	0	0	0	19
Georgien	19	170	41	0	0	230
Hongkong (China)	0	363	11	0	0	374
Indonesien	0	0	0	0	0	0
Jordanien	656	227	122	0	0	1 006
Kasachstan	0	0	0	0	0	0
Kosovo	28	37	104	0	0	174
Libanon	0	0	0	0	0	0
Litauen	40	1 000	10	0	0	1 050
Macau (China)	0	0	0	0	0	0
Malaysia	663	1 100	580	0	0	2 344
Malta	8	27	6	0	0	41
Moldau	66	51	1	0	0	118
Montenegro	27	38	6	0	261	332
Peru	224	520	0	0	0	745
Katar	76	110	7	0	0	193
Rumänien	31	63	26	0	0	120
Russ. Föderation	425	2 044	0	0	0	2 469
Singapur	22	115	43	0	0	179
Chinesisch Taipeh	78	568	0	0	0	647
Thailand	114	1 830	163	0	0	2 107
Trinidad und Tobago	0	0	0	0	0	0
Tunesien	0	0	61	0	0	61
Ver. Arab. Emirate	30	75	47	0	0	152
Uruguay	10	22	0	0	0	32
Vietnam	0	0	0	0	0	0

Ausschlusscodes:

Code 1: Funktionale Behinderung – die Schülerin/der Schüler hat eine mittelschwere bis schwere dauerhafte körperliche Behinderung.

Code 2: Kognitive Behinderung – die Schülerin/der Schüler hat eine mentale oder emotionale Behinderung und wurde nach entsprechenden Tests bzw. nach der professionellen Meinung qualifizierter Kräfte als kognitiv retardiert eingestuft.

Code 3: Unzureichende Kenntnis der Testsprache – keine der Testsprachen des jeweiligen Landes ist die Muttersprache der Schülerin/des Schülers und die Schülerin/der Schüler lebt seit weniger als einem Jahr im betreffenden Land.

Code 4: Sonstige – von den nationalen Zentren definierte und dem internationalen Konsortium genehmigte Ausschlussgründe.

Code 5: Keine Unterlagen in der Unterrichtssprache verfügbar.

Anmerkung: Wegen einer ausführlicheren Erklärung der in dieser Tabelle enthaltenen Einzelheiten vgl. PISA 2015 Technical Report (OECD, erscheint demnächst).

* Vgl. Anmerkungen am Anfang dieses Anhangs.

StatLink <http://dx.doi.org/10.1787/888933433129>

- **Spalte 7** zeigt die Zahl der an PISA 2015 teilnehmenden Schülerinnen und Schüler. Dabei ist zu beachten, dass die im Rahmen nationaler Optionen zusätzlich einbezogenen Schülerinnen und Schüler in diesem Wert in manchen Fällen nicht berücksichtigt sind.
- **Spalte 8** zeigt die gewichtete Zahl der teilnehmenden Schülerinnen und Schüler, d.h. die Zahl der Schülerinnen und Schüler in der definierten Grundgesamtheit auf Länderebene, die durch die PISA-Stichprobe repräsentiert wird.
- Jedes Land war bestrebt, den Grad der Erfassung der PISA-Zielpopulation innerhalb der in die Stichprobe einbezogenen Schulen zu maximieren. Für jede einbezogene Schule wurden zunächst alle in Betracht kommenden Schülerinnen und Schüler, d.h. alle 15-jährigen, unabhängig von der Klassenstufe, aufgelistet. Auch die von der Stichprobe auszuschließenden Schülerinnen und Schüler mussten in der Stichprobendokumentation aufgeführt sein; dazu musste eine Liste erstellt werden, in der die Gründe für ihren Ausschluss anzugeben waren. Spalte 9 informiert über die Gesamtzahl der ausgeschlossenen Schülerinnen und Schüler, die in Tabelle A2.2 näher beschrieben und in spezifische Kategorien unterteilt sind.
- Der in **Spalte 10** angegebene Wert entspricht der gewichteten Zahl der ausgeschlossenen Schülerinnen und Schüler, d.h. der Gesamtzahl der Schüler in der definierten Grundgesamtheit auf Länderebene, die durch die Zahl der aus der Stichprobe ausgeschlossenen Schülerinnen und Schüler repräsentiert wird, die ebenfalls in Tabelle A2.2 näher beschrieben und in verschiedene Kategorien unterteilt sind. Der Ausschluss der Schülerinnen und Schüler erfolgte auf der Basis von fünf Kategorien: Schülerinnen und Schüler mit kognitiver Behinderung (die eine mentale oder emotionale Behinderung aufweisen und kognitiv retardiert sind, so dass sie nicht in der Lage sind, an einer Erhebung unter den PISA-Testbedingungen teilzunehmen); Schülerinnen und Schüler mit funktionaler Behinderung (die an einer mittelschweren bis schweren dauerhaften körperlichen Behinderung leiden, so dass sie nicht in der Lage sind, an einer Erhebung unter den PISA-Testbedingungen teilzunehmen); Schülerinnen und Schüler mit unzureichender Kenntnis der Testsprache (die keine der Testsprachen des Landes hinreichend lesen oder sprechen können und daher die Sprachbarriere in der Testsituation nicht hätten überwinden können – in der Regel können Schülerinnen und Schüler, die weniger als ein Jahr Unterricht in der Testsprache hatten, von der Erhebung ausgeschlossen werden); sonstige Schülerinnen und Schüler (eine Kategorie, die von den nationalen Zentren definiert und vom Internationalen Konsortium genehmigt werden muss); sowie Schülerinnen und Schüler, die in dem Haupttestbereich in einer Unterrichtssprache unterrichtet werden, für die keine Unterlagen verfügbar waren.
- **Spalte 11** zeigt den Prozentsatz der Schülerinnen und Schüler, die innerhalb der Schulen ausgeschlossen wurden. Er ist berechnet als die gewichtete Zahl der ausgeschlossenen Schülerinnen und Schüler (Spalte 10), dividiert durch die gewichtete Zahl der ausgeschlossenen und der teilnehmenden Schülerinnen und Schüler (Spalte 8 plus Spalte 10), multipliziert mit 100.
- **Spalte 12** zeigt die Ausschlussrate insgesamt, die dem gewichteten Prozentsatz der angestrebten Grundgesamtheit der Zielpopulation auf Länderebene entspricht, die auf Schulebene oder auf innerschulischer Ebene von PISA ausgeschlossen wurde. Sie wird wie folgt berechnet: Ausschlussrate auf Schulebene (Spalte 6 dividiert durch 100), zuzüglich Ausschlussrate innerhalb der Schulen (Spalte 11 dividiert durch 100), multipliziert mit 1, abzüglich der Ausschlussrate auf Schulebene (Spalte 6 dividiert durch 100). Dieses Ergebnis wird dann mit 100 multipliziert.
- **Spalte 13** zeigt einen Index für den Erfassungsgrad der angestrebten Grundgesamtheit auf Länderebene durch die PISA-Stichprobe. Australien, Kanada, Dänemark, Estland, Lettland, Litauen, Luxemburg, Montenegro, Neuseeland, Norwegen, Schweden und das Vereinigte Königreich waren die einzigen Länder, in denen der Erfassungsgrad unter 95% lag.
- **Spalte 14** zeigt einen Index für den Erfassungsgrad der 15-jährigen Schülerinnen und Schüler durch die PISA-Stichprobe. Dieser Index misst den Gesamtanteil der nationalen Schülerpopulation, der durch die nach den Ausschlüssen verbleibende Schülerstichprobe erfasst wird. Dabei werden Ausschlüsse sowohl auf Schulebene als auch auf Schülerebene berücksichtigt. Nahe bei 100 liegende Werte deuten darauf hin, dass die PISA-Stichprobe das gesamte Schulsystem nach Definition für PISA 2015 repräsentiert. Der Index ergibt sich aus der gewichteten Zahl der teilnehmenden Schülerinnen und Schüler (Spalte 8), dividiert durch die gewichtete Zahl der teilnehmenden und der ausgeschlossenen Schüler (Spalte 8 plus Spalte 10), multipliziert mit der definierten Grundgesamtheit auf Länderebene (Spalte 5), dividiert durch die in Betracht kommende Population (Spalte 2) (mal 100).
- **Spalte 15** zeigt einen Index des prozentualen Anteils der erfassten Schülerpopulation. Dieser Index ergibt sich aus der gewichteten Gesamtzahl der teilnehmenden Schülerinnen und Schüler (Spalte 8), dividiert durch die Gesamtpopulation der 15-jährigen Schülerinnen und Schüler (Spalte 1).

Dieser hohe Erfassungsgrad trägt zur Vergleichbarkeit der Erhebungsergebnisse bei. Selbst wenn man z.B. annimmt, dass die ausgeschlossenen Schülerinnen und Schüler durchgehend schlechtere Ergebnisse erzielt hätten als die teilnehmenden und dass diese Korrelation mittelstark ausgeprägt ist, würde eine Ausschlussrate in einer Größenordnung von 5% wahrscheinlich nur zu einer Überzeichnung der nationalen Durchschnittsergebnisse um weniger als 5 Punkte führen (auf einer Skala mit einem internationalen Mittelwert von 500 Punkten und einer Standardabweichung von 100 Punkten). Diese Feststellung basiert auf folgenden Berechnungen: Beträgt die Korrelation zwischen der Ausschlussneigung und den Schülerleistungen 0,3, sind die resultierenden Durchschnittsergebnisse bei einer Ausschlussrate von 1% um 1 Punkt, von 5% um 3 Punkte und von 10% um 6 Punkte überzeichnet. Beträgt die Korrelation zwischen der Ausschlussneigung und den Schülerleistungen 0,5, sind die resultierenden Durchschnittsergebnisse bei einer Ausschlussrate von 1% um 1 Punkt, von 5% um 5 Punkte und von 10% um 10 Punkte überzeichnet. Für diese Berechnungen wurde ein Modell verwendet, in dem von einer bivariaten Normalverteilung der Leistungen und der Teilnahmeneigung ausgegangen wurde. Wegen näherer Einzelheiten vgl. *PISA 2015 Technical Report* (OECD, erscheint demnächst).



Stichprobenziehung und Beteiligungsquoten

Die Genauigkeit jeglicher Untersuchungsergebnisse hängt von der Qualität der Informationen ab, auf denen die nationalen Stichproben basieren, sowie von den Stichprobenziehungsverfahren. Für PISA wurden Qualitätsstandards, Verfahren, Instrumente und Verifikationsmechanismen entwickelt, die gewährleisten, dass die nationalen Stichproben vergleichbare Daten ergaben und die Ergebnisse ohne Vorbehalte miteinander verglichen werden können.

Die meisten PISA-Stichproben wurden nach dem Konzept der zweistufigen, geschichteten Stichproben konzipiert (soweit die Länder unterschiedliche Stichprobenkonzepte anwandten, sind diese dokumentiert im *PISA 2015 Technical Report*, OECD, erscheint demnächst). Im ersten Schritt wurde eine Stichprobe einzelner Schulen gezogen, die von 15-jährigen Schülerinnen und Schülern besucht wurden. Die Schulstichprobe wurde systematisch gezogen, mit Wahrscheinlichkeiten proportional zur Größe, die wiederum in Abhängigkeit zur geschätzten Zahl der in Betracht kommenden (15-jährigen) Schülerinnen und Schüler stand. In jedem Land wurden mindestens 150 Schulen (sofern vorhanden) ausgewählt, wobei die Anforderungen für die nationalen Analysen allerdings häufig eine etwas größere Stichprobe bedingten. Gleichzeitig mit der Stichprobenziehung wurden Ersatzschulen ermittelt für den Fall, dass eine der ausgewählten Schulen nicht an der PISA-2015-Erhebung teilnehmen wollte.

In Island, Luxemburg, Macau (China), Malta und Katar wurden alle Schulen und alle innerhalb der Schulen in Betracht kommenden Schülerinnen und Schüler in die Stichprobe aufgenommen.

Experten des PISA-Konsortiums führten in den meisten Teilnehmerländern die Auswahl der Stichproben durch; in den Ländern, die ihre Stichproben selbst auswählten, überwachten sie diesen Prozess. In der zweiten Phase des Auswahlprozesses wurden Schülerstichproben innerhalb der für die Stichprobe ermittelten Schulen gezogen. Nach der Auswahl der Schulen wurde eine Liste aller 15-jährigen Schülerinnen und Schüler an den für die Stichprobe gezogenen Schulen erstellt. Aus dieser Liste wurden 42 Schülerinnen und Schüler mit gleicher Wahrscheinlichkeit ausgewählt (wenn es weniger als 42 gab, wurden alle 15-jährigen Schülerinnen und Schüler ausgewählt). Die Zahl der für die Stichproben ausgewählten Schülerinnen und Schüler je Schule konnte von 42 abweichen, durfte jedoch 20 nicht unterschreiten.

Die Qualitätsstandards der PISA-Daten setzten eine Mindestteilnahmequote sowohl für die Schulen als auch für die Schülerinnen und Schüler voraus. Diese Standards wurden aufgestellt, um mögliche beteiligungsbedingte Verzerrungen zu minimieren. Bei den Ländern, die diese Standards erfüllten, dürften etwaige Verzerrungen infolge von Nichtbeteiligung unerheblich, d.h. in der Regel geringer als der Stichprobenfehler sein.

Für die ursprünglich ausgewählten Schulen war eine Mindestbeteiligungsquote von 85% erforderlich. Wenn die ursprüngliche Beteiligungsquote auf Schulebene zwischen 65% und 85% lag, konnte durch die Einbeziehung von Ersatzschulen immer noch eine akzeptable Quote erzielt werden. Dieses Verfahren war mit dem Risiko eines erhöhten Beteiligungsbias verbunden. Daher wurden die Teilnehmerländer ersucht, so viele der ursprünglich für die Stichprobe gezogenen Schulen wie möglich zur Teilnahme zu veranlassen. Wenn die Beteiligungsquote der Schülerinnen und Schüler zwischen 25% und 50% lag, galt die betreffende Schule nicht als teilnehmende Schule, die diesbezüglichen Daten wurden aber in die Datenbank aufgenommen und bei den verschiedenen Schätzungen berücksichtigt. Daten von Schulen mit einer Schülerbeteiligungsquote von weniger als 25% wurden nicht in die Datenbank aufgenommen.

Für PISA 2015 wurde auch eine Mindestschülerbeteiligungsquote von 80% innerhalb der teilnehmenden Schulen festgesetzt. Diese Mindestquote musste auf Landesebene und nicht zwangsläufig von jeder teilnehmenden Schule erreicht werden. Folgesitzungen waren in Schulen erforderlich, in denen zu wenig Schülerinnen und Schüler an der ersten Testsitzung teilgenommen hatten. Die Schülerbeteiligungsquoten wurden für alle ursprünglich ausgewählten Schulen sowie für alle Schulen, d.h. die ursprünglich ausgewählten und die Ersatzschulen, auf der Basis der Teilnahme der Schülerinnen und Schüler an der ersten Testsitzung wie auch an etwaigen Folgesitzungen errechnet. Als teilnehmende Schüler galten Schüler, die bei der ersten Testsitzung oder einer Folgesitzung anwesend waren. Schüler, die sich nur an der Beantwortung der Fragebogen beteiligt hatten, wurden in die internationale Datenbank aufgenommen und bei den Statistiken in dieser Veröffentlichung berücksichtigt, wenn sie zumindest Angaben zur beruflichen Tätigkeit des Vaters oder der Mutter gemacht hatten.

Tabelle A2.3 gibt Auskunft über die Beteiligungsquoten auf Schüler- und auf Schulebene, vor und nach Einbeziehung von Ersatzschulen.

- **Spalte 1** zeigt die gewichtete Beteiligungsquote der Schulen vor Einbeziehung von Ersatzschulen. Diese wird ermittelt, indem Spalte 2 durch Spalte 3 dividiert wird.
- **Spalte 2** zeigt die gewichtete Zahl der teilnehmenden Schulen vor Einbeziehung von Ersatzschulen (gewichtete nach der Schülerzahl).
- **Spalte 3** zeigt die gewichtete Zahl der für die Stichprobe gezogenen Schulen vor Einbeziehung von Ersatzschulen (sowohl effektiv teilnehmende als auch nicht teilnehmende Schulen, gewichtet nach der Schülerzahl).
- **Spalte 4** zeigt die ungewichtete Zahl der teilnehmenden Schulen vor Einbeziehung von Ersatzschulen.
- **Spalte 5** zeigt die ungewichtete Zahl der teilnehmenden und nicht teilnehmenden Schulen vor Einbeziehung von Ersatzschulen.
- **Spalte 6** zeigt die gewichtete Schulbeteiligungsquote nach Einbeziehung von Ersatzschulen. Diese wird ermittelt, indem Spalte 7 durch Spalte 8 dividiert wird.

[Teil 1/1]

Tabelle A2.3 Beteiligungsquoten

	Ursprüngliche Stichprobe – vor Einbeziehung von Ersatzschulen					Endgültige Stichprobe – nach Einbeziehung von Ersatzschulen					Endgültige Stichprobe – Schüler innerhalb der Schulen nach Einbeziehung von Ersatzschulen				
	Gewichtete Schulbeteiligungsquote vor Einbeziehung von Ersatzschulen (in %)	Gewichtete Zahl der teilnehmenden Schulen (gew. auch nach Schülerzahlen)	Gewichtete Zahl der Stichprobenschulen (effektiv teilnehmend und nicht teilnehmend), (gew. auch nach Schülerzahlen)	Ungewichtete Zahl der teilnehmenden Schulen	Ungewichtete Zahl der teilnehmenden und nicht teilnehmenden Schulen	Gewichtete Schulbeteiligungsquote nach Einbeziehung von Ersatzschulen (in %)	Gewichtete Zahl der teilnehmenden Schulen (gew. auch nach Schülerzahlen)	Gewichtete Zahl der Stichprobenschulen (effektiv teilnehmend und nicht teilnehmend), (gew. auch nach Schülerzahlen)	Ungewichtete Zahl der teilnehmenden Schulen	Ungewichtete Zahl der teilnehmenden und nicht teilnehmenden Schulen	Gewichtete Schülerbeteiligungsquote nach Einbeziehung von Ersatzschulen (in %)	Gewichtete Zahl der teilnehmenden Schüler	Gewichtete Zahl der Stichprobenschüler (effektiv teilnehmend und nicht teilnehmend)	Ungewichtete Zahl der teilnehmenden Schüler	Ungewichtete Zahl der Stichprobenschüler (effektiv teilnehmend und nicht teilnehmend)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
OECD-Länder	94	260 657	276 072	720	788	95	262 130	276 072	723	788	84	204 763	243 789	14 089	17 477
Australien	100	81 690	81 730	269	273	100	81 690	81 730	269	273	87	63 660	73 521	7 007	9 868
Österreich	83	98 786	118 915	244	301	95	113 435	118 936	286	301	91	99 760	110 075	9 635	10 620
Belgien	74	283 853	381 133	703	1 008	79	299 512	381 189	726	1 008	81	210 476	260 487	19 604	24 129
Kanada	92	215 139	232 756	207	232	99	230 749	232 757	226	232	93	189 206	202 774	7 039	7 515
Chile	98	86 354	87 999	339	344	98	86 354	87 999	339	344	89	73 386	82 672	6 835	7 693
Tschech. Rep.	90	57 803	63 897	327	371	92	58 837	63 931	331	371	89	49 732	55 830	7 149	8 184
Dänemark	100	11 142	11 154	206	207	100	11 142	11 154	206	207	93	10 088	10 822	5 587	5 994
Estland	100	58 653	58 782	167	168	100	58 800	58 800	168	168	93	53 198	56 394	5 882	6 294
Finnland	91	679 984	749 284	232	255	94	706 838	749 284	241	255	88	611 563	693 336	5 980	6 783
Frankreich	96	764 423	794 206	245	256	99	785 813	794 206	253	256	93	685 972	735 487	6 476	6 944
Deutschland	92	95 030	103 031	190	212	98	101 653	103 218	209	212	94	89 588	94 986	5 511	5 838
Griechenland	93	83 897	89 808	231	251	99	88 751	89 825	244	251	92	77 212	83 657	5 643	6 101
Ungarn	99	4 114	4 163	122	129	99	4 114	4 163	122	129	86	3 365	3 908	3 365	3 908
Island	99	61 023	61 461	167	169	99	61 023	61 461	167	169	89	51 947	58 630	5 741	6 478
Irland	91	105 192	115 717	169	190	93	107 570	115 717	173	190	90	98 572	108 940	6 598	7 294
Israel	74	383 933	516 113	414	532	88	451 098	515 515	464	532	88	377 011	430 041	11 477	12 841
Italien	94	1 087 414	1 151 305	189	200	99	1 139 734	1 151 305	198	200	97	1 096 193	1 127 265	6 647	6 838
Japan	100	612 937	615 107	168	169	100	612 937	615 107	168	169	99	559 121	567 284	5 581	5 664
Korea	86	14 122	16 334	231	269	93	15 103	16 324	248	269	90	12 799	14 155	4 845	5 368
Lettland	100	5 891	5 891	44	44	100	5 891	5 891	44	44	96	5 299	5 540	5 299	5 540
Luxemburg	95	1 311 608	1 373 919	269	284	98	1 339 901	1 373 919	275	284	95	1 290 435	1 352 237	7 568	7 938
Mexiko	63	121 527	191 966	125	201	93	178 929	191 966	184	201	85	152 346	178 985	5 345	6 269
Niederlande	71	40 623	56 875	145	210	85	48 094	56 913	176	210	80	36 860	45 897	4 453	5 547
Neuseeland	95	58 824	61 809	229	241	95	58 824	61 809	229	241	91	50 163	55 277	5 456	6 016
Norwegen	88	314 288	355 158	151	170	99	352 754	355 158	168	170	88	300 617	343 405	4 466	5 108
Polen	86	87 756	102 193	213	254	95	97 516	102 537	238	254	82	75 391	91 916	7 180	8 732
Portugal	93	50 513	54 499	272	295	99	53 908	54 562	288	295	92	45 357	49 103	6 342	6 900
Slowak. Rep.	98	16 886	17 286	332	349	98	16 896	17 286	333	349	92	15 072	16 424	6 406	7 009
Slovenien	99	404 640	409 246	199	201	100	409 246	409 246	201	201	89	356 509	399 935	6 736	7 540
Spanien	100	93 819	94 097	202	205	100	93 819	94 097	202	205	91	82 582	91 081	5 458	6 013
Schweden	93	75 482	81 026	212	232	98	79 481	81 375	225	232	92	74 465	80 544	5 838	6 305
Schweiz	97	1 057 318	1 091 317	175	195	99	1 081 935	1 091 528	187	195	95	874 609	918 816	5 895	6 211
Türkei	84	591 757	707 415	506	598	83	654 992	707 415	547	598	89	517 426	581 252	14 120	16 123
Ver. Königreich	67	2 601 386	3 902 089	142	213	83	3 244 399	3 893 828	177	213	90	2 629 707	2 929 771	5 712	6 376
Ver. Staaten															
Partnerländer/-volkswirtschaften	100	43 809	43 919	229	230	100	43 809	43 919	229	230	94	38 174	40 814	5 213	5 555
Albanien	96	341 463	355 216	159	166	96	341 463	355 216	159	166	92	274 121	296 434	5 494	5 934
Argentinien	89	508 448	572 941	212	238	97	556 478	572 941	231	238	90	345 508	382 352	6 311	7 016
Brasilien	93	2 509 198	2 692 686	806	889	94	2 533 711	2 693 137	815	889	87	1 996 574	2 286 505	22 791	26 586
P-S-J-G (China)	88	1 259 845	1 437 201	248	268	100	1 437 652	1 437 652	268	268	97	1 287 710	1 331 794	9 841	10 097
Bulgarien	100	56 265	56 483	179	180	100	56 600	56 600	180	180	95	50 931	53 685	5 928	6 240
Kolumbien	99	664 664	673 817	364	375	100	672 526	673 835	371	375	95	535 682	566 734	11 777	12 611
Costa Rica	99	66 485	67 073	204	206	99	66 485	67 073	204	206	92	47 494	51 369	6 846	7 411
Kroatien	100	34 575	34 652	160	162	100	34 575	34 652	160	162	91	37 275	40 803	5 809	6 354
Zypern*	97	8 830	9 126	122	132	97	8 830	9 126	122	132	94	8 016	8 526	5 561	5 957
Dominik. Rep.	99	136 669	138 187	193	195	99	136 669	138 187	193	195	94	122 620	130 700	4 731	5 026
eJR Mazedonien	100	16 426	16 472	106	107	100	16 426	16 472	106	107	95	14 999	15 802	5 324	5 617
Georgien	97	40 552	41 595	256	267	99	41 081	41 566	262	267	94	35 567	37 873	5 316	5 689
Hongkong (China)	75	45 603	60 716	115	153	90	54 795	60 715	138	153	93	48 222	51 806	5 359	5 747
Indonesien	98	3 126 468	3 176 076	232	236	100	3 176 076	3 176 076	236	236	98	3 015 844	3 092 773	6 513	6 694
Jordanien	100	119 024	119 024	250	250	100	119 024	119 024	250	250	97	105 868	108 669	7 267	7 462
Kasachstan	100	202 701	202 701	232	232	100	202 701	202 701	232	232	97	187 683	192 921	7 841	8 059
Kosovo	100	26 924	26 924	224	224	100	26 924	26 924	224	224	99	22 016	22 333	4 826	4 896
Libanon	67	40 542	60 882	208	308	87	53 091	60 797	270	308	95	36 052	38 143	4 546	4 788
Litauen	99	31 386	31 588	309	311	100	31 543	31 588	310	311	91	27 070	29 889	6 523	7 202
Macau (China)	100	4 414	4 414	45	45	100	4 414	4 414	45	45	99	4 476	4 507	4 476	4 507
Malaysia	51	229 340	446 237	147	230	98	437 424	446 100	224	230	97	393 785	407 396	8 843	9 097
Malta	100	4 341	4 343	59	61	100	4 341	4 343	59	61	85	3 634	4 294	3 634	4 294
Moldau	100	30 145	30 145	229	229	100	30 145	30 145	229	229	98	28 754	29 341	5 325	5 436
Montenegro	100	7 301	7 312	64	65	100	7 301	7 312	64	65	94	6 346	6 766	5 665	6 043
Peru	100	468 406	470 651	280	282	100	469 662	470 651	281	282	99	426 205	430 959	6 971	7 054
Katar	99	13 333	13 470	166	168	99	13 333	13 470	166	168	94	12 061	12 819	12 061	12 819
Rumänien	99	171 553	172 652	181	182	100	172 495	172 495	182	182	99	162 918	164 216	4 876	4 910
Russ. Föderation	99	1 181 937	1 189 441	209	210	99	1 181 937	1 189 441	209	210	97	1 072 914	1 108 068	6 021	6 215
Singapur	97	45 299	46 620	175	179	98	45 553	46 620	176	179	93	42 241	45 259	6 105	6 555
Chinesisch Taipeh	100	286 778	286 778	214	214	100	286 778	286 778	214	214	98	246 408	251 424	7 708	7 871
Thailand	99	739 772													



- **Spalte 7** zeigt die gewichtete Zahl der teilnehmenden Schulen nach Einbeziehung von Ersatzschulen (gewichtet nach der Schülerzahl).
- **Spalte 8** zeigt die gewichtete Zahl der für die Stichprobe gezogenen Schulen nach Einbeziehung von Ersatzschulen (sowohl effektiv teilnehmende als auch nicht teilnehmende Schulen, gewichtet nach der Schülerzahl).
- **Spalte 9** zeigt die ungewichtete Zahl der teilnehmenden Schulen nach Einbeziehung von Ersatzschulen.
- **Spalte 10** zeigt die ungewichtete Zahl der teilnehmenden und nicht teilnehmenden Schulen nach Einbeziehung von Ersatzschulen.
- **Spalte 11** zeigt die gewichtete Schülerbeteiligungsquote nach Einbeziehung von Ersatzschulen. Diese wird ermittelt, indem Spalte 12 durch Spalte 13 dividiert wird.
- **Spalte 12** zeigt die gewichtete Zahl der teilnehmenden Schülerinnen und Schüler.
- **Spalte 13** zeigt die gewichtete Zahl der Stichprobenschüler (am Test teilnehmende und nicht teilnehmende Schülerinnen und Schüler).
- **Spalte 14** zeigt die ungewichtete Zahl der teilnehmenden Schülerinnen und Schüler. Zu beachten ist, dass Schüler von Schulen mit einer Schülerbeteiligungsquote von weniger als 50% nicht in die (gewichteten und ungewichteten) Quoten einbezogen wurden.
- **Spalte 15** zeigt die ungewichtete Zahl der Stichprobenschüler (am Test teilnehmende und nicht teilnehmende Schülerinnen und Schüler). Zu beachten ist, dass Schüler von Schulen, in denen weniger als die Hälfte der in Betracht kommenden Schüler am Test teilgenommen hat, nicht in den (gewichteten und ungewichteten) Quoten berücksichtigt wurden.

Definition der Schulen

In einigen Ländern wurden anstelle von Schulen Untereinheiten von Schulen für die Stichprobe herangezogen, was die Schätzung der Anteile der zwischenschulischen Varianz beeinflussen kann. In Deutschland, Japan, Österreich, Rumänien, Slowenien, der Tschechischen Republik und Ungarn wurden Schulen mit mehr als einem Bildungsgang in Einheiten unterteilt, die den jeweiligen Bildungsgängen entsprechen. In den Niederlanden wurden Schulen, die Sekundarstufe I und II umfassen, in die jeweiligen Stufen unterteilt. Was Belgien betrifft, so wurden in der Flämischen Gemeinschaft bei auf mehrere Standorte verteilten Schulen die verschiedenen Einzeleinrichtungen in die Stichprobe einbezogen, während in der Französischen Gemeinschaft im Fall solcher Schulen die jeweils größere Verwaltungseinheit berücksichtigt wurde. In Australien wurden im Fall von Schulen mit mehr als einem Standort die verschiedenen Einzeleinrichtungen erfasst. In Argentinien und Kroatien wurden im Fall solcher Schulen die einzelnen Standorte berücksichtigt. In Spanien wurden die Schulen im Baskenland, die Mehrsprachenmodelle anbieten, für die Stichprobenziehung in die einzelnen Programme unterteilt. In Luxemburg wurde eine Schule, die an der Grenze zu Deutschland liegt, entsprechend dem Wohnsitzland der Schüler unterteilt. Darüber hinaus wurden die internationalen Schulen in Luxemburg in Schüler unterteilt, die in einer der drei Amtssprachen unterrichtet wurden, sowie in Schüler vom Teil der Schulen, die mangels verfügbarer Unterlagen in der Unterrichtssprache ausgeschlossen wurden. In den Vereinigten Arabischen Emiraten wurden einige Schulen nach Lehrplänen und in manchen Fällen nach Geschlecht unterteilt, während andere Schulen als Ganzes erfasst wurden. Aufgrund der Umstrukturierung wurden in Schweden manche Schulen in mehrere Teile unterteilt, wobei jeder Teil über eine eigene Schulleitung verfügt. In Portugal wurde eine Umstrukturierung in Schulkomplexe vorgenommen, bei der die Lehrkräfte und die Schulleitung allen Einheiten des Schulkomplexes gemeinsam sind.

Klassenstufen

Die Schülerinnen und Schüler, die an den Tests von PISA 2015 teilgenommen haben, besuchen unterschiedliche Klassenstufen. Der Anteil der Schülerinnen und Schüler in den einzelnen Klassenstufen ist in Tabelle A2.4a nach Ländern geordnet aufgeführt und in Tabelle A2.4b für die einzelnen Länder nach Geschlecht aufgeschlüsselt.

[Teil 1/1]

Tabelle A2.4a Prozentsatz der Schüler in den einzelnen Klassenstufen

		Alle Schüler											
		7. Klasse		8. Klasse		9. Klasse		10. Klasse		11. Klasse		12. Klasse und darüber	
		%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.
OECD-Länder	Australien	0.0	(0.0)	0.1	(0.0)	11.2	(0.3)	74.6	(0.4)	14.0	(0.4)	0.1	(0.0)
	Österreich	0.0	(0.0)	2.0	(0.6)	20.8	(0.9)	71.2	(1.0)	5.9	(0.3)	0.0	(0.0)
	Belgien	0.6	(0.1)	6.4	(0.5)	30.7	(0.7)	61.0	(0.9)	1.3	(0.1)	0.0	(0.0)
	Kanada	0.1	(0.0)	0.7	(0.1)	10.8	(0.5)	87.6	(0.6)	0.8	(0.1)	0.0	(0.0)
	Chile	1.7	(0.3)	4.1	(0.6)	24.0	(0.7)	68.1	(1.0)	2.1	(0.2)	0.0	(0.0)
	Tschech. Rep.	0.5	(0.1)	3.9	(0.3)	49.4	(1.2)	46.2	(1.2)	0.0	(0.0)	0.0	c
	Dänemark	0.2	(0.1)	16.4	(0.6)	81.9	(0.7)	1.4	(0.5)	0.0	c	0.0	c
	Estland	0.8	(0.2)	21.3	(0.6)	76.6	(0.6)	1.3	(0.3)	0.0	c	0.0	(0.0)
	Finnland	0.5	(0.1)	13.6	(0.4)	85.7	(0.4)	0.0	(0.0)	0.2	(0.1)	0.0	c
	Frankreich	0.0	(0.0)	1.0	(0.2)	23.1	(0.6)	72.5	(0.7)	3.2	(0.2)	0.1	(0.1)
	Deutschland	0.5	(0.1)	7.7	(0.4)	47.3	(0.8)	43.1	(0.8)	1.5	(0.5)	0.0	(0.0)
	Griechenland	0.2	(0.1)	0.7	(0.2)	3.8	(0.8)	95.3	(0.9)	0.0	c	0.0	c
	Ungarn	1.7	(0.3)	8.5	(0.5)	75.8	(0.7)	14.0	(0.5)	0.0	c	0.0	c
	Island	0.0	c	0.0	c	0.0	c	100.0	c	0.0	c	0.0	c
	Irland	0.0	(0.0)	1.8	(0.2)	60.6	(0.7)	26.5	(1.1)	11.1	(0.9)	0.0	c
	Israel	0.0	c	0.1	(0.0)	16.4	(0.9)	82.7	(0.9)	0.9	(0.3)	0.0	c
	Italien	0.1	(0.0)	1.0	(0.2)	15.2	(0.6)	77.2	(0.7)	6.6	(0.3)	0.0	c
	Japan	0.0	c	0.0	c	0.0	c	100.0	(0.0)	0.0	c	0.0	c
	Korea	0.0	c	0.0	c	9.1	(0.8)	90.4	(0.8)	0.5	(0.1)	0.0	c
	Lettland	0.9	(0.2)	11.7	(0.5)	84.4	(0.6)	2.9	(0.3)	0.0	(0.0)	0.0	c
	Luxemburg	0.3	(0.1)	7.9	(0.1)	50.9	(0.1)	40.3	(0.1)	0.6	(0.0)	0.0	c
	Mexiko	2.3	(0.3)	4.8	(0.4)	31.9	(1.4)	60.3	(1.6)	0.5	(0.1)	0.2	(0.0)
	Niederlande	0.1	(0.0)	2.8	(0.3)	41.6	(0.6)	54.8	(0.6)	0.8	(0.2)	0.0	(0.0)
	Neuseeland	0.0	c	0.0	c	0.0	(0.0)	6.2	(0.3)	88.8	(0.5)	5.0	(0.5)
	Norwegen	0.0	c	0.0	c	0.6	(0.1)	99.3	(0.2)	0.1	(0.1)	0.0	c
	Polen	0.6	(0.1)	4.9	(0.3)	93.8	(0.4)	0.6	(0.2)	0.0	c	0.0	c
	Portugal	3.2	(0.3)	8.4	(0.5)	22.9	(0.9)	65.1	(1.2)	0.4	(0.1)	0.0	c
	Slowak. Rep.	2.2	(0.4)	4.6	(0.4)	42.6	(1.3)	50.6	(1.2)	0.1	(0.0)	0.0	c
	Slowenien	0.0	c	0.3	(0.1)	4.8	(0.3)	94.6	(0.4)	0.3	(0.1)	0.0	c
	Spanien	0.1	(0.0)	8.6	(0.5)	23.4	(0.6)	67.9	(0.9)	0.1	(0.1)	0.0	c
	Schweden	0.1	(0.1)	3.1	(0.4)	94.9	(0.8)	1.8	(0.7)	0.1	(0.1)	0.0	c
	Schweiz	0.5	(0.1)	11.8	(0.7)	61.3	(1.2)	25.9	(1.3)	0.5	(0.1)	0.0	(0.0)
Türkei	0.6	(0.1)	2.6	(0.4)	20.7	(1.0)	72.9	(1.2)	3.0	(0.3)	0.1	(0.0)	
Ver. Königreich	0.0	c	0.0	c	0.0	c	1.6	(0.3)	97.4	(0.4)	1.0	(0.3)	
Ver. Staaten	0.0	(0.0)	0.5	(0.3)	9.6	(0.7)	72.4	(0.9)	17.3	(0.6)	0.1	(0.0)	
Partnerländer/-volkswirtschaften	Albanien	0.2	(0.1)	1.0	(0.2)	35.8	(2.3)	61.7	(2.3)	1.2	(0.7)	0.0	(0.0)
	Algerien	18.8	(1.0)	23.5	(1.1)	35.1	(1.5)	19.4	(2.1)	3.2	(0.7)	0.0	c
	Brasilien	3.5	(0.2)	6.4	(0.4)	12.5	(0.5)	35.9	(0.9)	39.2	(0.8)	2.5	(0.2)
	P-S-J-G (China)	1.1	(0.2)	9.2	(0.7)	52.7	(1.7)	34.6	(2.0)	2.2	(0.5)	0.1	(0.0)
	Bulgarien	0.5	(0.2)	3.0	(0.6)	92.2	(0.8)	4.3	(0.4)	0.0	c	0.0	c
	Kolumbien	5.3	(0.4)	12.3	(0.6)	22.7	(0.6)	40.2	(0.7)	19.5	(0.6)	0.0	c
	Costa Rica	6.2	(0.7)	14.0	(0.7)	33.0	(1.2)	46.5	(1.6)	0.2	(0.1)	0.1	(0.1)
	Kroatien	0.0	c	0.2	(0.2)	79.2	(0.5)	20.6	(0.4)	0.0	c	0.0	c
	Zypern*	0.0	c	0.3	(0.0)	5.8	(0.1)	93.1	(0.1)	0.7	(0.1)	0.0	c
	Dominik. Rep.	7.1	(0.8)	13.8	(1.2)	20.6	(0.8)	41.9	(1.1)	14.2	(0.7)	2.4	(0.3)
	eJR Mazedonien	0.1	(0.1)	0.1	(0.1)	70.2	(0.2)	29.7	(0.2)	0.0	c	0.0	c
	Georgien	0.1	(0.0)	0.8	(0.2)	22.0	(0.8)	76.0	(0.9)	1.1	(0.3)	0.0	c
	Hongkong (China)	1.1	(0.1)	5.6	(0.4)	26.0	(0.7)	66.7	(0.7)	0.6	(0.5)	0.0	c
	Indonesien	2.1	(0.3)	8.1	(0.7)	42.1	(1.5)	45.5	(1.6)	2.3	(0.4)	0.0	(0.0)
	Jordanien	0.2	(0.1)	0.6	(0.1)	6.6	(0.4)	92.6	(0.4)	0.0	c	0.0	c
	Kosovo	0.0	(0.1)	0.6	(0.1)	24.9	(0.8)	72.4	(0.9)	2.1	(0.2)	0.0	c
	Libanon	3.7	(0.5)	8.3	(0.8)	16.6	(1.1)	62.3	(1.4)	9.0	(0.8)	0.1	(0.1)
	Litauen	0.1	(0.0)	2.6	(0.2)	86.3	(0.4)	11.0	(0.4)	0.0	(0.0)	0.0	c
	Macau (China)	2.9	(0.1)	12.2	(0.2)	29.7	(0.2)	54.5	(0.1)	0.6	(0.1)	0.0	c
	Malta	0.0	c	0.0	c	0.3	(0.1)	6.1	(0.2)	93.6	(0.1)	0.1	(0.0)
	Moldau	0.2	(0.1)	7.6	(0.5)	84.5	(0.8)	7.5	(0.8)	0.0	(0.0)	0.0	c
	Montenegro	0.0	c	0.0	c	83.7	(0.1)	16.3	(0.1)	0.0	c	0.0	c
	Peru	2.5	(0.3)	6.6	(0.4)	15.9	(0.5)	50.2	(0.8)	24.8	(0.8)	0.0	c
	Katar	0.9	(0.1)	3.5	(0.1)	16.3	(0.1)	60.7	(0.1)	18.0	(0.1)	0.6	(0.0)
	Rumänien	1.4	(0.3)	8.9	(0.5)	74.8	(0.9)	14.9	(0.7)	0.0	c	0.0	c
	Russ. Föderation	0.2	(0.1)	6.6	(0.3)	79.7	(1.5)	13.4	(1.5)	0.1	(0.0)	0.0	c
	Singapur	0.0	(0.0)	1.9	(0.3)	7.9	(0.8)	90.0	(1.0)	0.1	(0.0)	0.1	(0.0)
	Chinesisch Taipeh	0.0	c	0.0	c	35.4	(0.7)	64.6	(0.7)	0.0	c	0.0	c
	Thailand	0.2	(0.1)	0.6	(0.2)	23.8	(1.0)	72.9	(1.0)	2.4	(0.4)	0.0	c
	Trinidad und Tobago	3.3	(0.2)	10.8	(0.3)	27.3	(0.3)	56.5	(0.3)	2.2	(0.2)	0.0	c
	Tunesien	4.3	(0.3)	10.6	(0.8)	19.6	(1.3)	60.9	(1.7)	4.6	(0.4)	0.0	c
	Ver. Arab. Emirate	0.6	(0.1)	2.5	(0.3)	10.6	(0.7)	53.4	(0.8)	31.4	(0.8)	1.5	(0.1)
Uruguay	7.5	(0.6)	9.7	(0.5)	20.7	(0.7)	61.3	(1.2)	0.8	(0.1)	0.0	c	
Vietnam	0.3	(0.1)	1.7	(0.4)	7.7	(1.8)	90.4	(2.2)	0.0	(0.0)	0.0	c	
Argentinien**	1.6	(0.4)	9.7	(0.8)	27.4	(1.2)	58.5	(1.6)	2.8	(0.3)	0.0	c	
Kasachstan**	0.1	(0.1)	2.7	(0.3)	60.4	(1.7)	36.2	(1.8)	0.6	(0.1)	0.0	c	
Malaysia**	0.0	c	0.0	c	3.2	(0.6)	96.4	(0.7)	0.4	(0.3)	0.0	c	

* Vgl. Anmerkungen am Anfang dieses Anhangs.

** Der Erfassungsgrad ist zu niedrig, um die Vergleichbarkeit zu gewährleisten (vgl. Anhang A4).

StatLink  <http://dx.doi.org/10.1787/888933433129>

ANHANG A3

TECHNISCHE HINWEISE ZU DEN IN DIESEM BAND ENTHALTENEN ANALYSEN

Methoden und Definitionen

Relatives Risiko

Das relative Risiko ist eine Messgröße für die kausale Abhängigkeit zwischen einem Antezedens- und einem Konsequenzfaktor. Das relative Risiko ist nichts anderes als das Verhältnis zwischen zwei Risiken, d.h. dem Risiko der Beobachtung eines Ergebnisses bei Anwesenheit des Antezedens und dem Risiko der Beobachtung eines Ergebnisses bei Abwesenheit des Antezedens. Abbildung A3.1 stellt die im Folgenden verwendete Einteilung dar.

Abbildung A3.1 • **Bezeichnungen in einer zweidimensionalen Tabelle**

p_{11}	p_{12}	$p_{1.}$
p_{21}	p_{22}	$p_{2.}$
$p_{.1}$	$p_{.2}$	$p_{..}$

p_{ij} bezeichnet die Wahrscheinlichkeit für jede Zelle und entspricht der Zahl der Beobachtungen in einer bestimmten Zelle, dividiert durch die Gesamtzahl der Beobachtungen. $p_{i.}$, $p_{.j}$ stehen jeweils für die bedingte Wahrscheinlichkeit in jeder Zeile bzw. jeder Spalte. Die bedingte Wahrscheinlichkeit entspricht der bedingten Häufigkeit, dividiert durch die Gesamtzahl der Schüler.

Es wird angenommen, dass die Zeilen dem Antezedensfaktor entsprechen – mit der ersten Zeile für das „Vorhandensein des Antezedensfaktors“ und der zweiten Zeile für das „Nichtvorhandensein des Antezedensfaktors“ –, und dass die Spalten dem Konsequenzfaktor entsprechen, wobei in der ersten Spalte das „Vorhandensein des Konsequenzfaktors“ und in der zweiten Spalte das „Nichtvorhandensein des Konsequenzfaktors“ vermerkt ist. Das relative Risiko entspricht dann:

$$RR = \frac{(p_{11}/p_{1.})}{(p_{21}/p_{2.})}$$

Quotenverhältnis (Odds Ratio)

Für die Definition des Quotenverhältnisses, einer anderen Messgröße der relativen Wahrscheinlichkeit des Eintretens eines bestimmten Ergebnisses zwischen Gruppen, kann die gleiche Einteilung verwendet werden. Das Quotenverhältnis für die Beobachtung des Ergebnisses bei Anwesenheit des Antezedens ist nichts anderes als

$$OR = \frac{(p_{11}/p_{12})}{(p_{21}/p_{22})}$$

wobei p_{11}/p_{12} den Odds (d.h. dem Chancenverhältnis) entspricht, das Ergebnis bei Anwesenheit des Antezedens zu beobachten und p_{21}/p_{22} den Odds entspricht, das Ergebnis bei Abwesenheit des Antezedens zu beobachten.

Zur Schätzung des Odds-Ratio kann eine logistische Regression verwendet werden: Der potenzierte Logit-Koeffizient für eine binäre Variable entspricht dem Odds Ratio. Durch die Einführung von Kontrollvariablen in die logistische Regression lässt sich nach Berücksichtigung der anderen Unterschiede zwischen den Gruppen ein „generalisiertes“ Odds Ratio schätzen.

Statistiken auf der Basis von Mehrebenenmodellen

Statistiken auf der Basis von Mehrebenenmodellen umfassen Varianzkomponenten (zwischen- und innerschulische Varianz), den aus diesen Komponenten abgeleiteten Index der Inklusion sowie Regressionskoeffizienten, sofern angegeben. Mehrebenenmodelle werden im Allgemeinen als zweistufige Regressionsmodelle (Schüler- und Schulebene) mit normalverteilten Residuen spezifiziert und nach der Maximum-Likelihood-Methode geschätzt. Wenn es sich bei der abhängigen Variablen um die Leistungen in Naturwissenschaften, Lesekompetenz oder Mathematik handelt, wurden für die Schätzung zehn plausible Werte für die Ergebnisse jedes Schülers auf der jeweiligen Skala verwendet. Die Modelle wurden anhand des „Mischmoduls“ von Stata® (Version 14.1) geschätzt.

Bei Mehrebenenmodellen werden sowohl auf der Schüler- als auch auf der Schulebene Gewichtungen vorgenommen. Das Ziel dieser Gewichtungen besteht darin, den Unterschieden in der Wahrscheinlichkeit Rechnung zu tragen, dass die Schüler



in der Stichprobe gezogen werden. Da bei PISA ein zweistufiges Stichprobenziehungsverfahren angewendet wird, sind diese Unterschiede auf Faktoren sowohl auf der Schul- als auch auf der Schülerebene zurückzuführen. Für die Mehrebenenmodelle wurden endgültige Schülergewichtungen (W_{FSTUWT}) herangezogen. Die innerschulischen Gewichtungen entsprechen der endgültigen Schülergewichtung, wurden jedoch umskaliert, um dem Stichprobenumfang innerhalb jeder Schule zu entsprechen. Die zwischenschulischen Gewichtungen entsprechen der Summe der endgültigen Schülergewichtungen (W_{FSTUWT}) innerhalb jeder Schule. Die Definition der zwischenschulischen Gewichtungen ist mit der Definition in den ersten Berichten zu den Ergebnissen von PISA 2012 identisch.

Der Index der Inklusion wird folgendermaßen definiert und geschätzt:

$$100 * \frac{\sigma_w^2}{\sigma_w^2 + \sigma_b^2}$$

wobei σ_w^2 und σ_b^2 den Schätzwerten für die inner- bzw. zwischenschulische Varianz entsprechen.

Die Ergebnisse in Mehrebenenmodellen und insbesondere die Schätzung der Anteile der zwischenschulischen Varianz hängen davon ab, wie die Schulen in den einzelnen Ländern definiert und organisiert sind und welche Einheiten für die Stichprobenziehung gewählt wurden. In einigen Ländern z.B. wurden einige der Schulen in der PISA-Stichprobe als Verwaltungseinheiten definiert (selbst wenn sie, wie in Italien, mehrere geografisch getrennte Einrichtungen umfassen); in anderen Ländern wurden sie als die Teile größerer Bildungseinrichtungen definiert, die von 15-Jährigen besucht werden; in manchen Ländern wurden Schulen wiederum als Schulgebäude definiert, in wieder anderen dagegen aus Sicht der Schulorganisation (z.B. als Einheiten, die eine eigene Schulleitung haben). Der *PISA 2015 Technical Report* (OECD, erscheint demnächst) und Anhang A2 liefern einen Überblick über die verschiedenen Methoden der Definition der Schulen. In Slowenien beispielsweise ist die primäre Stichprobeneinheit definiert als eine Gruppe von Schülerinnen und Schülern, die demselben Bildungsgang innerhalb einer Schule folgen. In diesem Fall ist die Varianz zwischen Schulen tatsächlich die Varianz innerhalb von Schulen, aber zwischen unterschiedlichen Bildungsgängen. Die Verwendung von Schichtungsvariablen bei der Auswahl der Schulen kann auch die Schätzung der Anteile der zwischenschulischen Varianz beeinflussen, insbesondere wenn die Schichtungsvariablen mit zwischenschulischen Unterschieden assoziiert sind.

Wegen der Art und Weise, in der die Schülerstichprobe erhoben wurde, ist in der Varianz innerhalb der Schulen sowohl die Varianz zwischen verschiedenen Klassen als auch zwischen verschiedenen Schülerinnen und Schülern enthalten.

Effektstärken

Es kann zuweilen nützlich sein, die Indexunterschiede zwischen Gruppen, z.B. zwischen Jungen und Mädchen, in den verschiedenen Ländern miteinander zu vergleichen. Ein Problem, das in solchen Fällen auftreten kann, ist, dass die Indexverteilung zwischen den Gruppen oder Ländern variiert. Eine Möglichkeit, dieses Problem zu lösen, besteht darin, eine Effektstärke zu berechnen, die den Verteilungsunterschieden Rechnung trägt. Eine Effektstärke misst z.B. den Unterschied bei der Selbstwirksamkeit im Bereich Naturwissenschaften zwischen Jungen und Mädchen in einem gegebenen Land in Relation zu der Durchschnittsvarianz der Punktwerte bei der Selbstwirksamkeit im Bereich Naturwissenschaften unter allen Schülerinnen und Schülern in demselben Land.

Gemäß der üblichen Praxis werden in Tabelle I.3.6 Effektstärken von unter 0,20 als gering, Effektstärken im Bereich um 0,50 als mittelgroß und Effektstärken von über 0,80 als groß angegeben.

Die Effektstärke zwischen zwei Untergruppen wird berechnet als:

$$\frac{m_1 - m_2}{\sqrt{\sigma^2}}$$

wobei m_1 und m_2 jeweils die mittleren Werte für die Untergruppen 1 und 2 darstellen und σ^2 für die Gesamtvarianz (zwischen den und innerhalb der Gruppen) steht.

Konzentrationsindizes

Index der aktuellen Konzentration

Der Index der aktuellen Konzentration von Schülern mit Migrationshintergrund in Schulen auf der Ebene der Länder bzw. Volkswirtschaften entspricht dem Mindestprozentsatz der Schülerinnen und Schüler sowohl mit als auch ohne Migrationshintergrund, die von einer Schule auf eine andere umverteilt werden müssten, damit alle Schulen einen identischen Anteil

an Schülern mit Migrationshintergrund und damit auch einen identischen Anteil an Schülern ohne Migrationshintergrund aufweisen. Er ist definiert als

$$CC = \frac{\sum_{i=1}^I N_i |p_i - p|}{N}$$

wobei N_i der Zahl der Schüler in Schule i , N der Zahl der Schüler in der Population und I der Zahl der Schulen entspricht. $p_i = A_i/N_i$ dem Anteil der Schülerinnen und Schüler mit Migrationshintergrund in Schule i und $p = A/N$ den Anteil der Schülerinnen und Schüler mit Migrationshintergrund in der Schülerpopulation darstellt.

Der Index der aktuellen Konzentration S steht mit dem von Gorard und Taylor (2002) entwickelten Segregationsindex in Zusammenhang, der dem Prozentsatz der Schülerinnen und Schüler mit Migrationshintergrund entspricht, die von einer Schule auf eine andere umverteilt werden müssten, damit alle Schulen unter Berücksichtigung der ursprünglichen Größe der Schulen einen identischen Prozentsatz an Schülern mit Migrationshintergrund aufweisen. Der Segregationsindex von Gorard und Taylor wird wie folgt definiert:

$$S = 0.5 \times \sum_{i=1}^I \left| \frac{A_i}{A} - \frac{N_i}{N} \right|$$

Der Index der aktuellen Konzentration S lässt sich direkt vom Segregationsindex als $CC = 2pS$ ableiten. Der Segregationsindex von Gorard und Taylor hängt stark vom Prozentsatz der Schülerinnen und Schüler mit Migrationshintergrund in der Schülerpopulation ab. Wenn der Anteil der Schülerinnen und Schüler mit Migrationshintergrund im Land sehr gering und diese Schüler meist in einer internationalen Schule eingeschrieben sind, dann würde der Prozentsatz der Schülerinnen und Schüler mit Migrationshintergrund, die es umzuverteilen gilt, nahe bei 100% liegen. Der Index der aktuellen Konzentration reagiert weniger stark auf diesen Extremfall, hängt aber weiterhin vom Gesamtanteil der Schüler mit Migrationshintergrund in der Schülerpopulation ab.

Wenn der Index der aktuellen Konzentration anhand einer repräsentativen Stichprobe berechnet wird, ist es wichtig, Stichprobengewichtungen und -fehler mit zu berücksichtigen. Der Index der aktuellen Konzentration kann auch als Schülerdurchschnitt dargestellt werden.

$$\frac{\sum_{i=1}^I N_i |p_i - p|}{N} = \frac{1}{N} \sum_{i=1}^I N_i |p_i - p| = \frac{1}{N} \sum_{i=1}^I \sum_{j=1}^{N_i} |p_i - p|$$

Er kann daher leicht verallgemeinert und auf gewichtete Stichproben übertragen werden, indem einfach das letzte Glied der Gleichung durch einen gewichteten Durchschnitt ersetzt wird.

$$\frac{\sum_{i=1}^I \sum_{j=1}^{n_i} w_{ij} |p_i - p|}{\sum_{i=1}^I \sum_{j=1}^{n_i} w_{ij}}$$

Der Index der aktuellen Konzentration kann dann auf Schülerebene als der absolute Unterschied zwischen dem prozentualen Migrantenanteil in Schulen und dem Migrantenanteil im Land, gewichtet durch die endgültige Schüलगewichtung, berechnet werden. Standardfehler für den Index werden ermittelt, indem die endgültige Gewichtung durch die 80 Replikationsgewichtungen ersetzt wird.

Index der maximalen Konzentration

Der Index der maximalen Konzentration ist ein theoretischer Wert der maximalen Konzentration von Schülerinnen und Schülern mit Migrationshintergrund in Schulen, in Anbetracht der Größe der Schulen und Anzahl der Migranten in einem Land. Er entspricht in der kontrafaktischen Situation, in der alle Schülerinnen und Schüler mit Migrationshintergrund zunächst in den größten Schulen untergebracht werden, dem Mindestanteil der Schülerinnen und Schüler sowohl mit als auch ohne Migrationshintergrund, die von einer Schule auf eine andere umverteilt werden müssten, damit alle Schulen einen identischen Anteil an Schülern mit Migrationshintergrund aufweisen. In diesem hypothetischen Szenario ist die Konzentration insofern maximal, als die Schülerinnen und Schüler mit Migrationshintergrund nur in der kleinstmöglichen Zahl an Schulen präsent sind (in Anbetracht der Größe der Schulen und der Migrantenpopulation).

Zur Berechnung des Index müssen die Schulen in einem ersten Schritt in jedem Land in absteigender Reihenfolge ihrer jeweiligen Schulgewichtung (berechnet als die Summe der endgültigen Schüलगewichtungen in der betreffenden Schule) sortiert werden. In einem zweiten Schritt werden alle Schülerinnen und Schüler mit Migrationshintergrund bis zur gewichteten Summe der Schülerinnen und Schüler mit Migrationshintergrund in dem bestimmten Land anhand dieser Sortierung auf die Schulen



verteilt. Danach wird der oben definierte Konzentrationsindex berechnet. Standardfehler für den Index werden ermittelt, indem die endgültige Gewichtung durch die 80 Replikationsgewichtungen ersetzt wird.

Definition von Schulen mit niedriger und Schulen mit hoher Konzentration

Die Einteilung der Schulen in Schulen mit hoher und niedriger Migrantenkonzentration basiert auf einem Schwellenwert, der für jedes Land bzw. jede Volkswirtschaft spezifisch ist, so dass die Zahl der Schulen mit hoher und niedriger Konzentration nicht vom Anteil der Schülerinnen und Schüler mit Migrationshintergrund in den einzelnen Bildungssystemen abhängt. Der Schwellenwert wird definiert als der (gewichtete) Median der Verteilung der Anteile an Schülerinnen und Schülern mit Migrationshintergrund auf die Schulen. In jedem Land befinden sich nahezu 50% der Schülerinnen und Schüler in Schulen mit hoher Konzentration und 50% der Schülerinnen und Schüler in Schulen mit niedriger Konzentration.

Standardfehler und Signifikanztests

Die in diesem Bericht enthaltenen Statistiken stellen Schätzwerte der nationalen Leistung auf der Basis der Schülerstichproben dar, und nicht etwa aus den Antworten sämtlicher Schülerinnen und Schüler eines Landes auf sämtliche Fragen errechnete Werte. Daher ist es wichtig, die mögliche Höhe des Messfehlers in diesen Schätzungen zu ermitteln. In PISA wird bei jeder Schätzung ein Messfehler angegeben, der durch den Standardfehler (S.E.) ausgedrückt ist. Mit Hilfe von Konfidenzintervallen können Schlüsse über die Populationsmittelwerte und -anteile gezogen werden, die die mit den Stichprobenschätzungen verbundene Unsicherheit widerspiegeln. Unter Annahme einer Normalverteilung kann davon ausgegangen werden, dass das tatsächlich beobachtete statistische Ergebnis einer gegebenen Population in 95 von 100 Wiederholungsmessungen mit unterschiedlichen Stichproben derselben Population innerhalb des Konfidenzintervalls liegen würde.

Die Leser sind häufig in erster Linie daran interessiert, ob sich ein bestimmter Wert für ein gegebenes Land von einem zweiten Wert für dasselbe Land oder für ein anderes Land unterscheidet, z.B. ob in einem bestimmten Land Mädchen bessere Leistungen erzielen als Jungen. In den Tabellen und Abbildungen dieses Berichts werden Unterschiede als statistisch signifikant bezeichnet, wenn ein Unterschied, der absolut mindestens ebenso groß ist, in weniger als 5% der Fälle beobachtet würde, obwohl die entsprechenden Populationswerte in Wirklichkeit nicht voneinander abweichen.

Für sämtliche Teile des Berichts wurden Signifikanztests durchgeführt, um die statistische Signifikanz der vorgenommenen Vergleiche zu prüfen.

Geschlechtsspezifische Unterschiede und Unterschiede zwischen den Mittelwerten der Untergruppen

Die geschlechtsspezifischen Unterschiede bei den Schülerleistungen oder anderen Indizes wurden auf ihre statistische Signifikanz hin getestet. Positive Unterschiede weisen auf höhere Punktzahlen für Jungen hin, negative Unterschiede auf höhere Punktzahlen für Mädchen. Die in den Tabellen dieses Bands fettgedruckten Unterschiede sind im Allgemeinen bei einem Konfidenzniveau von 95% statistisch signifikant.

Desgleichen wurden die Unterschiede zwischen anderen Schülergruppen (z.B. Schülerinnen und Schüler ohne und mit Migrationshintergrund) oder Schulkategorien (z.B. benachteiligte und begünstigte Schulen) auf ihre statistische Signifikanz hin getestet. Die Definitionen der Untergruppen finden sich im Allgemeinen in den Tabellen und im Begleittext zu den Analysen. Sozioökonomisch begünstigte (benachteiligte) Schulen sind Schulen im obersten (untersten) Quartil der Verteilung des durchschnittlichen PISA-Index des wirtschaftlichen, sozialen und kulturellen Status (ESCS) auf Schulebene in den einzelnen Ländern bzw. Volkswirtschaften. Alle in den Tabellen von Anhang B dieses Berichts fettgedruckten Unterschiede sind bei einem Konfidenzniveau von 95% statistisch signifikant.

Unterschiede zwischen den Mittelwerten der Untergruppen, nach Berücksichtigung der anderen Variablen

Für viele Tabellen wurden Vergleiche des beobachteten Unterschieds („vor Berücksichtigung der anderen Variablen“) und nach Berücksichtigung der anderen Variablen wie des PISA-Index des wirtschaftlichen, sozialen und kulturellen Status der Schüler (ESCS) zwischen den Untergruppen angestellt. Die bereinigten Unterschiede wurden mittels linearer Regression geschätzt und bei einem Konfidenzniveau von 95% auf ihre Signifikanz hin getestet. Signifikante Unterschiede sind durch Fettdruck gekennzeichnet.

Leistungsunterschiede zwischen dem obersten und dem untersten Quartil der PISA-Indizes und -Skalen

Unterschiede bei den Durchschnittsleistungen zwischen dem obersten und dem untersten Quartil auf den PISA-Indizes und -Skalen wurden auf ihre statistische Signifikanz hin getestet. Angaben in Fettdruck weisen darauf hin, dass bei einem Konfidenzniveau von 95% ein statistisch signifikanter Unterschied zwischen der Leistung des obersten und des untersten Quartils der Schüler auf dem jeweiligen Index besteht.

Veränderung der Leistung je Indexeinheit

Für viele Tabellen wurde die Veränderung der Schülerleistungen je Indexeinheit berechnet. Die fettgedruckten Werte zeigen an, dass die Veränderungen bei einem Konfidenzniveau von 95% statistisch signifikant von null abweichen.

Relative Risiken und Odds Ratio

Die fettgedruckten Werte in den Tabellen von Anhang B dieses Berichts weisen darauf hin, dass das relative Risiko/Odds Ratio bei einem Konfidenzniveau von 95% statistisch signifikant von 1 abweicht. Zur Berechnung der statistischen Signifikanz um den Wert von 1 (Nullhypothese) wird unterstellt, dass die statistische Maßzahl für das relative Risiko/Odds-Ratio bei einer Nullhypothese eher einer Log-Normalverteilung als einer Normalverteilung folgt.

Für viele Tabellen werden auch „generalisierte“ Odds Ratios (nach Berücksichtigung der anderen Variablen) präsentiert. Diese Odds Ratios wurden anhand einer logistischen Regression geschätzt und gegenüber der Nullhypothese eines Odds Ratio von 1 auf ihre Signifikanz getestet (d.h. gleiche Wahrscheinlichkeit nach Berücksichtigung der anderen Variablen).

Spannweite der Rangplätze

Zur Berechnung der Spannweite der Rangplätze der Länder werden für jedes relevante Land Daten unter Verwendung des Mittelwerts und des Standardfehlers des Mittelwerts simuliert, um eine Verteilung möglicher Werte zu generieren. Es werden etwa 10 000 Simulationsrechnungen durchgeführt, und auf der Basis dieser Werte werden für jedes Land 10 000 mögliche Rangplätze erstellt. Die Zählraten werden für jedes Land und für jeden Rangplatz vom höchsten bis zum niedrigsten Wert aggregiert, bis ein Betrag von mindestens 9 500 erreicht wird. Dann wird die Spannweite der Rangplätze ausgewiesen, einschließlich aller Rangplätze, die aggregiert wurden. Das bedeutet, dass die Spannweite der Rangplätze bei einem Konfidenzniveau von mindestens 95% bestimmt wird und dass eine unimodale Verteilung der Rangplätze unterstellt werden kann. Diese Methode wurde in allen PISA-Erhebungsrounds seit 2003, einschließlich PISA 2015, verwendet.

Der Hauptunterschied zwischen der Spannweite der Rangplätze (z.B. Abb. I.2.14) und dem Vergleich der Durchschnittsergebnisse der Länder (z.B. Abb. I.2.13) besteht darin, dass die Mehrfachvergleiche bei den Rangschätzungen im erstgenannten Vergleich berücksichtigt werden, während dies im letztgenannten Vergleich nicht der Fall ist. Deshalb besteht auf der Basis eines Paarvergleichs der Ergebnisse der ausgewählten Länder manchmal ein geringfügiger Unterschied zwischen der Spannweite der Rangplätze und der Anzahl der Länder über einem bestimmten Land. So haben P-S-J-G (China) und Korea beispielsweise gemäß Abbildung I.2.13 ein ähnliches Durchschnittsergebnis und die gleiche Ländergruppe, deren Mittelwert sich nicht statistisch signifikant von ihrem Mittelwert unterscheidet, das Ranking Koreas unter den OECD-Ländern kann jedoch mit 95%iger Wahrscheinlichkeit zwischen dem 9. und dem 14. Rangplatz eingeordnet werden, während die Spannweite der Rangplätze bei P-S-J-G (China) breiter ist (zwischen dem 8. und dem 16. Rangplatz) (Abb. I.2.14). Da mit Sicherheit davon ausgegangen werden kann, dass die Verteilung der Rangschätzungen bei jedem Land unimodal ist, sollten bei der Prüfung der Länderrankings die Ergebnisse der Spannweite der Rangplätze verwendet werden.

Standardfehler in Statistiken, die auf Mehrebenenmodellen basieren

Für Statistiken auf der Basis von Mehrebenenmodellen (wie z.B. die Schätzung der Varianzkomponenten und der Regressionskoeffizienten aus zweistufigen Regressionsmodellen) erfolgt die Schätzung der Standardfehler nicht mit der üblichen Replikationsmethode, die die Schichtung und den Auswahlatz endlicher Populationen berücksichtigt. Stattdessen basieren die Standardfehler auf Modellen: Bei ihrer Berechnung wird unterstellt, dass die Schulen sowie die Schülerinnen und Schüler in den Schulen durch Zufallsstichproben aus einer theoretischen unendlichen Population von Schulen sowie Schülerinnen und Schülern ausgewählt werden, die den unterstellten Parametern des Modells entspricht (wobei die Stichprobenwahrscheinlichkeit in der Gewichtung der Schulen und Schülerinnen und Schüler widerspiegelt wird).

Der Standardfehler für den geschätzten Index der Inklusion wird berechnet, indem dafür mittels der Delta-Methode aus den (modellbasierten) Standardfehlern für die Varianzkomponenten eine approximative Verteilung abgeleitet wird.

Standardfehler bei der Trendanalyse der Ergebnisse: Linking-Fehler

Standardfehler für Leistungsvergleiche im Zeitverlauf berücksichtigen die Unsicherheit im Equating-Verfahren, das die Darstellung der Ergebnisse verschiedener PISA-Erhebungen auf ein und derselben Skala ermöglicht. Dieser zusätzliche Unsicherheitsfaktor resultiert in konservativeren Standardfehlern (die größer sind als die Standardfehler, die vor der Einführung des Linking-Fehlers geschätzt wurden) (vgl. Anhang A5 wegen einer technischen Erörterung des Linking-Fehlers).

Die fettgedruckten Werte in den Tabellen zu Leistungstrends oder Veränderungen, die in Anhang B dieses Berichts aufgeführt sind, weisen darauf hin, dass die Leistungsveränderung dieser bestimmten Gruppe bei einem Konfidenzniveau von 95% statistisch signifikant von 0 abweicht. Die für die Berechnung der statistischen Signifikanz des angegebenen Leistungstrends oder der Leistungsveränderung verwendeten Standardfehler enthalten einen Linking-Fehler.

Literaturverzeichnis

Gorard, S. und C. Taylor (2002), "What is segregation? A comparison of measures in terms of 'strong' and 'weak' compositional invariance", *Sociology*, Vol.36/4, S. 875-895, <http://dx.doi.org/10.1177/003803850203600405>.

OECD (erscheint demnächst), *PISA 2015 Technical Report*, PISA, OECD Publishing, Paris.



ANHANG A4

QUALITÄTSSICHERUNG

Wie bereits in allen früheren PISA-Erhebungen wurden in sämtlichen Teilen von PISA 2015 Qualitätssicherungsverfahren durchgeführt. Die Technischen Standards von PISA 2015 (www.oecd.org/pisa/) enthalten genaue Angaben zu den Umsetzungsmodalitäten der PISA-Erhebung in den einzelnen Ländern, Volkswirtschaften und Regionen, deren Stichproben international überprüft wurden. Internationale Vertragspartner überwachen die Umsetzung an all diesen Standorten und überprüfen die Einhaltung der Standards.

Die einheitliche Qualität und sprachliche Äquivalenz der in PISA 2015 verwendeten Erhebungsinstrumente wurden dadurch gesichert, dass beurteilt wurde, wie leicht oder schwer die englische Originalfassung übersetzt werden konnte. Es wurden zwei Originalfassungen der Erhebungsinstrumente in Englisch und Französisch erstellt (mit Ausnahme der Erhebung zur finanziellen Allgemeinbildung und der Handbücher, die nur in englischer Sprache vorliegen), damit die Länder ein doppeltes Übersetzungsverfahren durchführen, d.h. zwei unabhängige Übersetzungen aus der/den Originalsprache(n) und einen Abgleich durch eine dritte Person erstellen lassen. Es wurden genaue Anweisungen für die „Lokalisierung“ (Anpassung, Übersetzung und Validierung) der Instrumente für den Feldtest und ihrer Überprüfung für die Haupterhebung gegeben und entsprechende Übersetzungs- und Bearbeitungsrichtlinien aufgestellt. Ein unabhängiges Team erfahrener Übersetzer, das vom PISA-Konsortium eingesetzt und geschult wurde, überprüfte die einzelnen nationalen Fassungen anhand der englischen und französischen Originalfassung. Die Muttersprache der Übersetzer war die Unterrichtssprache des betreffenden Landes, und die Übersetzer verfügten über ausreichende Kenntnisse bezüglich der Bildungssysteme. Nähere Informationen über die PISA-Übersetzungsverfahren enthält der *PISA 2015 Technical Report* (OECD, erscheint demnächst).

Die Erhebung wurde mit Hilfe standardisierter Verfahren durchgeführt. Das PISA-Konsortium stellte umfassende Handbücher zur Verfügung, die den Ablauf der Erhebung erklärten und u.a. präzise Anweisungen für die Arbeit der Schulkoordinatoren sowie Vorlagen für die Testleiter zum Gebrauch bei den Testsitzungen enthielten. Vorgeschlagene Anpassungen der Erhebungsmethoden oder vorgeschlagene Änderungen der Testsitzungen wurden vor der Überprüfung dem PISA-Konsortium zur Genehmigung vorgelegt. Das PISA-Konsortium überprüfte dann die Übersetzungen und die Anpassung der Handbücher für die verschiedenen Länder.

Um die Glaubwürdigkeit von PISA im Hinblick auf Validität und Unvoreingenommenheit zu gewährleisten und einen einheitlichen Ablauf der Testsitzungen zu fördern, wurden die Testleiterinnen und Testleiter in den Teilnehmerländern nach folgenden Kriterien ausgewählt: Die Leiterinnen und Leiter der vorgesehenen PISA-Testsitzungen durften nicht zugleich die Fachlehrer der teilnehmenden Schülerinnen und Schüler in den Bereichen Lesen, Mathematik und Naturwissenschaften sein; die Testleiterinnen und Testleiter sollten nicht Mitglieder des Kollegiums einer Schule sein, an der sie die Leitung der PISA-Tests übernahmen; nach Möglichkeit sollten die Testleiterinnen und Testleiter auch nicht zum Kollegium einer für die PISA-Stichprobe gezogenen Schule gehören. Die Teilnehmerländer organisierten für die Testleiterinnen und Testleiter eine individuelle Schulung.

Es war Aufgabe der Teilnehmerländer sicherzustellen, dass die Testleiterinnen und Testleiter mit den Schulkoordinatoren bei der Vorbereitung der Testsitzung zusammenarbeiteten, u.a. bei der Überprüfung und Aktualisierung der Unterlagen über den bisherigen Bildungsweg der Schülerinnen und Schüler sowie beim Ausfüllen des Sitzungsteilnahmeprotokolls, das die Teilnahme der Schüler und die Materialzuteilung bescheinigen soll, und des Formulars für den Sitzungsbericht, das die Zeiten der Testsitzungen, etwaige Störungen während der Sitzungen usw. festhält; ferner galt es zu gewährleisten, dass die Zahl der von den Schülerinnen und Schülern eingesammelten Testhefte und Fragebogen mit der Anzahl der an die Schule (in Ländern mit papiergestützten Tests) versandten Exemplare übereinstimmt oder sicherzustellen, dass die Zahl der für den Test verwendeten USB-Sticks (in Ländern mit computergestützten Tests) geprüft wurde und die Schul-, Schüler- und Eltern- und Lehrerfragebogen (falls zutreffend) sowie das gesamte Testmaterial (ausgefüllt und nicht ausgefüllt) nach dem Test an die nationale Zentrale gesandt wurden.

Das für die Überwachung der Erhebungsverfahren verantwortliche PISA-Konsortium führte alle Verfahrensschritte des PISA-Qualitätsmonitoring (PQM) durch und stellte in allen Teilnehmerländern für das Qualitätsmonitoring in PISA zuständige Vertreter (PQM-Vertreter) ein, organisierte deren Ausbildung, wählte die zu besuchenden Schulen aus und trug die von den PQM-Vertretern bei ihren Besuchen gesammelten Informationen zusammen. Bei den PQM-Vertretern handelt es sich um unabhängige Vertragspartner in den Teilnehmerländern, die vom internationalen Vertragspartner für die Überwachung der Erhebungsverfahren eingestellt wurden. Sie besuchen eine Stichprobe von Schulen, um die Testdurchführung zu beobachten und aufzuzeichnen, wie die in der Haupterhebung dokumentierten Verfahren in den Feldoperationen umgesetzt werden.



In der Regel wurden in jedem Land zwei oder drei PQM-Vertreter eingestellt, und in jedem Land wurden im Durchschnitt 15 Schulen besucht. In Ländern mit adjudizierten Regionen mussten generell zusätzliche PQM-Vertreter eingestellt werden, da in diesen Regionen mindestens fünf Schulen untersucht wurden.

Alle im Rahmen der PISA-Erhebung 2015 gesammelten Qualitätssicherungsdaten wurden in einer zentralen Datenadjudizierungs-Datenbank zur Qualität der Feldoperationen und Verfahren, wie Druck, Übersetzung, Schul- und Schülerstichprobenauswahl sowie Kodierung erfasst und verglichen. Danach wurden ausführliche Berichte für die PISA-Adjudizierungsgruppe erstellt. Diese Gruppe bestand aus der Technischen Beratergruppe und dem Stichproben-Sachverständigen. Ihre Aufgabe ist die Überprüfung der Adjudizierungsdatenbank und -berichte, um Empfehlungen für geeignete Methoden zur Aufrechterhaltung der Qualität von PISA-Daten zu formulieren. Wegen weiterer Informationen vgl. *PISA 2015 Technical Report* (OECD, erscheint demnächst).

Die Ergebnisse der Datenbeurteilung und nachfolgender weiterer Untersuchungen zeigen, dass die technischen Standards von PISA in allen PISA-2015-Teilnehmerländern und -volkswirtschaften eingehalten wurden; Ausnahmen bilden die nachstehend aufgelisteten Länder:

- In Albanien wurde die PISA-Erhebung gemäß den Verfahrensnormen und Leitlinien der OECD durchgeführt. Aufgrund der Modalitäten der Datenerfassung waren die Testdaten jedoch nicht mit den Angaben im Schülerfragebogen kompatibel. Infolgedessen konnte Albanien nicht in die Analysen einbezogen werden, in denen die Schülerantworten im Fragebogen mit den Testergebnissen in Zusammenhang gestellt wurden.
- In Argentinien wurde die PISA-Erhebung gemäß den Verfahrensnormen und Leitlinien der OECD durchgeführt. Allerdings war der Anteil der in der PISA-Stichprobe erfassten 15-Jährigen in absoluter und relativer Rechnung deutlich zurückgegangen. In Argentinien wurde im Bereich der Sekundarschulen eine Umstrukturierung vorgenommen, von der die Schulen in der Ciudad Autónoma de Buenos Aires (CABA) ausgenommen waren, was den Erfassungsgrad der im Stichprobenrahmen aufgelisteten, in Betracht kommenden Schulen beeinflusst haben dürfte. Folglich sind die Ergebnisse Argentiniens nicht mit denen anderer Länder oder den Ergebnissen Argentiniens aus den Vorjahren vergleichbar.
- In Kasachstan hatte sich herausgestellt, dass die nationalen Kodierer bei der Beurteilung zu nachsichtig waren. Daher entsprachen die vom Menschen kodierten Items nicht den PISA-Standards und wurden von den internationalen Daten ausgeschlossen. Da die vom Menschen kodierten Items einen bedeutenden Teil der in PISA getesteten Konstrukte bilden, führte die Ausklammerung dieser Items zu einem deutlich geringeren Erfassungsgrad des PISA-Tests. Folglich sind die Ergebnisse Kasachstans nicht mit denen anderer Länder oder den Ergebnissen Kasachstans aus den Vorjahren vergleichbar.
- In Malaysia wurde die PISA-Erhebung gemäß den Verfahrensnormen und Leitlinien der OECD durchgeführt. Allerdings erreichte die gewichtete Beantwortungsquote in den ursprünglich ausgewählten Schulen in Malaysia (51%) bei weitem nicht die PISA-Standardbeantwortungsquote von 85%. Folglich sind die Ergebnisse Malaysias nicht mit denen anderer Länder oder den Ergebnissen Malaysias aus den Vorjahren vergleichbar.

Literaturverzeichnis

OECD (erscheint demnächst), *PISA 2015 Technical Report*, OECD Publishing, Paris.



ANHANG A5

ÄNDERUNGEN BEI DER DURCHFÜHRUNG UND SKALIERUNG VON PISA 2015 UND AUSWIRKUNGEN AUF DIE TRENDANALYSEN

Vergleich der Schülerleistungen in den Bereichen Naturwissenschaften, Lesekompetenz und Mathematik in den verschiedenen PISA-Zyklen

In den PISA-Erhebungen 2006, 2009, 2012 und 2015 wurde jeweils dieselbe Gesamtskala Naturwissenschaften verwendet, so dass die Punkte auf dieser Skala im Zeitverlauf unmittelbar miteinander vergleichbar sind. Dasselbe gilt für die seit PISA 2000 verwendete Gesamtskala Lesekompetenz und die seit PISA 2003 verwendete Gesamtskala Mathematik. Vergleiche der Punktwerte im Zeitverlauf sind möglich, da einige Items in allen Erhebungen zum Einsatz kamen und die aus den verschiedenen Kalibrierungen der Itemparameter abgeleiteten Leistungsskalen durch ein Equating-Verfahren angepasst werden.

Alle Schätzungen statistischer Mengen sind mit statistischer Unsicherheit behaftet, und dies gilt auch für die zum Equating der PISA-Skalen im Zeitverlauf verwendeten Transformationsparameter. Ein Linking-Fehler, der dieser Unsicherheit Rechnung trägt, wird in die Schätzung des Standardfehlers für Schätzungen der PISA-Leistungstrends und -veränderungen im Zeitverlauf einbezogen. (Wegen weiterer Einzelheiten bezüglich der Linking-Fehler vgl. die folgenden Abschnitte.)

Die Unsicherheit beim Equating der Skalen ist das Ergebnis der Veränderungen der Art und Weise, wie der Test im Lauf der Jahre administriert (z.B. Differenzen im Zusammenhang mit der Testgestaltung) und skaliert (z.B. Differenzen im Zusammenhang mit den Kalibrierungsstichproben) wird. Sie ist außerdem auf die Veränderlichkeit der Rahmenkonzepte zurückzuführen. Im Rahmen von PISA werden die Rahmenkonzepte für Naturwissenschaften, Lesekompetenz und Mathematik alle neun Jahre im Rotationsverfahren überprüft, um die jüngsten Erkenntnisse in Bezug auf die Frage zu erfassen, welche Kenntnisse und Fertigkeiten 15-Jährige erwerben müssen, um voll am Leben der Gesellschaft von morgen teilzuhaben.

Veränderungen im Hinblick auf die Testdurchführung und die Testgestaltung können einen gewissen Einfluss darauf haben, wie die Schülerinnen und Schüler auf Testitems reagieren. Veränderungen der Stichproben und der für die Skalierung verwendeten Modelle führen zu unterschiedlichen Schätzungen des Schwierigkeitsgrads der Items. Daher ist die Erfassung der Ergebnisse einer Erhebung auf der auf einer früheren Erhebung beruhenden Skala mit einer gewissen Unsicherheit behaftet. Alle PISA-Erhebungen vor 2015 unterschieden sich beispielsweise im Hinblick auf die folgenden drei Aspekte voneinander:

- *Die Testkonzeption¹.* Die Testkonzeption kann die Schülerantworten auf unterschiedliche Art und Weise beeinflussen. So nehmen die Schülerinnen und Schüler den Schwierigkeitsgrad eines Items aus dem Bereich Lesekompetenz möglicherweise anders wahr, wenn dieses zu Beginn des Tests präsentiert wird, wie dies in PISA 2000 überwiegend der Fall war, als wenn dasselbe Item an anderen Stellen im Test präsentiert wird, wie dies in späteren Erhebungen der Fall war. Ferner wenden die Schülerinnen und Schüler möglicherweise nicht dieselbe Mühe auf, wenn das Item Teil einer 30-minütigen Sequenz in Lesekompetenz inmitten eines Mathematik- und Naturwissenschaftstests ist, wie dies zumeist der Fall war, als Lesekompetenz 2003, 2006 und 2012 einen untergeordneten Bereich darstellte, wie bei Erhebungen, in denen Lesekompetenz den Schwerpunktbereich bildet. In PISA sind diese Effekte unsystematisch und fallen in der Regel gering aus, sie machen jedoch einen Teil der Unsicherheit aus, mit der die Schätzungen behaftet sind.
- *Die Kalibrierungsstichproben.* In den PISA-Erhebungen vor 2015 wurde der Schwierigkeitsgrad der Items lediglich anhand der Antworten der Schülerinnen und Schüler gemessen, die an der jeweils letzten Erhebung teilnahmen. In PISA 2009 und PISA 2012 setzte sich die Kalibrierungsstichprobe aus einer nach dem Zufallsprinzip ausgewählten Gruppe von 500 Schülerinnen und Schülern je Land bzw. Volkswirtschaft zusammen. In PISA 2000, 2003 und 2006 umfasste die Kalibrierungsstichprobe lediglich Schülerinnen und Schüler aus den OECD-Ländern (500 je Land) (OECD, 2009). Dies lässt darauf schließen, dass es für jedes Trenditem so viele (unabhängige) Schätzungen des Schwierigkeitsgrads gab wie Erhebungen, in denen das betreffende Item verwendet wurde. Diese Schätzungen waren nicht identisch, und die Schwankungen zwischen dem geschätzten Schwierigkeitsgrad der Items tragen zu der Unsicherheit bei, mit der die Vergleiche zwischen den verschiedenen PISA-Erhebungen behaftet sind. Die Verwendung lediglich einer Teilstichprobe der im Rahmen von PISA erhobenen Schülerdaten je Land erhöht diese Unsicherheit weiter und lässt sich durch die begrenzte Rechenleistung rechtfertigen, die zum Zeitpunkt der ersten PISA-Erhebungen zur Verfügung stand.
- *Die Auswahl und Zahl der bereits in früheren Erhebungen verwendeten Items.* Ebenso wie die Unsicherheit, mit der die Durchschnittsergebnisse der Länder und die Itemparameter behaftet sind, durch die Einbeziehung einer größeren Zahl von Schulen und Schülern in die Stichprobe verringert wird, wird die Unsicherheit im Hinblick auf die Herstellung eines Zusammenhangs zwischen den Skalen (Linking) gesenkt, indem beim Linking eine größere Zahl von bereits in früheren Erhebungen verwendeten Items berücksichtigt wird. Was den Schwerpunktbereich (in PISA 2015 z.B. Naturwissenschaften) betrifft, so stellen die Items, die bereits in früheren Erhebungen verwendet wurden, eine Untergruppe der Gesamtheit der Items dar, aus denen sich die Erhebung zusammensetzt, da der Item-Pool in PISA nach und nach erneuert wird, um jeweils den neuesten Rahmenkonzepten

Rechnung zu tragen. Die Rahmenkonzepte beruhen auf den aktuellen Erkenntnissen darüber, welche Lese-, Mathematik- und naturwissenschaftlichen Kompetenzen 15-Jährigen abverlangt werden, damit sie voll an der Gesellschaft teilhaben können.

Im Rahmen von PISA 2015 wurden verschiedene Verbesserungen an der Testkonzeption und dem Skalierungsverfahren eingeführt, mit denen auf die Verringerung der drei oben beschriebenen Unsicherheitsquellen abgezielt wird. Insbesondere wurden die Unterschiede bei der Konstrukterfassung in den verschiedenen Bereichen und die Wahrnehmung der Schülerinnen und Schüler von bestimmten Bereichen als „Schwerpunktbereich“ oder als „untergeordneter Bereich“ durch die Testgestaltung im Rahmen von PISA 2015 verringert bzw. beseitigt. Bei der am häufigsten umgesetzten Testversion (der computergestützten Version in Ländern, die das Problemlösen im Team testeten) beispielsweise wurden 86% der Schülerinnen und Schüler lediglich in zwei Bereichen jeweils eine Stunde lang getestet (33% in Naturwissenschaften und Lesekompetenz, 33% in Naturwissenschaften und Mathematik und 22% in Naturwissenschaften und Problemlösen im Team, wobei für die Hälfte jeder Gruppe die Reihenfolge umgekehrt wurde) (vgl. OECD (erscheint demnächst) wegen näherer Einzelheiten). Die Zahl der bereits in früheren Erhebungen verwendeten Items wurde zudem für alle Bereiche deutlich erhöht, insbesondere in den untergeordneten Bereichen. Als Lesekompetenz ein untergeordneter Bereich war (2003 und 2006), wurde beispielsweise nur eine Zahl von Items, die einer Testdauer von einer Stunde bzw. zwei 30-minütigen Blöcken entsprach, verwendet, um das Linking mit PISA 2000 zu unterstützen; als Mathematik 2012 zum zweiten Mal den Schwerpunktbereich bildete, entsprach die Zahl der Items, mit denen das Linking mit PISA 2003 erfolgte, einer Testdauer von anderthalb Stunden. 2015 betrug die Testdauer sowohl für den Schwerpunktbereich Naturwissenschaften als auch für Lesekompetenz und Mathematik drei Stunden, um das Linking mit den bereits existierenden Skalen zu unterstützen.

Auch das Skalierungsverfahren wurde verbessert, indem die Kalibrierungsstichprobe auf der Grundlage aller Antworten der Schülerinnen und Schüler aus den vergangenen vier Erhebungsrounden gebildet wurde. Diese umfasst für alle Bereiche eine Erhebung, in der der jeweilige Bereich den Schwerpunktbereich bildete; für den Schwerpunktbereich reicht die Stichprobe bis zur vorherigen Erhebungsrounde zurück, in der er den Schwerpunktbereich bildete. Bei der nächsten PISA-Erhebung (2018) wird sich die Kalibrierungsstichprobe um bis zu rd. 75% mit der Erhebungsrounde von 2015 überschneiden. Infolgedessen verringert sich die Unsicherheit aufgrund der Neuschätzung der Itemparameter bei der Skalierung verglichen mit den Erhebungsrounden bis 2012 erheblich.

Diese Verbesserungen werden zwar voraussichtlich zu Verringerungen des Linking-Fehlers zwischen PISA 2015 und künftigen Erhebungsrounden führen, sie können jedoch die Unsicherheit erhöhen, die sich in den Linking-Fehlern zwischen 2015 und früheren Erhebungsrounden widerspiegelt, da die früheren Erhebungsrounden eine andere Testkonzeption aufwiesen und einem anderen Skalierungsverfahren folgten.

Zusätzlich wurden in PISA 2015 weitere Änderungen im Hinblick auf die Testadministration und die Skalierung eingeführt:

- Änderung des Testmodus. 2015 wurde der computergestützte Test zum wesentlichen Durchführungsmodus des PISA-Tests. Alle in PISA 2015 verwendeten Trenditems wurden für die Bearbeitung am Computer angepasst. Die Äquivalenz zwischen der papiergestützten und der computergestützten Version der Trenditems, die zur Messung der Schülerleistungen in Naturwissenschaften, Lesekompetenz und Mathematik verwendet wurden, wurde im Rahmen eines umfassenden Feldtests, der in allen Teilnehmerländern und -volkswirtschaften von PISA 2015 durchgeführt wurde, anhand einer heterogenen Schülerpopulation aus allen Ländern und Volkswirtschaften geprüft, die an der PISA-Erhebung 2015 teilnahmen. Auf den Ergebnissen dieser Moduseffektuntersuchung im Hinblick auf die Äquivalenz der Items („skalare“ Äquivalenz oder „metrische“ Äquivalenz; vgl. u.a. Meredith, 1993; Davidov, Schmidt und Billiet, 2011) beruht die Skalierung der Antworten der Schülerinnen und Schüler in der Haupterhebung. Die Parameter der skalar bzw. metrisch invarianten Items wurden so beschränkt, dass sie für die gesamte Kalibrierungsstichprobe identisch sind, einschließlich der Teilnehmer, die sie im papierbasierten bzw. computergestützten Modus bearbeiteten (vgl. den Abschnitt zum „Vergleich von PISA-Ergebnissen zwischen papierbasiertem und computergestütztem Erhebungsmodus“ wegen näherer Einzelheiten).
- Änderung des Skalierungsmodells. Bei der Skalierung der Itemparameter wurde ein flexibleres statistisches Modell auf die Antworten der Schülerinnen und Schüler zugeschnitten. Dieses Modell, dessen allgemeinste Form das generalisierte Modell abgestufter Punktwerte – Generalised Partial Credit Model – ist (d.h. ein zweiparametrisches Item-Response-Theorie-Modell, vgl. Birnbaum, 1968; Muraki, 1992), enthält Beschränkungen für Trenditems, um so viele Trenditems mit einparametrischen Likelihood-Funktionen zu berücksichtigen, wie die Daten unterstützen, und wird daher als „hybrides“ Modell bezeichnet. Die einparametrischen Modelle, auf denen die Skalierung in den früheren Erhebungsrounden beruhte (Rasch, 1960; Masters, 1982), stellen einen Sonderfall des gegenwärtig verwendeten Modells dar. Der Hauptunterschied zwischen dem gegenwärtig verwendeten hybriden Modell und den zuvor verwendeten einparametrischen Modellen besteht darin, dass das hybride Modell bei der Konstruktion einer Punktzahl nicht alle Items gleich gewichtet, sondern den Aufgaben vielmehr optimale Gewichte zuordnet, die auf ihrer Kapazität beruhen, zwischen Schülerinnen und Schülern mit großen bzw. geringen Fähigkeiten zu unterscheiden. Daher lässt sich der Vielfalt der in den PISA-Tests verwendeten Antwortformate in diesem Modell besser Rechnung tragen.
- Änderung der Berücksichtigung der differenziellen Itemfunktion in den einzelnen Ländern. In Tests wie PISA, bei denen die Items in viele verschiedene Sprachen übersetzt werden, funktionieren manche Items in einigen Ländern möglicherweise anders, als sie in der Mehrzahl der Länder funktionieren. So lassen sich etwa Begriffe, die sich in eine bestimmte Sprache schwerer übersetzen



lassen, nicht immer vermeiden. Die daraus folgenden Interaktionen zwischen Item und Land gefährden potenziell die Validität. In den vergangenen Erhebungsrounden wurden für alle Länder einheitliche Itemparameter verwendet, mit Ausnahme einer sehr geringen Zahl von Items, die als „fragwürdig“ eingestuft wurden und daher für einige Länder als „nicht administriert“ behandelt wurden (dies betraf in der Regel weniger als eine Handvoll Items, beispielsweise wenn nur spät im Verfahren Flüchtigkeitsfehler in der Übersetzung oder beim Druck entdeckt wurden). 2015 erlaubte die Kalibrierung eine begrenzte Zahl länder- und erhebungsspezifischer Abweichungen von den internationalen Itemparametern (Glas und Jehangir, 2014; Oliveri und von Davier, 2011; Oliveri und von Davier, 2014). Bei diesem Ansatz wird die Vergleichbarkeit der PISA-Punktzahlen zwischen den einzelnen Ländern und im Zeitverlauf gewahrt, die durch das Vorhandensein einer ausreichenden Zahl invarianter Items gesichert wird, und zugleich die (eingeschränkte) Abhängigkeit der Länderrangfolgen von der Auswahl der in der Erhebung berücksichtigten Items verringert, wodurch die Fairness erhöht wird. Im *Technical Report für PISA 2015* ist die Zahl der spezifischen Parameter für die einzelnen PISA-Teilnehmerländer und -volkswirtschaften angegeben (OECD, erscheint demnächst).

- Änderung der Berücksichtigung der nicht erreichten Items. Zu guter Letzt wurden die nicht erreichten Items (d.h. die unbeantworteten Items am Ende der Testhefte) in PISA 2015 als nicht administriert behandelt, wohingegen sie in den früheren PISA-Erhebungsrounden für die Schätzung des Leistungsniveaus der Schülerinnen und Schüler (d.h. bei der „Punktzahlvergabe“) als Falschantworten, bei der Schätzung der Itemparameter (der „Skalierung“) hingegen als nicht administriert gewertet worden waren. Diese Änderung sorgt für eine konsistente Berücksichtigung der Schülerantworten bei der Schätzung der Itemparameter und der Schülerleistungen und beseitigt den potenziellen Vorteil zugunsten von Ländern und Testteilnehmern, die Multiple-Choice-Fragen, die sie nicht rechtzeitig bearbeiten konnten, auf gut Glück beantworteten, gegenüber Testteilnehmern, die diese nicht erreichten Items unbeantwortet ließen³. Diese neue Berücksichtigung der nicht erreichten Items könnte für Länder mit vielen unbeantworteten Items jedoch zu höheren Punktzahlen führen, als in der Vergangenheit geschätzt worden wäre.

Die Verknüpfung der Ergebnisse von PISA 2015 mit den bereits existierenden Vergleichsskalen (Linking)

In diesem Abschnitt wird beschrieben, wie die Ergebnisse von PISA 2015 umgewandelt wurden, um die Ergebnisse von PISA 2015 auf den bereits existierenden PISA-Skalen (der in PISA 2000 definierten Gesamtskala Lesekompetenz, der in PISA 2003 definierten Gesamtskala Mathematik und der in PISA 2006 definierten Gesamtskala Naturwissenschaften) zu erfassen.

Bei der Schätzung der Itemparameter für 2015 auf der Grundlage der Antworten der Schülerinnen und Schüler, die an den Erhebungsrounden 2006, 2009, 2012 und 2015 teilnahmen, wurde unterstellt, dass diese Antworten von M verschiedenen Populationen stammten, wobei M die Gesamtzahl der Länder und Volkswirtschaften ist, die an PISA teilnahmen, multipliziert mit der Zahl der Erhebungsrounden, an denen sie teilnahmen (Mehrgruppenmodell). Jede Population mij (wobei i für das Land und j für die Erhebungsrunde steht) ist durch einen bestimmten Mittelwert und eine bestimmte Leistungsvarianz charakterisiert⁴. Die Leistungsdurchschnitte und Standardabweichungen gehörten zu den Parametern, die zusammen mit den Itemparametern anhand des Skalierungsmodells geschätzt wurden. (Wie in den vorherigen Erhebungsrounden wurden die einzelnen Schätzungen der Leistungen erst in einem zweiten Schritt berechnet, der für jedes Land bzw. jede Volkswirtschaft separat durchgeführt wurde. Dieser Schritt der „Punktzahlvergabe“ war nur für die Erhebungsrunde von 2015 erforderlich und wurde für diese durchgeführt.) Das Ergebnis der Skalierung ist eine verlinkte Skala, die auf der Annahme der Invarianz der Itemfunktionen zwischen den Erhebungsrounden von 2006, 2009, 2012 und 2015 beruht und auf der die Mittelwerte und Standardabweichungen der Länder im Zeitverlauf direkt vergleichbar sind.

Um die bei der Skalierung erstellte Skala an die bereits vorhandene numerische Skala anzupassen, die zur Erfassung der PISA-Ergebnisse aus früheren Erhebungsrounden verwendet wurde, wurde eine lineare Transformation auf die Ergebnisse angewendet. Die für diese Transformation verwendeten Parameter für den Achsenabschnitt und die Steigung wurden durch den Vergleich der während der Skalierung auf der Logit-Skala geschätzten Mittelwerte und Standardabweichungen der Länder und Volkswirtschaften mit den entsprechenden in den vergangenen Erhebungsrounden erzielten und in den PISA-Berichten veröffentlichten Mittelwerten und Standardabweichungen auf der PISA-Skala definiert. Konkret beruhte die Transformation im Bereich Naturwissenschaften auf dem Vergleich der mittleren Punktzahl im OECD-Durchschnitt und der Standardabweichung (innerhalb der einzelnen Länder) mit der mittleren Punktzahl im OECD-Durchschnitt und der Standardabweichung (innerhalb der einzelnen Länder) in PISA 2006. Bei dieser Transformation wurde die Definition der PISA-Skala als Skala „mit einem Mittelwert von 500 und einer Standardabweichung von 100 in den OECD-Ländern, wenn ein Bereich erstmals Schwerpunktbereich ist“ gewahrt. Ein ähnliches Verfahren wurde für Mathematik (wobei die durchschnittlichen Mittelwerte und Standardabweichungen der OECD-Länder an die letzte Erhebungsrunde angepasst wurden, in der Mathematik Schwerpunktbereich war, d.h. 2012) und Lesekompetenz (wobei die neu geschätzten Ergebnisse an die erfassten Ergebnisse von 2009 angepasst wurden) verwendet.

Untersuchung des Effekts der 2015 eingeführten Änderungen des Skalierungsansatzes auf die trendmäßigen Entwicklungen

Es ist möglich, zu schätzen, wie die Mittelwerte der Länder in der Vergangenheit ausgefallen wären, wenn der gegenwärtige Ansatz zur Skalierung der Schülerantworten auf die vergangenen Erhebungsrounden angewendet worden wäre. Dieser Abschnitt enthält Angaben zum Vergleich der in den früheren PISA-Berichten (z.B. OECD, 2014b) veröffentlichten Mittelwerte mit den bei der Skalierung in PISA 2015 berechneten Mittelwerten der Länder und Volkswirtschaften.

In Tabelle A5.1 sind die Korrelationen zwischen zwei Kategorien von Ländermittelwerten für 2006, 2009, 2012 und 2015 dargestellt: denjenigen, die in den Tabellen in Anhang B aufgeführt sind und im gesamten Bericht erörtert werden, und den geschätzten Mittelwerten, die auf denselben Daten beruhen, aber nach dem Skalierungsansatz von 2015 berechnet wurden und das Ergebnis des oben beschriebenen Mehrgruppenmodells sind. Die Unterschiede bei den Mittelwerten resultieren möglicherweise aus der Nutzung größerer Kalibrierungsstichproben, die Daten aus mehreren Erhebungsrounden poolen, aus der neuen Art und Weise, wie differenzielle Itemfunktionen in den einzelnen Ländern sowie nicht erreichte Items berücksichtigt werden, oder aus der Nutzung eines hybriden IRT-Modells anstelle der in den früheren Erhebungsrounden verwendeten einparametrischen Modelle. Die Spalte, die sich auf 2015 bezieht, veranschaulicht die Größenordnung der Differenzen aufgrund der Berechnung der Punktzahlen während des Schritts der Punktzahlvergabe, die zu vernachlässigen ist.

Tabelle A5.1. Korrelation der Ländermittelwerte nach alternativen Skalierungsansätzen

Im Vergleich aller an PISA 2015 teilnehmenden Länder und Volkswirtschaften

	2006	2009	2012	2015
Naturwissenschaften	0.9941	0.9961	0.9966	0.9997
Lesekompetenz	0.9850	0.9949	0.9934	0.9992
Mathematik	0.9953	0.9974	0.9973	0.9995

Anmerkung: Diese Tabelle gibt den Korrelationskoeffizienten zwischen den in Anhang B enthaltenen geschätzten Mittelwerten auf der Grundlage der erhebungsrundenspezifischen Skalierungsansätze und den nach dem Skalierungsansatz von 2015 generierten Mittelwerten für die A-posteriori-Verteilungen an.

StatLink  <http://dx.doi.org/10.1787/888933433162>

Die für die Jahre 2006, 2009 und 2012 in dieser Tabelle aufgeführten starken Korrelationen (allesamt stärker als 0,993, mit der Ausnahme der Lesekompetenz im Jahr 2006, wofür die Korrelation 0,985 betrug) lassen darauf schließen, dass die relative Position der Länder auf der PISA-Skala von den 2015 im Skalierungsansatz eingeführten Änderungen kaum beeinträchtigt wird. Die Größenordnung dieser Korrelationen zwischen nach unterschiedlichen Methoden berechneten Schätzwerten ist zudem auch höher als die Größenordnung der Korrelationen der Mittelwerte zwischen den aufeinanderfolgenden PISA-Erhebungen und deutlich höher als die Größenordnung der Korrelationen der Mittelwerte zwischen zwei Erhebungsrounden, in denen derselbe Bereich Schwerpunktbereich ist (in Neunjahresintervallen)⁵. Dies bedeutet, dass Änderungen der Methodik bestenfalls einen kleinen Teil der in PISA erfassten Veränderungen und Trends ausmachen.

Vergleich der Ländermittelwerte nach einem einheitlichen Skalierungsansatz

Nachdem die während der Skalierung der Itemparameter berechneten Ländermittelwerte auf die im vorigen Abschnitt beschriebene Art und Weise umgewandelt wurden, können sie verwendet werden, um für jedes Land die Reagibilität der im Haupttext beschriebenen und in den in Anhang B enthaltenen Tabellen dargestellten Trends im Hinblick auf die 2015 eingeführten Änderungen des Skalierungsansatzes und der Kalibrierungsstichproben zu untersuchen⁶. Diese umgewandelten Mittelwerte sind für Naturwissenschaften in Tabelle A5.3, für Lesekompetenz in Tabelle A5.4 und für Mathematik in Tabelle A5.5 aufgeführt.

Für eine große Mehrheit der Länder und Volkswirtschaften liegen die Differenzen zwischen den in Anhang B angegebenen Mittelwerten und den in Tabelle A5.3, A5.4 und A5.5 angegebenen Mittelwerten eindeutig innerhalb des mit dem Linking-Fehler verbundenen Konfidenzintervalls (s.u.). Es gibt jedoch einige nennenswerte Ausnahmen (Abb. A5.1, A5.2 und A5.3). Insbesondere treten bei der Betrachtung der Veränderungen zwischen 2015 und dem letzten Mal, als ein Bereich den Schwerpunktbereich bildete, die folgenden Beobachtungen zutage:

Naturwissenschaften

- Die für Kolumbien angegebene Verbesserung bei den durchschnittlichen Leistungen in Naturwissenschaften ist fast vollständig auf die Änderungen beim Skalierungsansatz zurückzuführen. Die Verbesserung des Durchschnittsergebnisses hätte lediglich drei Punkte betragen (und wäre damit nicht signifikant gewesen), wenn der Ansatz und die Kalibrierungsstichprobe von 2015 für die Skalierung der Ergebnisse von 2006 verwendet worden wären. In geringerem Maße sind auch die nicht signifikanten Anstiege der für Chile, Brasilien, Indonesien und Uruguay angegebenen Durchschnittsergebnisse durch die Veränderungen der Kalibrierungsstichprobe und des Skalierungsansatzes bedingt. Diese vier Länder hätten weniger positive (aber höchstwahrscheinlich immer noch nicht signifikante) Trends verzeichnet, wenn die in der Vergangenheit ermittelten Durchschnittsergebnisse auf der Grundlage des Skalierungsansatzes von PISA 2015 angegeben worden wären. Es ist nicht möglich, mit Sicherheit zu ermitteln, welche Unterschiede zwischen der ursprünglichen Skalierung der Daten von PISA 2006 und der neuen Skalierung von PISA 2015 diese Ergebnisse bewirkt haben. Ein wahrscheinlicher Grund für diese Unterschiede ist jedoch die neue Berücksichtigung der nicht erreichten Items. In all diesen Ländern erreichten viele Schülerinnen und Schüler nicht die am Ende der Testhefte oder Testformulare befindlichen Items.
- Die Vereinigten Staaten weisen zwischen 2006 und 2015 eine nicht signifikante Verbesserung (von sieben Punkten) im Bereich Naturwissenschaften auf. Die Verbesserung wäre etwas größer ausgefallen und wäre höchstwahrscheinlich als signifikant eingestuft worden (+15 Punkte), wenn der Ansatz und die Kalibrierungsstichprobe von 2015 zur Skalierung der Ergebnisse von



2006 verwendet worden wären. Die nach dem Skalierungsansatz von 2015 beobachtete Veränderung ist zwar größer als die angegebene Veränderung, liegt aber nach wie vor im Konfidenzintervall für die angegebene Veränderung.

Lesekompetenz

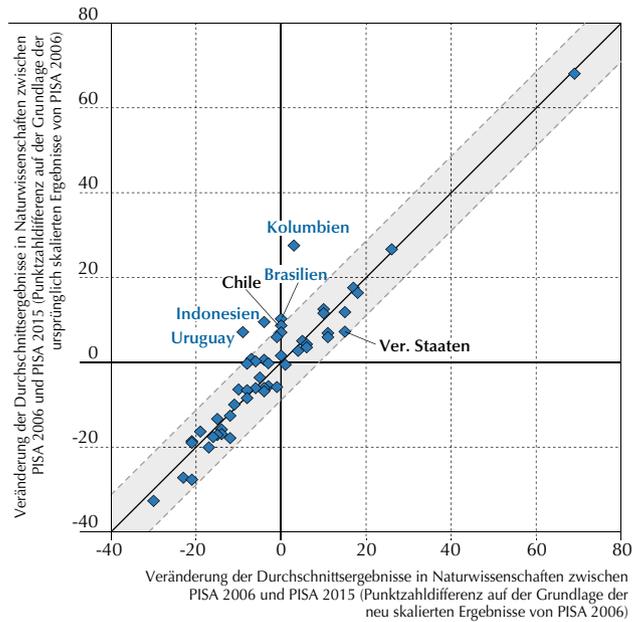
- Die für Korea angegebene negative Veränderung zwischen PISA 2009 und PISA 2015 (-22 Punkte) ist zum großen Teil auf den Unterschied beim Skalierungsansatz zurückzuführen. Wären die Ergebnisse von PISA 2009 für den Bereich Lesekompetenz mit der Kalibrierungsstichprobe und dem Skalierungsansatz von PISA 2015 skaliert worden, hätte die Differenz bei den Ergebnissen für Korea nur -9 Punkte betragen, was wahrscheinlich nicht als signifikant eingestuft worden wäre. Dem Skalierungsmodell von PISA 2015 zufolge wurden die in der Vergangenheit erfassten Ergebnisse in Lesekompetenz für Korea etwas zu hoch angegeben. Es ist aufgrund dieser Ergebnisse nicht möglich, mit Sicherheit zu ermitteln, welcher Aspekt des Ansatzes von PISA 2015 für die Differenz maßgeblich ist. Ein wahrscheinlicher Grund liegt jedoch in der neuen Art und Weise der Berücksichtigung differenzieller Itemfunktionen. Tatsächlich handelt es sich bei den meisten Items, die ein mittleres Niveau an differenzieller Itemfunktion für Korea aufwiesen und daher bei der Kalibrierung im Rahmen von PISA 2015 länderspezifische Parameter erhielten, um Items, bei denen die Schülerinnen und Schüler in Korea in den vergangenen PISA-Erhebungsrunden erfolgreicher waren, als die internationalen Parameter vorhergesagt hatten. In geringerem Maße weist Thailand ein ähnliches Muster auf. Die erfasste negative Veränderung (-12 Punkte) wäre als nicht signifikant eingestuft worden (-3 Punkte), wenn der Vergleich mit neu skalierten Ergebnissen für 2009 erfolgt wäre.
- Dänemark weist zwischen PISA 2009 und PISA 2015 eine nicht signifikante Verbesserung (von fünf Punkten) auf. Nach dem Ansatz von PISA 2015 hätte die Verbesserung jedoch 15 Punkte betragen und wäre höchstwahrscheinlich als signifikant eingestuft worden.
- Estland weist eine signifikante Verbesserung von 18 Punkten auf, die Verbesserung hätte jedoch lediglich 10 Punkte betragen, wenn die Ergebnisse von PISA 2009 anhand des Skalierungsmodells von PISA 2015 ermittelt worden wären.
- Die Niederlande weisen zwischen PISA 2009 und PISA 2015 eine nicht signifikante Verschlechterung (von fünf Punkten) auf. Nach dem Ansatz von PISA 2015 hätten die Niederlande jedoch einen (höchstwahrscheinlich nicht signifikanten) Anstieg um 4 Punkte verzeichnet.
- Die für Kolumbien, Trinidad und Tobago und Uruguay angegebene Verbesserung bei den durchschnittlichen Leistungen in Lesekompetenz ist höchstwahrscheinlich auf Änderungen des Skalierungsansatzes zurückzuführen. Die Veränderung der mittleren Punktzahl hätte in der Nähe von 0 Punkten gelegen (und wäre als nicht signifikant eingestuft worden), wenn der Ansatz und die Kalibrierungsstichprobe von 2015 für die Skalierung der Ergebnisse von 2009 verwendet worden wären. Auch die Erhöhung der mittleren Punktzahl für Peru und Moldau hätte nach einem einheitlichen Skalierungsansatz lediglich 15 Punkte bzw. 21 Punkte betragen (gegenüber einem angegebenen Anstieg um 28 Punkte). Ein wahrscheinlicher Grund für diese Unterschiede ist die neue Art und Weise der Berücksichtigung der nicht erreichten Items. In all diesen Ländern erreichten viele Schülerinnen und Schüler nicht die am Ende der Testhefte oder Testformulare befindlichen Items.

Mathematik

- Die negativen Veränderungen zwischen PISA 2012 und PISA 2015, die für Chinesisch Taipeh (-18 Punkte) und Vietnam (-17 Punkte) angegeben wurden, sind weitgehend auf die Verwendung eines anderen Skalierungsansatzes zurückzuführen. Wären die Ergebnisse von PISA 2012 für den Bereich Mathematik mit der Kalibrierungsstichprobe und dem Skalierungsansatz von PISA 2015 skaliert worden, hätte die Differenz bei den Ergebnissen für Chinesisch Taipeh und Vietnam nur -3 Punkte bzw. -4 Punkte betragen, was höchstwahrscheinlich nicht als signifikant eingestuft worden wäre. Die neue Art und Weise der Berücksichtigung differenzieller Itemfunktionen könnte der Hauptgrund für diese Differenzen sein.
- Die erfasste Veränderung für die Türkei zwischen PISA 2012 und PISA 2015 (-28 Punkte) hätte lediglich -18 Punkte betragen, wenn alle Ergebnisse nach dem Skalierungsansatz von 2015 generiert worden wären. Der dargestellte Trend verstärkt zwar den Umfang der Veränderung, die Richtung und die Signifikanz der Veränderung sind nach den beiden Ergebniskategorien jedoch ähnlich.
- Die Verbesserung bei der mittleren Punktzahl von Albanien in Mathematik zwischen PISA 2012 und PISA 2015 (+19 Punkte) wäre geringer ausgefallen und höchstwahrscheinlich als nicht signifikant eingestuft worden (+7 Punkte), wenn alle Ergebnisse nach einem einheitlichen Skalierungsansatz generiert worden wären. Ein wahrscheinlicher Grund für diese Differenz ist die neue Art und Weise der Berücksichtigung der nicht erreichten Items. Auch der für Uruguay angegebene nicht signifikante Anstieg (+9 Punkte) hätte bei einem einheitlichen Skalierungsansatz noch näher bei null gelegen (+1 Punkt).
- Singapur weist eine Verschlechterung der mittleren Punktzahl von 9 Punkten auf, die in Anbetracht des geringeren Stichprobenfehlers für dieses Land als signifikant eingestuft wird. Wären die Ergebnisse von PISA 2012 anhand des Skalierungsmodells von PISA 2015 ermittelt worden, wären sie jedoch 7 Punkte niedriger als die veröffentlichten Ergebnisse ausgefallen; folglich hätte der Unterschied zu den Ergebnissen von PISA 2015 bei einem einheitlichen Skalierungsansatz nur -2 Punkte betragen.

Alle anderen Differenzwerte zwischen den angegebenen Veränderungen und jenen, die darauf basieren, ob der Skalierungsansatz von PISA 2015 auf die früheren PISA-Erhebungen angewendet wird, sind geringer als die Differenzwerte, die aufgrund der in den folgenden Abschnitten dieses Anhangs angegeben Linking-Fehler zu erwarten sind.

Abbildung A5.1 ■ **Veränderungen bei den Leistungen in Naturwissenschaften zwischen 2006 und 2015, basierend auf den ursprünglich bzw. neu skalierten Ergebnissen**

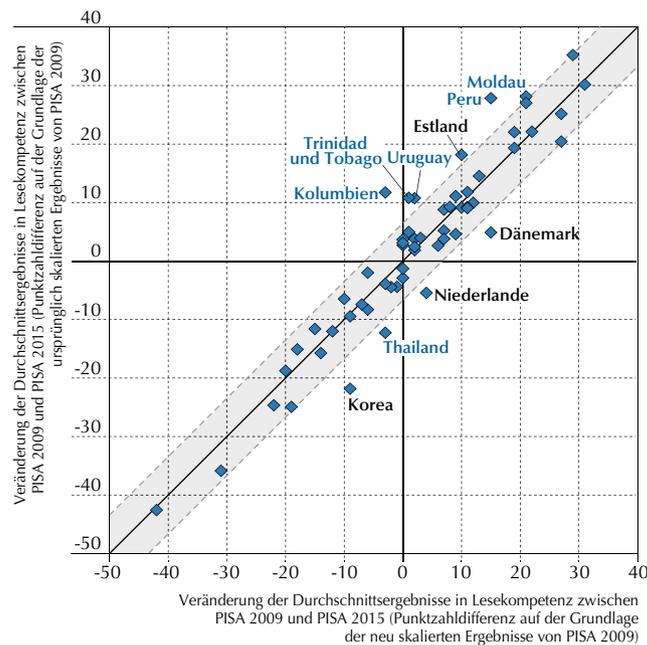


Anmerkung: Die durchgezogene Linie gibt die Diagonale an, auf der beide Veränderungen gleich ausfallen. Der grau unterlegte Bereich gibt das Konfidenzintervall der Diagonalen an, das auf dem Linking-Fehler für Vergleiche zwischen den ursprünglich skalierten Ergebnissen von 2006 und den Ergebnissen von 2015 beruht (vgl. Tabelle A5.2).

Quelle: OECD, PISA-2015-Datenbank, Tabelle I.2.4a und A5.3.

StatLink <http://dx.doi.org/10.1787/888933433132>

Abbildung A5.2 ■ **Veränderungen bei den Leistungen in Lesekompetenz zwischen 2009 und 2015, basierend auf den ursprünglich bzw. neu skalierten Ergebnissen**



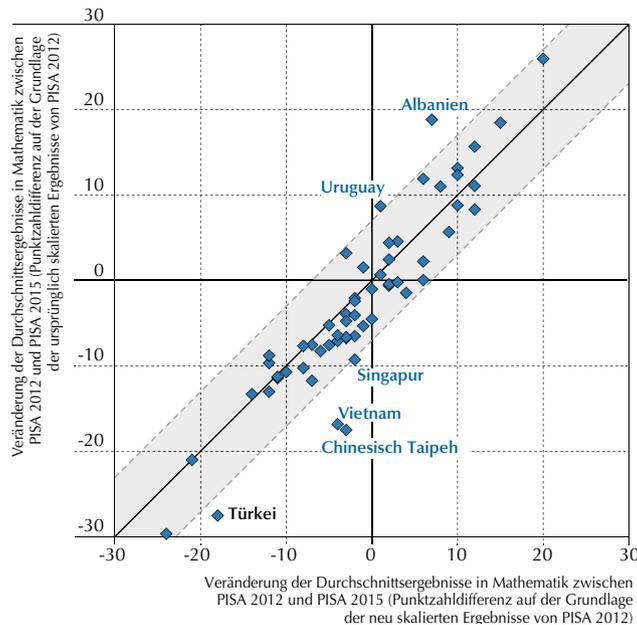
Anmerkung: Die durchgezogene Linie gibt die Diagonale an, auf der beide Veränderungen gleich ausfallen. Der grau unterlegte Bereich gibt das Konfidenzintervall der Diagonalen an, das auf dem Linking-Fehler für Vergleiche zwischen den ursprünglich skalierten Ergebnissen von 2009 und den Ergebnissen von 2015 beruht (vgl. Tabelle A5.2).

Quelle: OECD, PISA-2015-Datenbank, Tabelle I.4.4a und A5.4.

StatLink <http://dx.doi.org/10.1787/888933433149>



Abbildung A5.3 ■ **Veränderungen bei den Leistungen in Mathematik zwischen 2012 und 2015, basierend auf den ursprünglich bzw. neu skalierten Ergebnissen**



Anmerkung: Die durchgezogene Linie gibt die Diagonale an, auf der beide Veränderungen gleich ausfallen. Der grau unterlegte Bereich gibt das Konfidenzintervall der Diagonalen an, das auf dem Linking-Fehler für Vergleiche zwischen den ursprünglich skalierten Ergebnissen von 2012 und den Ergebnissen von 2015 beruht (vgl. Tabelle A5.2).

Quelle: OECD, PISA-2015-Datenbank, Tabelle I.5.4a und A5.5.

StatLink <http://dx.doi.org/10.1787/888933433156>

Vergleich von PISA-Ergebnissen zwischen papierbasiertem und computergestütztem Erhebungsmodus

Die – auf internationaler Ebene untersuchte – Äquivalenz von Link-Items wurde in der umfangreichen Moduseffektuntersuchung ermittelt, die Teil des Feldtests für PISA 2015 war. Diese Resultate untermauern die Annahme, dass die Ergebnisse für verschiedene Testmodi auf derselben Skala erfasst werden können. Darüber hinaus können mögliche länder- und erhebungsspezifische Parameter bis zu einem gewissen Grad nationale Abweichungen von der internationalen Norm begründen.

Die Äquivalenz der Link-Items wurde zuerst während des 2014 durchgeführten Feldtests an äquivalenten Populationen untersucht, die schulintern nach dem Zufallsprinzip ausgewählt wurden. Über 40 000 Schülerinnen und Schüler aus den Ländern und Volkswirtschaften, in denen eine computergestützte Durchführung von PISA 2015 geplant war, wurden in den einzelnen Schulen nach dem Zufallsprinzip für den computergestützten oder den papierbasierten Erhebungsmodus ausgewählt, so dass die Verteilung der Fähigkeiten der Schülerinnen und Schüler in beiden Modi vergleichbar war. Dadurch war es möglich, zwischen den verschiedenen Modi festgestellte Unterschiede in den Antwortmustern der Schülerinnen und Schüler, insbesondere Unterschiede, die über das aufgrund von Zufallsvariationen zu erwartende Maß hinausgingen, auf einen Effekt des Testmodus auf das Item anstatt auf die Fähigkeit der Schüler, den Testmodus zu nutzen, zurückzuführen. Der Feldtest war darauf ausgelegt, Moduseffekte auf internationaler Ebene, nicht aber für die einzelnen nationalen Stichproben oder Teilstichproben innerhalb eines Landes zu untersuchen.

Die Moduseffektuntersuchung ging zwei zentralen Fragestellungen nach:

- Messen die in früheren PISA-Zyklen zur Durchführung im papierbasierten Modus entwickelten Items dieselben Kompetenzen, wenn sie am Computer durchgeführt werden? Messen beispielsweise alle naturwissenschaftlichen Items, die für die Bearbeitung am Computer angepasst wurden, ausschließlich naturwissenschaftliche Kompetenzen, oder messen sie eine Kombination aus naturwissenschaftlichen Kompetenzen und Computerkenntnissen?
- Haben die papierbasierten Versionen dieser Items denselben Schwierigkeitsgrad wie die computergestützten Versionen?

Nur wenn ein Item in beiden Modi dieselben Kompetenzen erfasste und denselben Schwierigkeitsgrad hatte, wurde es als vollkommen äquivalentes (d.h. als skalar invariantes) Item betrachtet, das aussagekräftige Leistungsvergleiche verschiedener Erhebungsmodi ermöglicht. Diese Analyse der Testäquivalenz beruhte auf gepoolten Daten aus allen Ländern und Volkswirtschaften unter Verwendung von Erklärungsmodellen der Item-Response-Theorie (IRT). In diesen Modellen schätzen zwei unterschiedliche Parametersets, wie informativ die Antworten der Schülerinnen und Schüler im Hinblick auf ihre Kompetenz

auf der vorgesehenen Skala sind und welches Kompetenzniveau sie signalisieren. Bei der Analyse wurden drei Gruppen von Items identifiziert:

- Gruppe 1: Items, die in beiden Modi dieselben geschätzten Schwierigkeits- und Diskriminationsparameter hatten und daher als auf Papier und am Computer vollkommen äquivalent eingestuft wurden (*skalare Invarianz*).
- Gruppe 2: Items mit gleichem Diskriminationsparameter, aber unterschiedlichem Schwierigkeitsparameter (*metrische Invarianz*). Aus einer erfolgreichen Bearbeitung dieser Items ließen sich generell Rückschlüsse auf die Kompetenz in diesem Bereich ziehen; die Schwierigkeit der Items war jedoch je nach Modus unterschiedlich, oft aufgrund von Schnittstellenproblemen, wie z.B. Antwortformaten, die Freihandzeichnungen oder das Aufstellen von Gleichungen verlangten. Mehrere Items waren am Computer schwieriger, einige waren am Computer einfacher zu lösen.
- Gruppe 3: Items, die den Feldtestschätzungen zufolge je nach Modus unterschiedliche Kompetenzen prüften (keine *metrische Invarianz*).

Die Items in Gruppe 3 wurden im computergestützten Test in der Haupterhebung nicht verwendet (zwei Items in Mathematik wurden nur im papierbasierten Test eingesetzt). Die Items aus Gruppe 1 und 2 wurden verwendet, und die Stabilität der Item-Parameter für verschiedene Erhebungszyklen und -modi wurde während der Skalierungsverfahren für die Haupterhebung weiter geprüft. Letztlich stützten die Daten die vollkommene (skalare) Äquivalenz zwischen den Modi für bis zu 61 Items in Naturwissenschaften, 51 Items in Mathematik und 65 Items im Bereich Lesekompetenz⁷. Diese Items fungieren als Anker-Items oder Link-Items für Skalierungszwecke und bilden die Basis für Vergleiche der Leistungen in verschiedenen Testmodi und im Zeitverlauf. Für die übrigen Trend-Items, die Bestandteil der Haupterhebung von PISA 2015 waren (24 in Naturwissenschaften, 38 in Mathematik und 30 im Bereich Lesekompetenz), wurde metrische Äquivalenz festgestellt, jedes dieser Items erhielt jedoch einen modusspezifischen Schwierigkeitsparameter. Beim Vergleich von Schülerinnen und Schülern, die den PISA-Test in unterschiedlichen Modi absolvierten, liefert diese Gruppe von metrisch-invarianten Items nur Informationen über die Rangplätze der Schülerleistungen innerhalb eines bestimmten Modus (und trägt somit zur Messgenauigkeit bei), sie liefert jedoch keine Informationen für ein modusübergreifendes Ranking von Schülern und Ländern. Items mit skalarer Äquivalenz haben identische Item-Parameter für PBA (*paper-based assessment* – papierbasierte Erhebung) und CBA (*computer-based assessment* – computergestützte Erhebung) in Tabelle C2.1, C2.3 und C2.4; Items mit lediglich metrischer Äquivalenz haben dieselben Steigungsparameter, aber unterschiedliche Schwierigkeitsparameter.

Die anhand einer für alle PISA-Teilnehmer, die den Test am Computer absolvierten, repräsentativen Population ermittelten vollkommene Äquivalenz der Link-Items in den verschiedenen Erhebungsmodi stellt sicher, dass die Ergebnisse des papierbasierten und computergestützten Erhebungsmodus verglichen werden können und dass die Verknüpfung zwischen diesen Ergebniskategorien solide ist. Sie impliziert unter anderem, dass die Durchschnittsergebnisse aller Schülerinnen und Schüler, die PISA 2015 am Computer absolvierten, sowie die prozentualen Anteile der Schülerinnen und Schüler auf den verschiedenen Kompetenzstufen nicht signifikant unterschiedlich ausgefallen wären, wenn sie denselben Test auf Papier abgelegt hätten.

Anhang A6 liefert weitere Informationen zur exploratorischen Analyse von Modus-nach-Gruppen-Interaktionen, die an den Feldtestdaten durchgeführt wurde. Die Ergebnisse dieser Analyse sind zwar vielversprechend, insbesondere im Hinblick auf Modus-nach-Geschlecht-Interaktionen, bei der Interpretation der Ergebnisse müssen jedoch die Unzulänglichkeiten von Feldtestdaten für diese Zwecke bedacht werden.

Evaluierung der Vergleichbarkeit neuer naturwissenschaftlicher Items und Trend-Items

Für PISA 2015 wurden neue naturwissenschaftliche Items entwickelt, um den Veränderungen des PISA-Rahmenkonzepts für den Naturwissenschaftstest und dem geänderten Haupttestmodus Rechnung zu tragen. Änderungen des Rahmenkonzepts, die mit der Entwicklung neuer Items einhergehen, finden in PISA regelmäßig statt: das Rahmenkonzept für Lesekompetenz wurde 2009 überarbeitet, das Rahmenkonzept für Mathematik 2012. Die Entwicklung neuer Items in Naturwissenschaften wurde von der Notwendigkeit geleitet, eine ausgewogene Abdeckung aller Aspekte des Rahmenkonzepts sicherzustellen, insbesondere bei Aspekten, die im Rahmenkonzept von PISA 2015 gegenüber dem Rahmenkonzept von PISA 2006 weiterentwickelt worden waren oder einen größeren Stellenwert erhalten hatten. Dazu zählt die Unterscheidung zwischen epistemischem und prozeduralem Wissen, die im vorherigen Rahmenkonzept lediglich implizit war, sowie die aktivere Komponente der naturwissenschaftlichen Grundbildung. Letztere kommt in dem neuen Konzept der naturwissenschaftlichen Grundbildung zum Ausdruck, das auf den Kompetenzen *Naturwissenschaftliche Forschung bewerten und naturwissenschaftliche Untersuchungen planen* sowie *Daten und Evidenz naturwissenschaftlich interpretieren* (neben *Phänomene naturwissenschaftlich erklären*) aufbaut. Diese Kompetenzen ähneln zwar jenen Kompetenzen, die zuvor als *Naturwissenschaftliche Fragestellungen erkennen* und *Naturwissenschaftliche Beweise heranziehen* bezeichnet wurden, sind jedoch eindeutig nicht vollkommen deckungsgleich mit diesen.

Nach der Haupterhebung 2015 wurde die Möglichkeit, Ergebnisse auf der bestehenden, 2006 eingeführten Gesamtskala Naturwissenschaften darzustellen, durch eine Dimensionalitätsprüfung getestet. Wurden neue und bestehende Naturwissenschafts-Items als unterschiedlichen latenten Dimensionen angehörig behandelt, war die Median-Korrelation (aller Länder/Sprachgruppen) zwischen diesen Dimensionen mit 0,92 relativ hoch (ähnlich hoch wie die Korrelation zwischen Subskalen aus ein



und demselben Bereich). Modellfit-Statistiken bestätigten, dass ein eindimensionales Modell die Daten besser abbildet als ein zweidimensionales Modell, was die Schlussfolgerung stützt, dass neue und bestehende Naturwissenschafts-Items eine kohärente eindimensionale Skala mit guter Reliabilität bilden. Weitere Einzelheiten zu den Skalierungsergebnissen finden sich in *PISA 2015 Technical Report* (OECD, erscheint demnächst).

Quantifizierung der Unsicherheit der Skalenvergleichbarkeit im Linking-Fehler

Die Standardfehler für Schätzungen der Veränderung der Leistungen und Trends über verschiedene PISA-Zyklen berücksichtigen die Unsicherheit, die durch die Verknüpfung von unter getrennten Kalibrierungen erstellten Skalen zustande kommt. Diese konservativeren Standardfehler (die größer sind als die Standardfehler, die vor der Einführung des Linking-Fehlers geschätzt wurden) spiegeln nicht nur die Messpräzision und Stichprobenvarianz wie für die herkömmlichen PISA-Ergebnisse, sondern auch den in Tabelle A5.2 dargestellten Linking-Fehler wider. Für PISA 2015 trägt der Linking-Fehler nicht nur der Unsicherheit aufgrund der Auswahl der Link-Items, sondern auch der Unsicherheit aufgrund der 2015 eingeführten Veränderungen der Skalierungsmethode Rechnung.

Wie in früheren Zyklen kommt nur die Unsicherheit bezüglich der Position der Ergebnisse früherer PISA-Zyklen auf der Vergleichsskala von 2015 im Linking-Fehler zum Ausdruck. Da diese Unsicherheit über die Position in der Verteilung (eine Veränderung des Interzept) bei ortsinvarianten Schätzungen (wie z.B. Schätzungen für die Varianz, den Interquartilbereich, Geschlechterdifferenzen, Regressionskoeffizienten, Korrelationskoeffizienten usw.) aufgehoben wird, ist der Linking-Fehler in den Standardfehlern für diese Schätzungen nicht berücksichtigt.

Linking-Fehler beim Vergleich von Punktzahlen zwischen zwei PISA-Erhebungen

Die Linking-Fehler für PISA 2015 wurden auf Basis des Vergleichs der neu skalierten Durchschnittsergebnisse der einzelnen Länder und Volkswirtschaften für die verschiedenen Bereiche (wie in Tabelle A5.3, Tabelle A5.4 und Tabelle A5.5 aufgeführt) mit den entsprechenden Durchschnittsergebnissen geschätzt, die aus Public-Use-Files stammen und auf der ursprünglichen Skalierung des jeweiligen Erhebungszyklus beruhen. Dieser neue Ansatz für die Schätzung der Linking-Fehler wurde erstmals in PISA 2015 verwendet. Die Anzahl der Beobachtungen, die zur Berechnung der einzelnen Linking-Fehler verwendet wurden, entspricht der Anzahl der Länder mit Ergebnissen in beiden Erhebungszyklen. Aufgrund der spärlichen Daten, die der Berechnung des Linking-Fehlers zugrunde liegen, wurde eine robuste Schätzung der Standardabweichung auf Basis der S_n -Statistik verwendet (Rousseeuw und Croux, 1993).

Tabelle A5.2. Linking-Fehler bei Vergleichen zwischen PISA 2015 und früheren Erhebungen

Vergleich	Naturwissenschaften	Lesekompetenz	Mathematik
PISA 2000 mit 2015		6.8044	
PISA 2003 mit 2015		5.3907	5.6080
PISA 2006 mit 2015	4.4821	6.6064	3.5111
PISA 2009 mit 2015	4.5016	3.4301	3.7853
PISA 2012 mit 2015	3.9228	5.2535	3.5462

Anmerkung: Vergleiche zwischen den Ergebnissen in PISA 2015 und früheren Erhebungen sind erst ab dem Erhebungszyklus möglich, bei dem der entsprechende Bereich erstmals Schwerpunktbereich war. Folglich kann beispielsweise kein Vergleich der Leistungen in Naturwissenschaften zwischen PISA 2000 und PISA 2015 angestellt werden.

StatLink  <http://dx.doi.org/10.1787/888933433162>

Linking-Fehler bei anderen Vergleichsformen von Schülerleistungen

Der Linking-Fehler für regressionsbasierte Leistungstrends und für Vergleiche auf Basis nichtlinearer Transformationen von Skalenwerten kann durch Simulation auf Basis des Linking-Fehlers für den Vergleich von Punktzahlen zwischen zwei PISA-Erhebungen geschätzt werden. Im Einzelnen sind in Tabelle A5.6 die Schätzungen des Linking-Fehlers für den Vergleich der prozentualen Anteile der Schülerinnen und Schüler mit Leistungen unter Stufe 2 bzw. auf oder über Stufe 5 dargestellt, während Tabelle A5.7 die Größenordnung des Linking-Fehlers angibt, der mit der Schätzung des durchschnittlichen Dreijahrestrends assoziiert ist.

Die Schätzung der Linking-Fehler für den Prozentsatz der Schülerinnen und Schüler, die unter Stufe 2 und auf oder über Stufe 5 abschneiden, geht von der Annahme aus, dass die Größenordnung der mit dem Linking von Skalen verbundenen Unsicherheit einer Normalverteilung folgt, mit einem Mittelwert von 0 und einer Standardabweichung, die dem in Tabelle A5.2 aufgeführten Linking-Fehler entspricht. Aus dieser Verteilung werden 500 Fehler gezogen und zum ersten plausiblen Wert der Schülerinnen und Schüler der einzelnen Länder und Volkswirtschaften in PISA 2015 hinzugefügt, um die 500 möglichen Szenarien darzustellen, in denen die einzige Ursache für Differenzen in Bezug auf 2015 die Unsicherheit in der Verlinkung ist.

Durch Berechnung des gesuchten Schätzwerts (z.B. prozentualer Anteil der Schülerinnen und Schüler auf einer bestimmten Kompetenzstufe) für jede der 500 Replikationen lässt sich ermesen, wie der Linking-Fehler diese Schätzung beeinflusst. Die Standardabweichung der 500 Replikationsschätzungen wird als der Linking-Fehler für die Veränderung des prozentualen Anteils der Schülerinnen und Schüler mit Leistungen auf einer bestimmten Kompetenzstufe verwendet. Da der Einfluss des Linking-Fehlers auf diese Schätzung von der genauen Form und Dichte der Leistungsverteilung um die Schwellenwerte herum abhängt, sind die Linking-Fehler für Vergleiche der Kompetenzstufen für jedes Land sowie innerhalb der einzelnen Länder für Jungen und Mädchen unterschiedlich.

In ähnlicher Weise beruht die Schätzung der Linking-Fehler für regressionsbasierte Trends auf der Annahme, dass die Unsicherheit der Verlinkung einer Normalverteilung folgt, mit einem Mittelwert von 0 und einer Standardabweichung, die dem in Tabelle A5.2 ausgewiesenen Linking-Fehler entspricht. Da hier jedoch Trends über mehr als zwei Erhebungsjahre untersucht werden sollen, muss zusätzlich zu den in Tabelle A5.2 aufgeführten Linking-Fehlern auch die Kovarianz zwischen Linking-Fehlern berücksichtigt werden. Um Daten aus mehreren PISA-Erhebungen zu simulieren, wurden 2 000 Beobachtungen aus einer multivariaten Normalverteilung gezogen, bei der alle Mittelwerte gleich 0 sind und deren Varianz-Kovarianz-Struktur durch den in Tabelle A5.2 ausgewiesenen Linking-Fehler sowie die in Tabelle 12.31 des *PISA 2012 Technical Report* (OECD, 2014a) veröffentlichten Linking-Fehler zwischen früheren PISA-Skalen bestimmt wird. Diese Beobachtungen repräsentieren 2 000 mögliche Szenarien, in denen der reale Trend 0 ist und der geschätzte Trend ausschließlich die Unsicherheit bei der Vergleichbarkeit der Leistungen zwischen verschiedenen Skalen widerspiegelt. Die Linking-Fehler für Vergleiche des durchschnittlichen Dreijahrestrends zwischen PISA 2015 und früheren Erhebungen hängen von der Anzahl der in der Schätzung berücksichtigten Erhebungszyklen ab, sind jedoch unabhängig von der Form der Leistungsverteilung innerhalb der einzelnen Länder.

Leistungsvergleiche: Unterschiede zwischen zwei Erhebungen und durchschnittlicher Dreijahrestrend

Um die Leistungsentwicklung zu evaluieren, werden in den Analysen die Leistungsveränderungen zwischen zwei Erhebungszyklen und der durchschnittliche Dreijahrestrend der Leistungen erfasst. Für den Bereich Lesekompetenz, bei dem bis zu sechs Datenpunkte verfügbar sind, werden auch kurvilineare Trendverläufe geschätzt.

Vergleiche zwischen zwei Erhebungen (z.B. die Leistungsveränderungen eines Landes bzw. einer Volkswirtschaft zwischen PISA 2006 und PISA 2015 oder die Leistungsveränderungen einer Untergruppe) werden wie folgt berechnet:

$$\Delta_{2015-t} = PISA_{2015} - PISA_t$$

wobei Δ_{2015-t} für den Leistungsunterschied zwischen PISA 2015 und einer früheren PISA-Erhebung steht (Vergleiche sind erst ab dem Erhebungszyklus möglich, bei dem der entsprechende Bereich erstmals Schwerpunktbereich war; folglich ist ein Vergleich der Mathematikleistungen zwischen PISA 2015 und PISA 2000 oder ein Vergleich der Leistungen in Naturwissenschaften zwischen PISA 2015 und PISA 2000 oder PISA 2003 nicht möglich.) $PISA_{2015}$ steht für die in PISA 2015 in Mathematik, Lesekompetenz oder Naturwissenschaften erzielte Punktzahl und $PISA_t$ für die in einer früheren Erhebung in Mathematik, Lesekompetenz oder Naturwissenschaften verzeichnete Punktzahl. Der Standardfehler der Leistungsveränderung $\sigma(\Delta_{2015-t})$ entspricht:

$$\sigma(\Delta_{2015-t}) = \sqrt{\sigma_{2015}^2 + \sigma_t^2 + error_{2015,t}^2}$$

wobei σ_{2015} für den bei $PISA_{2015}$ beobachteten Standardfehler, σ_t für den bei $PISA_t$ beobachteten Standardfehler und $error_{2015,t}$ für den Linking-Fehler bei Vergleichen der Ergebnisse in Naturwissenschaften, Lesekompetenz oder Mathematik zwischen der PISA-Erhebung 2015 und einer früheren Erhebung (t) steht. Der Wert für $error_{2015,t}$ ist für die meisten Vergleiche in Tabelle A5.2 und für Vergleiche der Kompetenzstufen in Tabelle A5.6 aufgeführt.

Ein zweiter in PISA wiedergegebener Analyse katalog bezieht sich auf den durchschnittlichen Dreijahrestrend der Leistungen. Der durchschnittliche Dreijahrestrend ist die durchschnittliche Veränderungsrate, die im Verlauf der PISA-Teilnahme eines Landes bzw. einer Volkswirtschaft je Dreijahreszeitraum beobachtet wird – ein Zeitintervall, das dem üblichen Intervall zwischen zwei aufeinanderfolgenden PISA-Erhebungen entspricht. Folglich signalisiert ein positiver durchschnittlicher Dreijahrestrend von x Punkten, dass das Land bzw. die Volkswirtschaft seine/ihre Leistungen seit seinen/ihren ersten vergleichbaren PISA-Ergebnissen um x Punkte je Dreijahreszeitraum verbessert hat. Bei Ländern und Volkswirtschaften, die nur an PISA 2012 und PISA 2015 teilgenommen haben, entspricht der durchschnittliche Dreijahrestrend der Differenz zwischen den beiden Erhebungen⁸.

Die Berechnung des durchschnittlichen Dreijahrestrends der Leistungen erfolgt durch eine Regression der Form

$$PISA_{i,t} = \beta_0 + \beta_1 time_t + \varepsilon_{i,t}$$

wobei $PISA_{i,t}$ die Position von Land i auf der Gesamtskala Naturwissenschaften, Lesekompetenz oder Mathematik in Jahr t (mittlere Punktzahl oder Perzentil der Punktzahlverteilung), $time_t$ eine Variable der Zeitmessung in Dreijahreseinheiten und $\varepsilon_{i,t}$ einen Fehlerterm darstellt, der die Stichproben- und Messunsicherheit für $PISA_{i,t}$ angibt. In der Schätzung wird unterstellt, dass Stichprobenfehler und Messfehler im Zeitverlauf unabhängig sind. Bei dieser Spezifikation gibt die Schätzung für β_1 die



durchschnittliche Veränderungsrate je Dreijahreszeitraum an. Ebenso wie bei Vergleichen zwischen zwei PISA-Erhebungen ein Linking-Fehler hinzugerechnet wird, enthalten auch die Standardfehler für β_1 einen Linking-Fehler:

$$\sigma(\beta_1) = \sqrt{\sigma_{s,i}^2(\beta_1) + \sigma_t^2(\beta_1)}$$

wobei $\sigma_{s,i}(\beta_1)$ für den Stichproben- und Imputationsfehler steht, der mit der Schätzung von β_1 assoziiert ist, und $\sigma_t^2(\beta_1)$ für den Linking-Fehler steht, der mit dem durchschnittlichen Dreijahrestrend assoziiert ist. Dieser ist in Tabelle A5.7 dargestellt.

Der durchschnittliche Dreijahrestrend ist eine robustere Messgröße der Fortschritte, die ein Land bzw. eine Volkswirtschaft bei den Bildungsergebnissen erzielt hat, da er auf Informationen aus allen Erhebungen basiert. Er reagiert dementsprechend weniger empfindlich auf anormale Messungen, die bei Vergleichen, die lediglich auf zwei Erhebungen beruhen, Veränderungen verursachen können. Berechnet wird der durchschnittliche Dreijahrestrend als die Linie, die den Verlauf der PISA-Teilnahme eines Landes bzw. einer Volkswirtschaft am besten abbildet. Die PISA-Ergebnisse werden auf das Jahr regressiert, in dem das Land an PISA teilnahm (gemessen in Dreijahres-Zeiteinheiten). Der durchschnittliche Dreijahrestrend trägt auch der Tatsache Rechnung, dass der Zeitraum zwischen zwei PISA-Erhebungen in einigen Ländern und Volkswirtschaften weniger als drei Jahre beträgt. Dies ist bei jenen Ländern und Volkswirtschaften der Fall, die im Rahmen von PISA+ an PISA 2000 oder PISA 2009 teilnahmen: Sie führten die Erhebung in den Jahren 2001, 2002 oder 2010 anstatt 2000 oder 2009 durch.

Die kurvilinearen Trends bei der Lesekompetenz werden in ähnlicher Weise geschätzt, indem eine quadratische Regressionsfunktion auf die PISA-Ergebnisse für Land i bei allen mit t indexierten Erhebungen angewendet wird:

$$PISA_{i,t} = \beta_2 + \beta_3 year_t + \beta_4 year_t^2 + \varepsilon_{i,t}$$

wobei $year_t$ eine Variable ist, die die Zeit in Jahren seit 2015 misst, und $year_t^2$ das Quadrat von $year_t$ ist. Da $year$ so skaliert ist, dass es 2015 gleich 0 ist, gibt β_3 die geschätzte jährliche Veränderungsrate im Jahr 2015 an und β_4 die Beschleunigung/Verlangsamung des Trends. Ist β_4 positiv, deutet dies darauf hin, dass der beobachtete Trend U-förmig ist, und die in den näher bei 2012 liegenden Jahren bei den Leistungen beobachteten Veränderungsrate höher (positiver) sind als die in früheren Jahren beobachteten. Bei negativem β_4 verläuft der beobachtete Trend umgekehrt U-förmig, und die in den näher bei 2012 liegenden Jahren bei den Leistungen beobachteten Veränderungsrate sind niedriger (negativer) als die in früheren Jahren beobachteten. Ebenso wie in der Schätzung der Standardfehler für den durchschnittlichen Dreijahrestrend ein Linking-Fehler hinzugerechnet wird, enthalten auch die Standardfehler für β_3 und β_4 einen Linking-Fehler (Tabelle A5.8). Kurvilineare Trends werden nur für den Bereich Lesekompetenz und nur für Länder und Volkswirtschaften geschätzt, bei denen ein Leistungsvergleich über mindestens fünf Erhebungen möglich ist, um eine Überanpassung an die Daten zu vermeiden.

Bereinigte Trends

PISA behält seine technischen Standards im Zeitverlauf bei. Das bedeutet zwar, dass Trends für konsistent definierte Populationen berechnet werden können, der dadurch repräsentierte Anteil der Population von 15-Jährigen und/oder die demografischen Merkmale 15-jähriger Schülerinnen und Schüler können sich aber auch verändern, beispielsweise aufgrund von Migrationsbewegungen.

Da Trendanalysen das Fortschrittstempo sukzessiver Schülerkohorten veranschaulichen, ist es wichtig zu untersuchen, inwieweit diese Ergebnisse auf Veränderungen beim Erfassungsgrad der Stichprobe und bei den demografischen Merkmalen der in der Stichprobe enthaltenen Schülerinnen und Schüler zurückzuführen sind, um zuverlässige Schlüsse aus den Ergebnissen ziehen zu können. Dementsprechend wurden drei Kategorien von Trendergebnissen entwickelt: unbereinigte Trends, um Veränderungen der Schulbesuchsquoten bereinigte Trends und um Veränderungen der demografischen Merkmale der Stichprobe bereinigte Trends. Bei den bereinigten Trends handelt es sich um Leistungstrends, die nach Neutralisierung des Effekts gleichzeitiger Veränderungen der demografischen Merkmale der Stichprobe geschätzt wurden.

Um Veränderungen der Schulbesuchsquoten bereinigte Trends

Um den Effekt von Veränderungen der Schulbesuchsquoten (genauer gesagt des Erfassungsgrads der PISA-Stichprobe bezogen auf die Gesamtpopulation der 15-Jährigen: vgl. Erfassungsindex 3 in Anhang A2) zu neutralisieren, wurde unterstellt, dass die Leistungen aller nicht von der Erhebung erfassten 15-Jährigen unter dem Medianeiveau für alle 15-Jährigen liegen würden. Auf Basis dieser Annahme war es möglich, den Medianwert für alle 15-Jährigen (für Länder, in denen der Erfassungsgrad der Stichprobe mindestens 50% betrug) und höhere Perzentile ohne Angaben zum Leistungsniveau der nicht erfassten 15-Jährigen zu berechnen.

In der Praxis ist es zur Schätzung um Veränderungen der Schulbesuchsquoten bereinigter Trends zunächst erforderlich, je Land bzw. Volkswirtschaft einen einzigen Fall zur Datenbank hinzuzufügen, der alle nicht von der PISA-Stichprobe erfassten 15-Jährigen repräsentiert. Die endgültige Schülergewichtung für diesen Fall wird berechnet als die Differenz zwischen der Gesamtpopulation der 15-Jährigen (vgl. Tabelle I.6.1 und Anhang A2) und der Summe der endgültigen Schülergewichtungen für die in der Stichprobe berücksichtigten Beobachtungen (die gewichtete Anzahl der teilnehmenden Schülerinnen und Schüler).

Analog dazu werden die einzelnen Replikationsgewichtungen für diesen Fall als Differenz zwischen der Gesamtpopulation der 15-Jährigen und der Summe der entsprechenden Replikationsgewichtungen berechnet. Etwaige aus diesem Verfahren resultierende negative Gewichtungen werden durch 0 ersetzt. Ein Wert unter einem der plausiblen Werte in der PISA-Stichprobe wird für die Leistungsvariablen dieses Falls eingesetzt.

In einem zweiten Schritt werden der Median und die oberen Perzentile der Verteilung für die erweiterte Stichprobe berechnet. In einigen Fällen, in denen der Erfassungsgrad unter 50% liegt, wird die Schätzung für den bereinigten Median als fehlend ausgewiesen.

Um Veränderungen der demografischen Merkmale der Stichprobe bereinigte Trends

Es wird ein der Poststratifikation entsprechendes Neugewichtungsverfahren eingesetzt, um die Stichprobenmerkmale früherer Stichproben an die beobachtete Zusammensetzung der PISA-2015-Stichprobe anzupassen.

In einem ersten Schritt wird die in jedem Erhebungszyklus berücksichtigte Stichprobe in diskrete Zellen unterteilt, die durch Migrationsstatus (vier Kategorien: ohne Migrationshintergrund, erste Generation, zweite Generation, fehlend), Geschlecht (zwei Kategorien: Junge, Mädchen) und relatives Alter (vier Kategorien für vier Dreimonatszeiträume) der Schülerinnen und Schüler definiert sind. Die wenigen in früheren PISA-Datensätzen enthaltenen Beobachtungen mit fehlenden Geschlechts- oder Altersangaben werden gelöscht. Auf diese Weise werden maximal 32 diskrete Zellen für die gesamte Population definiert. In allen Fällen aber, in denen die Anzahl der Beobachtungen in einer dieser 32 Zellen für ein bestimmtes Land bzw. eine bestimmte Volkswirtschaft und eine bestimmte PISA-Erhebung weniger als 10 beträgt, wird die entsprechende Zelle nach einem sequenziellen Algorithmus mit einer anderen, ähnlichen Zelle zusammengefasst, bis alle Zellen eine Mindeststichprobengröße von 10 haben⁹.

In einem zweiten Schritt werden die Zellen neugewichtet, so dass die Summe der endgültigen Schülergewichtungen in den einzelnen Zellen zwischen verschiedenen Erhebungen konstant ist und der Summe der endgültigen Schülergewichtungen in der PISA-2015-Stichprobe entspricht. Im Anschluss daran werden für diese neugewichteten Stichproben Schätzungen des Durchschnitts und der Verteilung der Schülerleistungen durchgeführt, die die (kontrafaktischen) Leistungen repräsentieren, die beobachtet worden wären, wenn die Stichproben früherer Jahre im Hinblick auf die bei dieser Neugewichtung verwendeten Variablen dieselbe Zusammensetzung gehabt hätten wie die Stichprobe in PISA 2015.

Tabelle A5.9 weist für jedes Land bzw. jede Volkswirtschaft die Anzahl der für die Poststratifikation verwendeten Zellen sowie für jeden Erhebungszyklus die Anzahl der Beobachtungen aus, die aus den um Veränderungen der demografischen Merkmale der Stichprobe bereinigten Trends ausgeklammert wurden.

Tabelle A5.10 weist für jedes Land bzw. jede Volkswirtschaft die Durchschnittswerte der für die Anpassung verwendeten Hintergrundvariablen aus.

Vergleich von Items und nicht leistungsbezogenen Skalen zwischen verschiedenen PISA-Erhebungsrunden

Um Informationen über die Merkmale der Schülerinnen und Schüler und der Schulen zu sammeln, werden sowohl die Schülerinnen und Schüler als auch die Schulleitungen in PISA gebeten, einen Hintergrundfragebogen auszufüllen. Da zwischen PISA 2006 und PISA 2015 mehrere Fragen gleich blieben, ist ein Vergleich der Antworten auf diese Fragen im Zeitverlauf möglich. Sofern nicht anders angegeben, wurden Fragen mit geringfügigen oder größeren Änderungen im Wortlaut im Zeitverlauf nicht verglichen, weil unmöglich festzustellen ist, ob beobachtete Änderungen in einer Antwort auf Änderungen im gemessenen Konstrukt oder auf Änderungen in der Art, wie das Konstrukt gemessen wird, zurückzuführen sind.

Darüber hinaus werden, wie in Anhang A1 beschrieben, Fragebogen-Items in PISA für die Konstruktion von Indizes verwendet. In PISA werden zwei Arten von Indizes verwendet: einfache Indizes und Skalenindizes.

Bei einfachen Indizes wird eine Reihe von Antworten auf Fragebogen-Items umkodiert. Für Trendanalysen werden die in PISA 2006 beobachteten Werte ebenso wie einfache Antworten auf Fragebogen-Items direkt mit PISA 2015 verglichen. Dies gilt für Indizes wie die Schüler/Lehrer-Quote oder den Migrationsstatus.

Skalenindizes hingegen sind als Warm-Likelihood-Schätzungen (Warm Likelihood Estimates – WLE; Warm, 1989) in der Datenbank berücksichtigt und basieren auf einem generalisierten Partial-Credit-Modell (GPCM; vgl. Muraki, 1992). In allen Fällen, in denen zumindest ein Teil der bei der Konstruktion der Indizes verwendeten Fragen in PISA 2006 und PISA 2015 intakt bleibt, basiert die Skalierung des entsprechenden Index auf einer gleichzeitigen Kalibrierung mit PISA-2006- und PISA-2015-Daten, gefolgt von einer linearen Transformation zur Abbildung der resultierenden Skala auf der ursprünglichen PISA-2006-Skala für den Index, die auf einem Partial-Credit-Modell (PCM; vgl. OECD, 2009) beruht. Durch dieses Verfahren, das dem für kognitive Skalen verwendeten Verfahren entspricht, wird sichergestellt, dass die korrespondierenden Indexwerte verglichen werden können.

Um Veränderungen bei diesen Items und Skalen zu evaluieren, werden Schätzwertänderungen zwischen zwei Erhebungen, in der Regel PISA 2006 und PISA 2015, in den Analysen erfasst. Vergleiche zwischen zwei Erhebungen (z.B. die Veränderungen



eines Landes/einer Volkswirtschaft beim Index der Freude am naturwissenschaftlichen Lernen zwischen PISA 2006 und PISA 2015 oder die Veränderungen bei diesem Index für eine Untergruppe) werden berechnet als:

$$\Delta_{2015,2006} = PISA_{2015} - PISA_{2006}$$

wobei $\Delta_{2015,t}$ für die Differenz bei diesem Index zwischen PISA 2015 und einer früheren Erhebung, $PISA_{2015}$ für den in PISA 2015 beobachteten Indexwert und $PISA_{2006}$ für den 2006 verzeichneten Indexwert steht. Der Standardfehler der Veränderung des Indexwerts $\sigma(\Delta_{2015-2006})$ ist:

$$\sigma(\Delta_{2015-2006}) = \sqrt{\sigma_{2015}^2 + \sigma_{2006}^2}$$

wobei σ_{2015} für den bei $PISA_{2015}$ und σ_{2006} für den bei $PISA_{2006}$ beobachteten Standardfehler steht. Standardfehler für Veränderungen der Indexwerte berücksichtigen keine Messunsicherheit oder Unsicherheit aufgrund des Equating-Verfahrens und sind folglich etwas unterzeichnet. Standardfehler für Veränderungen der Antworten auf einzelne Items unterliegen keiner Messunsicherheit oder Equating-bedingten Unsicherheit.

OECD-Durchschnitt

Der OECD-Durchschnitt wird im gesamten Bericht als Referenzgröße verwendet. Berechnet wird er als der Durchschnitt der OECD-Länder, in dem alle Länder gleich gewichtet sind. Einige OECD-Länder haben an bestimmten Erhebungen nicht teilgenommen, andere OECD-Länder verfügen für einige Erhebungen nicht über vergleichbare Ergebnisse, wieder andere haben bestimmte Fragen nicht in ihre Fragebogen aufgenommen oder sie von Erhebung zu Erhebung bedeutend geändert. In Tabellen und Abbildungen, die Trendentwicklungen darstellen, wird der OECD-Durchschnitt anhand einer konsistenten Auswahl von OECD-Ländern ausgewiesen. Beispielsweise umfasst der „OECD33-Durchschnitt“ nur 33 OECD-Länder mit nicht fehlenden Beobachtungen für die Erhebungen, für die dieser Durchschnitt selbst keinen Fehlwert aufweist. Diese Einschränkung ermöglicht valide Vergleiche des OECD-Durchschnitts im Zeitverlauf.

Online verfügbare Tabellen (nicht auf Deutsch verfügbar) <http://dx.doi.org/10.1787/888933433162>

Tabelle A5.3.	Mean scores in science since 2006 produced with the 2015 approach to scaling
Tabelle A5.4.	Mean scores in reading since 2006 produced with the 2015 approach to scaling
Tabelle A5.5.	Mean scores in mathematics since 2006 produced with the 2015 approach to scaling
Tabelle A5.6.	Link error for comparisons of proficiency levels between PISA 2015 and previous assessments
Tabelle A5.7.	Link error for comparisons of the average three-year change between PISA 2015 and previous assessments
Tabelle A5.8.	Link error for the curvilinear trend between PISA 2015 and previous assessments
Tabelle A5.9.	Cells used to adjust science, reading and mathematics scores to the PISA 2015 samples
Tabelle A5.10.	Descriptive statistics for variables used to adjust science, reading and mathematics scores to the PISA 2015 samples

Anmerkungen

1. Vgl. auch Carstensen (2013) wegen des Einflusses der Testkonzeption auf die Trendmessung.
2. Die eingeschränkte Berücksichtigung differenzieller Itemfunktionen in den früheren Erhebungsrounden wurde zusammen mit der erhebungsroundenspezifischen Kalibrierungsstichprobe dafür kritisiert, dass sie zu Trendschätzungen führt, die nicht mit den nationalen Kalibrierungen im Einklang stehen, bei denen entsprechende Stichproben verwendet werden (Urbach, 2013).
3. Die Zahl der nicht erreichten Items wird in PISA 2015 für die Generierung plausibler Werte (plausible values) als Quelle für Hintergrundinformationen verwendet, so dass die Korrelation zwischen den nicht erreichten Items und den Leistungen in den Ergebnissen modelliert und berücksichtigt ist.
4. Das Modell erlaubt, dass einige Länder und Volkswirtschaften Daten für weniger als vier Erhebungsjahre liefern.
5. Die Korrelation der Mittelwerte von PISA 2009 und PISA 2012 beträgt für die Länder und Volkswirtschaften, die an PISA 2015 teilnahmen, in Naturwissenschaften 0,985 (wobei beide Erhebungen auf Jahre fielen, in denen Naturwissenschaften einen untergeordneten Bereich bildeten, so dass exakt dieselben Aufgaben verwendet wurden), in Lesekompetenz 0,972 (wobei in PISA 2012 nur ein Teil der Aufgaben von PISA 2009 verwendet wurde) und in Mathematik 0,981 (wobei PISA 2012 zeitlich mit einer Revision des Rahmenkonzepts und einer Ausweitung des Katalogs an Testaufgaben zusammenfiel). PISA 2009 und PISA 2012 sind die beiden Erhebungen, bei denen die Testkonzeption und der Skalierungsansatz am ähnlichsten ausfielen. Die Korrelation der Mittelwerte in Lesekompetenz von PISA 2000 und PISA 2009 beträgt (für die Länder und Volkswirtschaften, die an PISA 2015 teilnahmen) 0,955, die Korrelation der Mittelwerte in Mathematik von PISA 2003 und PISA 2012 beträgt 0,953 und die Korrelation der Mittelwerte in Naturwissenschaften von PISA 2006 und PISA 2015 beträgt 0,947 (0,944 auf der Grundlage der Ergebnisse in Tabelle A5.3, die nach einem einheitlichen Skalierungsansatz berechnet wurden).
6. Die im Rahmen der Skalierung berechneten Ländermittelwerte sind diejenigen, die beobachtet worden wären, wenn nur Schülerinnen und Schüler berücksichtigt worden wären, für die Antwortdaten für die entsprechenden Bereiche vorliegen. Da im Rahmen von PISA jedoch Daten für alle Schülerinnen und Schüler in allen Bereichen berechnet werden, die in einem Land bzw. einer Volkswirtschaft untersucht werden, unabhängig davon, ob ein Schüler auch tatsächlich ein Testheft erhalten hat, das Einheiten für einen bestimmten Bereich enthält, können sich die im Zuge der Skalierung berechneten modellbasierten Mittelwerte von den in Anhang B angegebenen Mittelwerten unterscheiden. Der Effekt der berechneten Punktzahlen auf die Mittelwerte ist jedoch zu vernachlässigen, wie der Vergleich der Ergebnisse für 2015 zwischen den auf dem Skalierungsmodell beruhenden in Tabelle A5.3, A5.4 und A5.5 angegebenen Schätzungen und den auf dem Gesamtpopulationsmodell beruhenden in Tabelle I.2.3, I.4.3 und I.5.3 angegebenen Schätzungen erkennen lässt.
7. Bei der Untersuchung der Ergebnisse für ein bestimmtes Land oder eine bestimmte Volkswirtschaft sind diese Zahlen wegen der Möglichkeit länderspezifischer Abweichungen von der internationalen Norm als Obergrenze für die tatsächliche Anzahl der skalar invarianten Items zu betrachten.
8. Der durchschnittliche Dreijahrestrend hängt mit der „annualisierten Veränderung“ zusammen, auf die in früheren PISA-Berichten Bezug genommen wurde (OECD, 2014b). Durch Multiplikation der annualisierten Veränderung mit dem Faktor 3 ergibt sich der durchschnittliche Dreijahrestrend.
9. Die Stichproben werden immer zuerst nach dem Migrationsstatus unterteilt (es sei denn, dies würde zu Gruppen mit weniger als 10 Beobachtungen führen), danach innerhalb der nach Migrationsstatus definierten Gruppen nach Geschlecht (es sei denn, dies würde zu Gruppen mit weniger als 10 Beobachtungen führen) und zuletzt nach Altersgruppen. In jeder dieser Etappen werden Zellen wie nachstehend beschrieben zusammengefasst, wenn es Gruppen mit weniger als 10 Beobachtungen gibt; dabei endet die Sequenz der Zellfusionen in den einzelnen Etappen jeweils dann, wenn alle Gruppen eine Mindestgröße von 10 erreichen. 1. Etappe (Migrationsstatus, innerhalb zuvor definierter Sprachgruppen): Zusammenfassung der Kategorien „fehlend“ und „ohne Migrationshintergrund“; Zusammenfassung der Kategorien „erste Generation“ und „zweite Generation“; Zusammenfassung aller Kategorien. 2. Etappe (Geschlecht, innerhalb der zuvor definierten Migrationsstatuskategorien): Zusammenfassung von Jungen und Mädchen. 3. Etappe (Alter, innerhalb der zuvor definierten Migrations-/Geschlechterkategorien): Zusammenfassung des ersten und zweiten Quartals; Zusammenfassung des dritten und vierten Quartals; Zusammenfassung aller Kategorien.



Literaturverzeichnis

- Birnbaum, A.** (1968), *On the Estimation of Mental Ability*, Series Report 15, USAF School of Aviation Medicine, Randolph Air Force Base (TX).
- Carstensen, C.H.** (2013), "Linking PISA Competencies over Three Cycles – Results from Germany", S. 199-213 in M. Prenzel et al. (Hrsg.), *Research on PISA*, Springer, Niederlande, http://dx.doi.org/10.1007/978-94-007-4458-5_12.
- Davidov, E., P. Schmidt und J. Billiet** (Hrsg.) (2011), *Cross-Cultural Analysis: Methods and Applications*. Routledge, New York.
- Glas, C. und K. Jehangir** (2014), "Modeling Country Specific Differential Item Functioning", in D. Rutkowski (Hrsg.), *Handbook of International Large-Scale Assessment*, CRC Press, Boca Raton (FL).
- Masters, G.N.** (1982), "A Rasch Model for Partial Credit Scoring." *Psychometrika*, Vol.47/2, S. 149-174, <http://dx.doi.org/10.1007/BF02296272>.
- Meredith, W.** (1993), "Measurement invariance, factor analysis and factorial invariance", *Psychometrika*, Vol. 58/4, S. 525-543, <http://dx.doi.org/10.1007/BF02294825>.
- Muraki, E.** (1992), "A Generalized Partial Credit Model: Application of an EM Algorithm" *Applied Psychological Measurement*, Vol. 16/2, S. 159-176, <http://dx.doi.org/10.1177/014662169201600206>.
- OECD** (erscheint demnächst), *PISA 2015 Technical Report*, PISA, OECD Publishing, Paris.
- OECD** (2014a), *PISA 2012 Ergebnisse: Was Schülerinnen und Schüler wissen und können (Band I, Überarbeitete Ausgabe, Februar 2014)*, W. Bertelsmann Verlag, Bielefeld, <http://dx.doi.org/10.1787/9789264208858-de>.
- OECD** (2014b), *PISA 2012 Technical Report*, OECD Publishing, Paris, <http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>.
- OECD** (2009), *PISA 2006 Technical Report*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264048096-en>.
- Oliveri, M.E. und M. von Davier** (2014), "Toward Increasing Fairness in Score Scale Calibrations Employed in International Large-Scale Assessments" *International Journal of Testing*, Vol. 14/1, S. 1-21, <http://dx.doi.org/10.1080/15305058.2013.825265>.
- Oliveri, M.E. und M. von Davier** (2011), "Investigation of Model Fit and Score Scale Comparability in International Assessments" *Psychological Test and Assessment Modeling*, Vol. 53/3, S. 315-333.
- Rasch, G** (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, Nielsen & Lydiche, Kopenhagen.
- Rousseuw, P.J. und C. Croux** (1993), "Alternatives to the Median Absolute Deviation", *Journal of the American Statistical Association*, Vol. 88/424, S. 1273-1283, <http://dx.doi.org/10.1080/01621459.1993.10476408>.
- Urbach, D.** (2013), "An Investigation of Australian OECD PISA Trend Results", in M. Prenzel et al. (Hrsg.), *Research on PISA*, S. 165-179, Springer Niederlande, http://dx.doi.org/10.1007/978-94-007-4458-5_10.
- Warm, T.A.** (1989), "Weighted likelihood estimation of ability in item response theory", *Psychometrika*, Vol. 54/3, S. 427-450, <http://dx.doi.org/10.1007/BF02294627>.

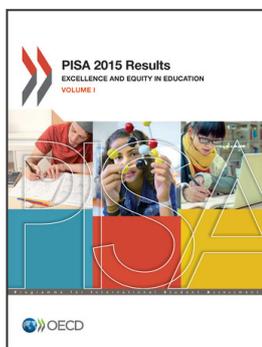


ANHANG A6

THE PISA 2015 FIELD TRIAL MODE-EFFECT STUDY

Nicht auf Deutsch verfügbar

Englische Fassung unter: www.oecd.org/pisa



From:
PISA 2015 Results (Volume I)
Excellence and Equity in Education

Access the complete publication at:
<https://doi.org/10.1787/9789264266490-en>

Please cite this chapter as:

OECD (2016), "PISA 2015 - Technische Hinweise", in *PISA 2015 Results (Volume I): Excellence and Equity in Education*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264267879-13-de>

Das vorliegende Dokument wird unter der Verantwortung des Generalsekretärs der OECD veröffentlicht. Die darin zum Ausdruck gebrachten Meinungen und Argumente spiegeln nicht zwangsläufig die offizielle Einstellung der OECD-Mitgliedstaaten wider.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.