

Heike Wendt
Benjamin Euen
Wilfried Bos



TIMSS 2007 Nutzermanual

Dokumentation zur Arbeit mit den Datensätzen von TIMSS 2007

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen PLI3069 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Mai 2013

Die *Trends in International Mathematics and Science Study* (TIMSS) ist eine international-vergleichende Schulleistungsuntersuchung, die von der *International Association for the Evaluation of Educational Achievement* (IEA) verantwortet wird. Ein zentrales Anliegen ist die Dokumentation langfristiger Entwicklungen in den teilnehmenden Bildungssystemen. Mit TIMSS 2007 wurden Fachleistungen von Schülerinnen und Schülern in den Domänen Mathematik und Naturwissenschaften unter Berücksichtigung curricularer Vorgaben und anderer zentraler Rahmenbedingungen schulischer Lernumgebungen im internationalen Vergleich untersucht.

Die vorliegende Handreichung erläutert – ergänzend zu Dokumentationen der internationalen Studienleitung sowie der für Deutschland vorliegenden Berichtslegung und Dokumentation der Erhebungsinstrumente (vgl. Abschnitt 1.7) – Besonderheiten der Datenerhebung und -aufbereitung zu TIMSS 2007 und gibt damit spezifische Hinweise zur Nutzung und Analyse der Daten. Da sich Deutschland an TIMSS 2007 ausschließlich mit der 4. Jahrgangsstufe beteiligt hat (Bos et al., 2008), bezieht sich die nachfolgende Darstellung ausschließlich auf die Daten und Materialien dieser Studienkomponente.

Die Handreichung ist wie folgt aufgebaut: Zunächst wird ein Überblick zu TIMSS 2007 gegeben. Dieser beinhaltet neben grundlegenden Informationen zu der Studie, der nationalen und internationalen Stichprobe, den nationalen und internationalen Instrumenten auch Übersichten zu den zentralen Publikationen und Materialien. Diese Übersichten verweisen gezielt auf weiterführende Informationen zu Ergebnissen, dem Studiendesign, technischen Dokumentationen, verfügbaren Daten und wichtigen Analysehinweisen. Im zweiten Teil der Handreichung werden die bereitgestellten Datensätze in ihren Strukturen und Besonderheiten erläutert. Im dritten Teil werden in Ergänzung Informationen und Hinweise für den Umgang mit den TIMSS-Daten und ihre Analyse gegeben. Hier wird die Verwendung von GewichtungsvARIABLEN erläutert, auf die Besonderheiten bezüglich der Berechnung von Standardfehlern hingewiesen sowie der Umgang mit *Plausible Values* und Benchmark-VARIABLEN beschrieben.

Dokumentation zu den Datensätzen von TIMSS 2007

I Überblick zu TIMSS 2007	4
1.1 Teilnehmer 2007	4
1.2 Internationale Stichprobe	4
1.3 Stichprobe und Beteiligungsquoten in Deutschland	5
1.4 Erhebungsinstrumente	6
1.4.1 Erweiterung von TIMSS in Deutschland	6
1.4.2 Leistungstests	6
1.4.3 Kontextfragebögen	8
1.5 Beteiligte Institutionen	9
1.6 An TIMSS 2007 angegliederte Studien	10
1.7 Zentrale Publikationen, Dokumentationen und Informationsquellen	10
II Datensätze	15
2.1 Inhalt und Struktur des Datenpaketes	15
2.2 Besonderheiten der Leistungsdatensätze	19
2.3 Besonderheiten der Fragebogendatensätze	19
2.4 Identifikationsvariablen	20
2.5 Fehlende Werte	21
2.6 Zusätzliche Items	22
III Besondere Hinweise für die Datenanalyse	23
3.1 Verwendung der Gewichtungsvariablen	23
3.2 Berechnung von Standardfehlern	26
3.3 Berechnungen mit Plausible Values	27
3.4 Berechnungen mit Benchmarks/Kompetenzstufen	28
Literatur	30

I Überblick zu TIMSS 2007

Die *Trends in International Mathematics and Science Study* (TIMSS) ist eine international-vergleichende Schulleistungsuntersuchung, die von der *International Association for the Evaluation of Educational Achievement* (IEA) verantwortet und seit 1995 alle vier Jahre durchgeführt wird.

Ein zentrales Anliegen von TIMSS ist es, langfristige Entwicklungen in den teilnehmenden Bildungssystemen zu dokumentieren, was eine regelmäßige Teilnahme an der Studie voraussetzt. Staaten, die erstmalig an der Untersuchung teilnehmen, erhalten die Gelegenheit ihre Leistungen im internationalen Vergleich zu verorten. Mit TIMSS werden unter Berücksichtigung curricularer Vorgaben und anderer zentraler Rahmenbedingungen schulischer Lernumgebungen im internationalen Vergleich Fachleistungen von Schülerinnen und Schülern in den Domänen Mathematik und Naturwissenschaften am Ende der vierten und achten Jahrgangsstufe sowie am Ende der Sekundarstufe II (*TIMSS Advanced*) untersucht.

Mit der Erhebung im Jahr 2007 beteiligte sich Deutschland zum ersten Mal an der Grundschuluntersuchung von TIMSS und nahm auch am darauffolgenden Studienzyklus 2011 mit Schülerinnen und Schülern der 4. Jahrgangsstufe an der Erhebung teil (Wendt, Tarelli, Bos, Frey & Vennemann, 2012).

Darüber hinaus hat Deutschland an der ersten TIMS-Studie im Jahr 1995 teilgenommen, damals *Third International Mathematics and Science Study*, jedoch nur Schülerinnen und Schüler der Sekundarstufen I und II einbezogen.

1.1 Teilnehmer 2007

Weltweit waren insgesamt 187 673 Schülerinnen und Schüler der vierten Jahrgangsstufe aus 38 Staaten an TIMSS 2007 beteiligt. Um sich mit internationalen Standards zu messen, haben weitere sieben Regionen als sogenannte Benchmark-Teilnehmer an der Studie teilgenommen (Martin & Mullis, 2004).

1.2 Internationale Stichprobe

Für TIMSS 2007 liegen insgesamt Daten für 187 673 Schülerinnen und Schüler, 6 902 Schulen und 13 612 Lehrkräfte vor. Eine Befragung der Eltern war nicht Teil der internationalen Studienkonzeption.

Für Analysen mit den internationalen Datensätzen ist zu berücksichtigen, dass sinnvolle Vergleiche von Bildungssystemen und ihren Erträgen die Einhaltung strenger methodischer Standards voraussetzen (Mullis et al., 2005). In der internationalen Berichterstattung sind umfangreiche Informationen zu den Teilnahmebedingungen einzelner Staaten und Regionen gegeben, um für eine international-vergleichende Interpretation von Ergebnissen Einschränkungen der Vergleichbarkeit, die aus Besonderheiten nationaler Stichproben resultieren, berücksichtigen zu können (Martin et al.,

2008; Mullis, Martin & Foy et al., 2008; Olson, Martin & Mullis, 2008). Während geringe Abweichungen von den Teilnahmestandards in der regulären Berichterstattung mit Fußnoten am Staatennamen gekennzeichnet werden, führen umfangreiche Abweichungen zum Ausschluss vom regulären internationalen Vergleich. In TIMSS 2007 war die Mongolei von einem entsprechenden Ausschluss betroffen: Die Mongolei hatte sich zwar an TIMSS 2007 beteiligt, es lagen jedoch nicht alle von der internationalen Studienleitung vorgesehenen Dokumentationen zur Stichprobenziehung, Testdurchführung und Testauswertung vor.

Die Ergebnisse der Mongolei werden daher in der internationalen Berichtserstattung (Martin et al., 2008; Mullis, Martin & Foy et al., 2008) zu TIMSS 2007 lediglich gesondert im Anhang berichtet und in der Berichtslegung für Deutschland (Bos et al., 2008) gar nicht berücksichtigt. In die Ergebnisse der 37 Staaten und sieben Regionen, die sich regulär an TIMSS 2007 beteiligen konnten, fließen somit Daten für 183 150 Schülerinnen und Schüler, 6 753 Schulen und 13 459 Lehrkräfte ein.

1.3 Stichprobe und Beteiligungsquoten in Deutschland

Mit der Erhebung im Jahr 2007 beteiligte sich Deutschland zum ersten Mal an der Grundschuluntersuchung von TIMSS. Die Teilnahme erfolgte auf Beschluss der *Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland* (KMK) und einer Vereinbarung zwischen der KMK und dem *Bundesministerium für Bildung und Forschung* (BMBF). Ein Vergleich der Länder der Bundesrepublik Deutschland war in TIMSS 2007 nicht vorgesehen und ist daher mit den vorliegenden Datensätzen nicht möglich.

Die TIMSS-2007-Schülerstichprobe in Deutschland umfasst 5 464 Schülerinnen und Schüler. Von diesen nahmen 5 200 Schülerinnen und Schüler an 246 Schulen tatsächlich am TIMSS-Test teil (Bonsen, Lintorf, Bos & Frey, 2008). Darüber hinaus wurden entsprechend der internationalen Studienkonzeption 246 Schulleitungen und 373 Lehrkräfte mittels standardisierter Fragebögen schriftlich befragt. Der Anteil der Klassen, für die mindestens ein Lehrerfragebogen vorliegt, beträgt 96,8 Prozent und die Rücklaufquote der Schulfragebögen liegt bei 94,8 Prozent (Bonsen et al., 2008). Zur nationalen Ergänzung der Studie zählte unter anderem die Administration eines Elternfragebogens. Die Beteiligungsquote für den Elternfragebogen liegt bei 77 Prozent.

Weitere Informationen zur Stichprobenziehung und den Teilnahmequoten in Deutschland sind in dem nationalen Ergebnisbericht zu TIMSS 2007 (Bos et al., 2008) zu finden.

1.4 Erhebungsinstrumente

In TIMSS kommen zwei Arten von Erhebungsinstrumenten zum Einsatz: Zur Ermittlung von Schülerkompetenzen in Mathematik und den Naturwissenschaften werden Leistungstests eingesetzt. Zur Erfassung umfangreicher Hintergrundinformationen werden Fragebögen auf System-, Schul-, Lehrer- und Schülerebene eingesetzt.

1.4.1 Erweiterung von TIMSS in Deutschland

Die für TIMSS 2007 international vorgesehene Instrumentierung wurde in Deutschland zur genaueren Erfassung national relevanter Faktoren ergänzt. Zu dieser nationalen Ergänzung zählen die Administration einer Skala des *Kognitiven Fähigkeitstests* (KFT; Heller & Perleth, 2000) und eines Elternfragebogens sowie zusätzliche Inhalte in der Schülerteilnahmeliste und den Schüler-, Lehrer- und Schulfragebögen.

Die Datenschutzbeauftragten der Länder der Bundesrepublik Deutschland wurden über alle im Rahmen von TIMSS 2007 implementierten Verfahren der Datenerhebung ausführlich informiert. Durch spezifische Anpassung der Testinstrumente und Erhebungsverfahren wurde sichergestellt, dass der Datenschutz den gesetzlichen Vorgaben entsprechend eingehalten wurde.

Alle Erweiterungen der Kontextfragebögen und der Schülerteilnahmeliste sind in der 2009 erschienenen Dokumentation der Erhebungsinstrumente (Bos, Bensen, Kummer, Lintorf & Frey, 2009) veröffentlicht und als nationale Ergänzungen gekennzeichnet.

1.4.2 Leistungstests

Die in TIMSS eingesetzten Leistungstests dienen der Ermittlung von Leistungsständen der Schülerinnen und Schüler in den Bereichen Mathematik und Naturwissenschaften mit dem Ziel, im internationalen Vergleich Aussagen über Leistungsunterschiede zwischen gesamten Schülerpopulationen sowie zwischen spezifischen Gruppen von Schülerinnen und Schülern für die übergeordneten Kompetenzskalen und die Subdomänen treffen zu können. In den Testheften, welche die Schülerinnen und Schüler während der Testsitzung erhielten und bearbeiteten, sind jeweils sowohl Mathematik- als auch Naturwissenschaftsaufgaben enthalten. Der studienzyklenspezifische Aufgabenpool von TIMSS 2007 umfasst 179 Mathematik- und 174 Naturwissenschaftsaufgaben. Die Zusammenstellung der Testhefte erfolgt über domänenspezifische Blöcke von Aufgaben. Jeder Aufgabenblock umfasst zwischen 10 und 15 Einzelaufgaben. Um die teilnehmenden Kinder nicht übermäßig zu belasten und dennoch einen umfassenden Einblick in ihre fachspezifischen Kompetenzen zu erhalten, liegt der Testkonzeption ein Multi-Matrix-Design zugrunde, welches eine Rotation von 28 verschiedenen Testblöcken (jeweils 14 für Mathematik und die Naturwissenschaften) über 14 unterschiedliche Testhefte vorsieht (Bos et al., 2008.). Durch das Rotationsdesign wird auch

sichergestellt, dass gut die Hälfte der Schülerinnen und Schüler zuerst zwei Aufgabenblöcke aus dem Kompetenzbereich Mathematik bearbeitet und im Anschluss die Aufgabenblöcke für den Bereich Naturwissenschaften, während für die andere Hälfte die Administration der domänenspezifischen Testblöcke in umgekehrter Reihenfolge erfolgt.

Alle Testaufgaben sind kindgerecht aufbereitet und zum Teil illustriert. Etwas mehr als die Hälfte der Testaufgaben wird im Multiple-Choice-Format gehalten, bei dem die Schülerinnen und Schüler aus zwei, vier oder fünf vorgegebenen Antworten die richtige Lösung auswählen müssen. Etwas weniger als die Hälfte der Aufgaben ist durch ein offenes Antwortformat gekennzeichnet, welches den Schülerinnen und Schülern die Möglichkeit gibt, ihre Antworten frei zu formulieren. Die Entwicklung der Testaufgaben erfolgt in einem kooperativen Prozess der TIMSS-Teilnehmerstaaten. Die Aufgabenvorschläge stammen aus verschiedenen Teilnehmerstaaten und werden von Expertinnen und Experten aus verschiedenen Fachdisziplinen und Ländern begutachtet. Ein maßgebliches Kriterium bei der Entwicklung und Überprüfung von Testaufgaben besteht darin, dass die Aufgaben die in der TIMSS-Rahmenkonzeption formulierten Testziele, Inhalte und Anforderungen repräsentieren (Mullis et al., 2005).

Der Testentwicklung liegt entsprechend ein Kompetenzmodell zugrunde, das für die Domänen Mathematik und naturwissenschaftliche Grundbildung getrennt spezifiziert, welche grundlegenden Stoffgebiete (Inhaltsbereiche) und fachspezifisch zu aktivierenden kognitiven Denkprozesse und -tätigkeiten (kognitive Anforderungsbereiche) durch die Leistungstests abgebildet werden sollen (Mullis et al., 2005). Inhaltlich wird für die Erfassung mathematischer Kompetenzen am Ende der vierten Jahrgangsstufe zwischen den Bereichen *Arithmetik*, *Geometrie/Messen* und *Daten* unterschieden. Für die Naturwissenschaften werden die Bereiche *Biologie*, *Physik* und *Geographie* differenziert, wobei der Inhaltsbereich *Physik* auch Themenstellungen umfasst, die in Deutschland traditionell der Chemie zugeordnet werden. Die einzelnen Inhaltsbereiche werden zudem fachbezogen in verschiedene Themengebiete aufgegliedert. Um grundlegende Denkprozesse zu beschreiben, die für eine erfolgreiche Aufgabenbearbeitung zu leisten sind, werden die kognitiven Anforderungsbereiche *Reproduzieren*, *Anwenden* und *Problemlösen* unterschieden, denen wiederum konkrete Verhaltensweisen zuzuordnen sind.

Eine Auswahl an Testaufgaben ist auf der Seite der internationalen Studienleitung veröffentlicht (vgl. Tabelle 3). Da TIMSS eine Trendstudie ist, wird nur ein Teil der Testaufgaben veröffentlicht, um die anderen Aufgaben auch in kommenden Studienzyklen einsetzen zu können. Eine Einsicht der unveröffentlichten Testaufgaben ist daher nicht möglich.

Weitere Informationen zum Testumfang, zu Aufgabenanteilen am Test sowie zum Rotationsverfahren sind in der nationalen Berichtslegung zu TIMSS 2007 (Bos et al., 2008) sowie in der Beschreibung der Rahmenkonzeption der Studie zu finden (Mullis et al., 2005).

In Deutschland wurde zusätzlich zu dem Mathematik- und Naturwissenschaftstest eine Subskala (Skala N2 Form A) des KFT administriert. Die Auswertung erfolgte entsprechend des Manuals.

1.4.3 Kontextfragebögen

Die an TIMSS teilnehmenden Schülerinnen und Schüler sowie die zentral an ihrem Lernprozess beteiligten Akteure werden zusätzlich mit Hilfe von Fragebögen um Auskunft zu Kontextmerkmalen des schulischen und häuslichen Lernens gebeten. Die von Seiten der internationalen Studienleitung vorgesehenen Inhalte der Kontextfragebögen stellen für alle Teilnehmerstaaten verbindliche Vorgaben dar (Mullis et al., 2009). In den in Deutschland eingesetzten TIMSS-2007-Fragebögen werden unter anderem folgende Aspekte thematisiert:

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

Mit dem *Elternfragebogen*, der jedoch nur in Deutschland eingesetzt wurde, können die individuellen Schülerangaben um Informationen zur familiären Lernumwelt ergänzt und erweitert werden. Dieser fokussiert unter anderem familiäre Ressourcen, soziodemographische Daten (z.B. sozialer Status, höchster Bildungsabschluss, Fragen zur Berufstätigkeit, Informationen über die Zusammensetzung der Familie), Bildungshintergrund der Eltern, Bildungsaspirationen der Eltern, lernunterstützende Aktivitäten und den kognitiven Anregungsgehalt der familiären Lernumwelt.

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

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

Der *Curriculumfragebogen*, der von Fachexpertinnen und Fachexperten ausgefüllt wird, enthält Fragen dazu, inwieweit die Inhalte der TIMSS-Tests in

den jeweils intendierten Curricula vorgesehen sind. Daneben werden auch Rahmenbedingungen der Umsetzung des Curriculums erfragt, zum Beispiel wer über das jeweilige Curriculum entscheidet, ob und in welcher Form die Implementation des Curriculums evaluiert wird oder wie Lehrkräfte bei der Umsetzung des Curriculums unterstützt werden. Die mittels des *Curriculumfragebogens* erhobenen Daten sind nicht Teil der Datensätze, sondern werden in aufbereiteter Form als Teil der internationalen Berichterstattung und Dokumentation zur Verfügung gestellt (Martin et al., 2008; Mullis, Martin & Foy et al., 2008; Mullis, Martin & Olson et al., 2008).

1.5 Beteiligte Institutionen

Die Realisierung internationaler Vergleichsstudien wie TIMSS erfordert die Zusammenarbeit verschiedener Organisationen, Institutionen und Personen auf nationaler und internationaler Ebene. Die IEA hat mit der internationalen Leitung der Studie das *TIMSS & PIRLS International Study Center* (ISC) am Boston College in Chestnut Hill, Massachusetts, USA beauftragt. Die Durchführung von TIMSS 2007 in Deutschland wurde zu gleichen Teilen vom *Bundesministerium für Bildung und Forschung* (BMBF) und durch die *Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland* (KMK) finanziert. Für die Analyse der Studienergebnisse und die Berichterlegung in Deutschland war ein nationales Konsortium unter Federführung von Prof. Dr. Wilfried Bos, Direktor des *Instituts für Schulentwicklungsforschung* (IFS), verantwortlich. Tabelle 1 gibt einen Überblick über die an TIMSS 2007 beteiligten Institutionen.

Tabelle 1: Übersicht zu beteiligten Institutionen (international und national) an TIMSS 2007

Beteiligte Institutionen
<p>international</p> <p>International Association for the Evaluation of Educational Achievement (IEA) Onlineresource: http://www.iea.nl/timss_2007.html</p> <p>International Study Center, Boston College (BC) Onlineresource: http://timssandpirls.bc.edu/TIMSS2007/index.html</p> <p>Data Processing and Research Center (DPC) Onlineresource: http://www.iea-dpc.de/timss_national200700.html?&L=1</p> <p>national</p> <p>Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) Onlineresource: http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsmonitoring/internationale-schulleistungsvergleiche/timss.html</p> <p>Bundesministerium für Bildung und Forschung (BMBF) Onlineresource: http://www.bmbf.de/de/6628.php</p> <p>Institut für Schulentwicklungsforschung (IFS) Onlineresource: http://www.ifs-dortmund.de/index.php?id=173</p> <p>Forschungsdatenzentrum (FDZ) Onlineresource: http://www.iqb.hu-berlin.de/fdz/</p>

1.6 An TIMSS 2007 angegliederte Studien

An die Datenerhebung von TIMSS 2007 wurden zwei weitere Bildungsstudien angegliedert (Bos et al., 2008). Dabei handelte es sich um die von der *Deutschen Forschungsgemeinschaft* (DFG) geförderte Studie „Übergang – Der Übergang von der Grundschule in die weiterführende Schule – Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten“ (Maaz, Baumert, Gresch & McElvany, 2010) vom *Berliner Max-Planck-Institut für Bildungsforschung* (MPIB) und um eine Studie vom *Institut zur Qualitätsentwicklung im Bildungswesen* (IQB) zur Normierung von Testaufgaben zu den Bildungsstandards in den Fächern Deutsch und Mathematik in der 3. und 4. Klasse (IQB - Institut zur Qualitätsentwicklung im Bildungswesen). Die vom *Forschungsdatenzentrum* (FDZ) zur Verfügung gestellten Daten und die hier vorliegende Dokumentation beziehen sich ausschließlich auf die Schülerinnen und Schüler sowie die Test- und Fragebogeninstrumente, die Teil von TIMSS 2007 waren.

1.7 Zentrale Publikationen, Dokumentationen und Informationsquellen

Die vorliegende Handreichung erläutert Besonderheiten der Datenerhebung und -aufbereitung zu TIMSS 2007 und gibt damit wichtige Hinweise zur Nutzung und Analyse der Daten. Diese Handreichung versteht sich dabei als Ergänzung zu bereits vorliegenden Dokumentationen und Publikationen zu TIMSS 2007. Die Tabellen 2 und 3 geben einen umfassenden Überblick zu zentralen Publikationen, Dokumentationen und Materialien, die bereits zu TIMSS 2007 vorliegen. Vertiefende Informationen und Ergebnisse zu TIMSS 2007 finden sich in den nationalen (Bos et al., 2008) und den internationalen (Martin et al., 2008; Mullis, Martin & Foy et al., 2008) Berichtsbänden. Diese werden ergänzt durch die Beschreibung der Rahmenkonzeption der Studie (Mullis et al., 2005), die Enzyklopädie zu den nationalen Kontextbedingungen des Lehrens und Lernens der TIMSS-2007-Teilnehmerstaaten im Bereich Mathematik und Naturwissenschaften (Mullis et al., 1998) sowie den *Technical Report* zu TIMSS 2007 (Olson et al., 2008). Umfassende Hinweise und Anleitungen zur Datenanalyse finden sich im TIMSS 2007 *User Guide* (Foy & Olson, 2009).

Die Dokumentation der internationalen (TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, 2007) und nationalen Erhebungsinstrumente (Bos et al., 2009) bieten einen Überblick über die in den Datensätzen enthaltenen Variablen und damit Einblick in das Analysepotential der Untersuchung.

Ein Teil der Testaufgaben bleibt zur Ermöglichung von Trendanalysen unveröffentlicht. Die freigegebenen Items aus TIMSS 2007 sind online abrufbar.¹

¹ Siehe hierzu die Internetseiten des TIMSS & PIRLS International Study Center [<http://timss.bc.edu>] und der IEA [<http://www.iea.nl>]

Für Sekundäranalysen haben Wissenschaftlerinnen und Wissenschaftler über das ISC und die IEA Zugriff auf alle internationalen Daten, die im Rahmen von TIMSS erhoben wurden.² Über das FDZ besteht darüber hinaus Zugriff auf die im Rahmen der nationalen Erweiterung (vgl. Abschnitt 1.4.1) erhobenen Daten aus TIMSS 2007.³

² Siehe hierzu die Internetseiten des TIMSS & PIRLS International Study Center [<http://timss.bc.edu>] und der IEA [<http://www.iea.nl>]

³ Siehe hierzu die Internetseiten des FDZ: <http://www.iqb.hu-berlin.de/fdz>

Tabelle 2: Übersicht zentraler Publikationen zu TIMSS 2007 (international)

Zentrale Publikationen
<p>Ergebnisberichte</p> <p>International Mathematics Report Mullis, I. V. S., Martin, M. O. & Foy, P. (2008). <i>TIMSS 2007. International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades</i>. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Korrigierte Onlineresource¹: http://timssandpirls.bc.edu/TIMSS2007/mathreport.html</p> <p>International Science Report Martin, M. O., Mullis, I. V. S., Foy, P. (2008). <i>TIMSS 2007. International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades</i>. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Korrigierte Onlineresource¹: http://timssandpirls.bc.edu/TIMSS2007/sciencereport.html</p> <p>Die internationalen Berichte - der <i>International Mathematics Report</i> und der <i>International Science Report</i> - fassen die Ergebnisse des internationalen Vergleichs von 67 Teilnehmerstaaten und Regionen zu den mathematischen beziehungsweise naturwissenschaftlichen Kompetenzen von Schülerinnen und Schülern der vierten und achten Jahrgangsstufe zusammen und beschreiben die Rahmenbedingungen mathematischer beziehungsweise naturwissenschaftlicher Lehr-Lernsituationen in den Teilnehmerstaaten. Die Leistungen der Schülerinnen und Schüler werden zudem nach den Jahrgangsstufen, internationalen TIMSS-Benchmarks sowie nach Inhaltsbereichen und kognitiven Anforderungsprozessen differenziert dargestellt. Für Staaten und Regionen, die sich bereits an vorangegangenen Studienzyklen beteiligt haben, werden darüber hinaus Trends über den Zeitraum der Erhebungszyklen 1995, 1999, 2003 und 2007 abgebildet. Die Berichte bieten zudem Analysen zu den ausgewählten individuellen und schulischen Rahmenbedingungen.</p> <p>Beschreibung des Studiendesigns</p> <p>Framework Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A. & Erberber, E. (2009). <i>TIMSS 2007. Assessment Frameworks</i>. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Onlineresource: http://timssandpirls.bc.edu/TIMSS2007/frameworks.html</p> <p>Die <i>TIMSS 2007 Assessment Frameworks</i> wurden in Zusammenarbeit von Experten aus über 60 Ländern erstellt. Die Frameworks beinhalten drei Rahmenkonzepte für die Umsetzung von TIMSS 2007 - das mathematische, das naturwissenschaftliche sowie das kontextbezogene Rahmenkonzept für die Fragebögen. Es liefert zudem eine Übersicht über das Studiendesign und grundsätzliche Informationen zur Testentwicklung.</p> <p>Beschreibung der Rahmenbedingungen in den Teilnehmerländern</p> <p>Encyclopedia Mullis, I. V. S., Martin, M. O., Olson, J. F., Berger, D. R., Milne, D. & Stanco, G. M. (Hrsg.). (2008). <i>TIMSS 2007. Encyclopedia: A Guide to Mathematics and Science Education Around the World, Volumes 1 and 2</i>. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Onlineresource: http://timssandpirls.bc.edu/TIMSS2007/encyclopedia.html</p> <p>Die <i>TIMSS 2007 Encyclopedia</i> beschreibt die nationalen Rahmenbedingungen des Lehrens und Lernens in den Bereichen Mathematik und Naturwissenschaften in den Teilnehmerstaaten aus TIMSS 2007. Für nahezu alle Teilnehmerstaaten und Regionen werden in einem eigenen Kapitel die Struktur des Schulsystems, die mathematischen und naturwissenschaftlichen Curricula, die Bedingungen und Anforderungen der Lehrerbildung sowie der Einsatz von Lernstandserhebungen und Prüfungen in der Schulzeit zusammenfassend beschrieben. Damit bietet die in zwei Bände (A-L und M-Z) aufgeteilte TIMSS 2007 Encyclopedia eine wichtige Grundlage zur Einordnung relevanter Merkmale und Rahmenbedingungen mathematischer und naturwissenschaftlicher Lehr- und Lernbedingungen mit einem Fokus auf die Grundschulzeit bis zur achten Jahrgangsstufe weltweit.</p> <p>Technische Dokumentation</p> <p>Technical Report Olson, J. F., Martin, M. O. & Mullis, I. V. S. (Hrsg.). (2008). <i>TIMSS 2007. Technical Report</i>. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Onlineresource: http://timssandpirls.bc.edu/TIMSS2007/techreport.html</p> <p>Der <i>TIMSS 2007 Technical Report</i> beinhaltet technische Informationen zum Design und der Durchführung der Leistungserhebung sowie Details zur Erstellung der Erhebungsinstrumente von TIMSS 2007. Er beschreibt die technischen und methodischen Vorgehensweisen zur Stichprobenziehung, Datenerhebung, dem Skalierungsverfahren, der Datenanalyse und der Berichtslegung. Zudem liefert der TIMSS 2007 Technical Report detaillierte Informationen zum Vorgehen und den Methoden, welche in TIMSS 2007 angewendet wurden, um international vergleichbare und qualitativ hochwertige Daten bereit zu stellen. Der Bericht erläutert die Schritte der Qualitätssicherung und -kontrolle, die von der Überarbeitung der Rahmenkonzeption bis zur Veröffentlichung der internationalen Datensätze sowie des User Guides implementiert wurden.</p>

¹ Die Onlinedokumente beinhalten im Vergleich zur Printversion korrigierte Werte für die mathematischen und naturwissenschaftlichen kognitiven Anforderungsbereiche Reproduzieren und Anwenden.

Tabelle 2 Fortsetzung: Übersicht zu zentralen Publikationen zu TIMSS 2007 (national)

Zentrale Publikationen
<p>Ergebnisbericht</p> <p>TIMSS 2007 Ergebnisbericht <i>Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.). (2008). TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Münster: Waxmann.</i> Dieser Band erschließt die Ergebnisse des internationalen Vergleichs für die deutsche Bildungsdiskussion. Die Ergebnisdarstellung erfolgt dabei vor dem Hintergrund zentraler Themen, die den Diskurs in den letzten Jahren besonders dominiert haben. Die Kompetenzen werden nach zentralen Merkmalen wie soziale Herkunft, Migrationsstatus und Gesellschaft analysiert.</p> <p>Dokumentation der Erhebungsinstrumente</p> <p>TIMSS 2007 Dokumentation der Erhebungsinstrumente <i>Bos, W., Bonsen, M., Kummer, N., Lintorf, K. & Frey, K. (Hrsg.). (2009). TIMSS 2007. Dokumentation der Erhebungsinstrumente zur Trends in International Mathematics and Science Study. Münster: Waxmann.</i> Mit dem Handbuch werden die im Rahmen von TIMSS 2007 in Deutschland eingesetzten Befragungsinstrumente dokumentiert. Das Handbuch enthält sowohl die internationalen Instrumente, die in allen Teilnehmerstaaten eingesetzt wurden, als auch die nationalen Ergänzungen.</p>

Tabelle 3: Weitere Dokumentationen und Materialien sowie Datensätze zu TIMSS 2007 (international)

Weitere Materialien
<p>International Database</p> <p>Foy, P. & Olson, J. F. (Hrsg.). (2009). <i>TIMSS 2007. International Database and User Guide</i>. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Online resource: http://timss.bc.edu/timss2007/idb_ug.html Die <i>TIMSS 2007 International Database</i> beinhaltet die Daten der Schülerinnen und Schüler, ihrer Lehrkräfte, der Schule und des curricularen Hintergrundes.</p> <p>Almanacs Die <i>Almanacs</i> beinhalten deskriptive Statistiken für alle Items und Hintergrundvariablen.</p> <p>Codebooks Die <i>Codebooks</i> beinhalten Beschreibungen aller Variablen der TIMSS 2007 International Database.</p> <p>SAS Data¹ <i>SAS Data</i> beinhaltet die TIMSS 2007 Daten im SAS Format.</p> <p>SPSS Data¹ <i>SPSS Data</i> beinhaltet die TIMSS 2007 Daten im SPSS Format.</p> <p>Curriculum Unter <i>Curriculum</i> sind Informationen aus den Curriculum Questionnaires zu finden.</p> <p>Items Der Punkt <i>Items</i> umfasst Iteminformationen, IRT Itemparameter und alle für den TIMSS 2007 Zyklus veröffentlichten Testaufgaben.</p> <p>Programs Programme und Makros für SAS, SPSS und der <i>IEA International Database Analyzer</i> (IDB Analyzer) sind unter <i>Programs</i> zu finden. Der IDB Analyzer ist eine seit 2004 vom DPC entwickelte Anwendung, welche die Analyse von Daten aus Large Scale Assessments erleichtert. Der IDB Analyzer erzeugt SPSS Syntaxen, um statistische Analysen unter Berücksichtigung der komplexen Samplingstruktur der Daten durchzuführen.</p> <p>TCMA Die <i>Test-Curriculum Matching Analyse</i> (TCMA) gibt Auskunft darüber, inwiefern die einzelnen Testaufgaben durch die Curricula der jeweiligen Teilnehmer abgedeckt werden. Ihre Ergebnisse sind in einer Tabelle zusammengestellt.</p> <p>User Guide</p> <p>Foy, P. & Olson, J. F. (Hrsg.). (2009). <i>TIMSS 2007. International Database and User Guide</i>. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Online resource: http://timss.bc.edu/timss2007/idb_ug.html Der <i>TIMSS 2007 User Guide</i> beschreibt den Inhalt und das Format der Daten der TIMSS 2007 International Database. Er stellt den IDB Analyzer vor und führt mit Beispielen in die Nutzung der TIMSS 2007 Daten in Verbindung mit SPSS ein. Außerdem erläutert er die Nutzung der Database und beschreibt ihre Struktur und ihren Inhalt.</p> <p>Contextual Questionnaires</p> <p>Online resource: http://timssandpirls.bc.edu/TIMSS2007/context.html Die <i>TIMSS 2007 Contextual Questionnaires</i> umfassen die in TIMSS 2007 eingesetzten Erhebungsinstrumente. Enthalten sind Fragebögen, mit denen umfangreiche Informationen zu Lern- und Lebensbedingungen über die Befragung der Schülerinnen und Schüler, Lehrkräfte, Schulleitungen und Curriculum Experten erfasst wurden.</p>

¹ Um Veränderungen über verschiedene Messzeitpunkte hinweg beobachten zu können, wurde in TIMSS 2011 ein neuer Skalierungsansatz für die Subskalen (mathematische und naturwissenschaftliche Inhalts- und kognitive Anforderungsbereiche) eingeführt. Diese Vorgehensweise wurde rückwirkend auch auf die entsprechenden Subskalen in TIMSS 2007 angewendet, so dass diese Daten Trendvergleiche zu TIMSS 2011 ermöglichen. Diese Änderung wurde auch auf die SAS und SPSS Daten übertragen. Die Leistungswerte auf den Gesamtskalen Mathematik und Naturwissenschaften bleiben hiervon jedoch unberührt.

II Datensätze

In den folgenden Abschnitten werden die bereitgestellten Datensätze zu TIMSS 2007 in ihren Strukturen und Besonderheiten erläutert.

2.1 Inhalt und Struktur des Datenpaketes

Das vollständige Datenpaket zu TIMSS 2007, welches dem FDZ vorliegt, umfasst 12 SPSS-Dateien (vgl. Tabelle 4), die sich in zwei Pakete (mit jeweils sechs Teildatensätzen) untergliedern lassen:

- (1) Internationale Datensätze, die mit den internationalen Instrumenten aus TIMSS 2007 erhobene Daten für alle Teilnehmerstaaten umfassen und
- (2) Datensätze für Deutschland, welche zusätzlich zu den internationalen Daten die im Rahmen der nationalen Erweiterung erhobenen Daten enthalten (vgl. Abschnitt 1.4.1).

(1) Internationale Datensätze

Während ein Zugriff auf die Datensätze (2) nur über das FDZ möglich ist, ist der Teil des internationalen Datenpakets (1) in Form von Einzeldatensätzen für jeden Teilnehmer online frei über die *International Database* verfügbar (vgl. Tabelle 3).

Die Teildatensätze sind entsprechend der Datenstruktur der internationalen Studienleitung nach Personengruppen gegliedert. Die Bezeichnung der internationalen Datensätze wurde analog zur *International Database* (vgl. Tabelle 3) vorgenommen. Die online verfügbare *International Database* enthält für jedes Teilnehmerland jeweils sechs Datensätze: einen Schulleitungsdatensatz (ACGxxxM4), einen Lehrerdatsatz (ATGxxxM4), einen Schülerdatensatz (ASGxxxM4), einen Leistungsdatensatz (ASAxM4), einen Datensatz mit Informationen zum *Reliability-Scoring* für die *Constructed-Response-Items* (ASRxxxM4) sowie eine Schüler-Lehrer-Linkdatei (ASTxxxM4). Zu beachten ist, dass die online verfügbare *International Database* für TIMSS 2007 nach der Berichtslegung zu TIMSS 2011 aktualisiert wurde und für die auf die Subdomänen bezogenen Leistungsvariablen „neue Schätzer“ enthält (vgl. Abschnitt 2.2).

Das internationale Datenpaket, welches über das FDZ zur Verfügung steht, umfasst sechs Datensätze. In diesen sind die Daten aller TIMSS 2007 Teilnehmer enthalten. Vier der sechs Datensätze aus dem Datenpaket (1) enthalten die Werte aus der *International Database [Stand 05.2013]*; dazu zählen: ein Schulleitungsdatensatz (T07_ACG), ein Schülerdatensatz (T07_ASG), ein Leistungsdatensatz (T07_ASA) sowie ein kombinierter Lehrer-Schülerdatensatz (T07_ATG).

Außerdem befinden sich in dem Datenpaket zwei Datensätze, die den Stand der *International Database* aus dem Jahr 2008 entsprechen und damit jene auf die Subdomänen bezogenen Leistungsvariablen beinhalten, die für die Berichtslegung zu TIMSS 2007 genutzt wurden (vgl. Abschnitt 2.2). Dazu

zählen ein kombinierter Lehrer-Schülerdatensatz (T07_ATG_Report07) und ein Datensatz (T07_Report07), der nur diese ausgewählten Leistungswerte enthält und es über Identifikationsvariablen (vgl. Abschnitt 2.4) ermöglicht, diese an weitere Datensätze zu spielen.

(2) Datensätze für Deutschland

Neben dem internationalen Datenpaket werden über das FDZ sechs weitere Datensätze für Deutschland bereitgestellt, die neben Daten zu den internationalen Instrumenten auch jene zu den nationalen Erweiterungen enthalten (vgl. Abschnitt 1.4.1). Vier Datensätze des nationalen Pakets enthalten die Werte aus der aktuellen *International Database* und sind um die nationalen Variablen ergänzt; dazu zählen: ein Schulleitungsdatensatz (T07_SLFB), ein Schülerdatensatz (T07_SEFB), ein Leistungsdatensatz (T07_Leistung) sowie ein kombinierter Lehrer-Schülerdatensatz (T07_LSFB). Außerdem befinden sich in dem Datenpaket zwei Datensätze, die den Stand der *International Database* aus dem Jahr 2008 entsprechen und damit jene auf die Subdomänen bezogenen Leistungsvariablen beinhalten, die für die Berichtslegung zu TIMSS 2007 genutzt wurden (vgl. Abschnitt 2.2). Dazu zählen ein kombinierter Lehrer-Schülerdatensatz (T07_LSFB_Report07) und ein Datensatz (T07_DEU_Report07), der nur diese ausgewählten Leistungswerte enthält und es über Identifikationsvariablen (vgl. Abschnitt 2.4) ermöglicht, diese an weitere Datensätze zu spielen.

Tabelle 4 gibt einen Überblick zu Struktur, Inhalt und Umfang der verschiedenen Teildatensätze.

Tabelle 4: Übersicht zu den Datensätzen

(1) Gesamtdatensätze international Datensatz & Beschreibung	N (Fälle/ Variablen)
T07_ASG.sav Schülerdatensatz Internationale Daten aus dem Schülerfragebogen, zusätzlich sind die Plausible Values (PV) für die Gesamt- und Subskalen enthalten. <i>Besonderheiten:</i> Die Plausible Values (PV) der Subdomänen entsprechen der Berichtslegung zu TIMSS 2011 (Foy et al., 2011; Mullis et al., 2012; Martin et al., 2012). Der Datensatz enthält Fälle für die Mongolei, die in der Berichtslegung zu TIMSS 2007 gesondert aufgeführt wurden.	187 673 / 203
T07_ASA.sav Leistungsdatensatz Internationale Leistungsdaten auf Itemebene, zusätzlich sind die Plausible Values (PV) für die Gesamt- und Subskalen enthalten. <i>Besonderheiten:</i> Die Plausible Values (PV) der Subdomänen entsprechen der Berichtslegung zu TIMSS 2011 (Foy et al., 2011; Mullis et al., 2012; Martin et al., 2012). Der Datensatz enthält Fälle für die Mongolei, die in der Berichtslegung zu TIMSS 2007 gesondert aufgeführt wurden.	187 673 / 475
T07_ATG.sav Lehrerdatsatz Internationale Daten aus dem Lehrerfragebogen, zusätzlich sind die Plausible Values (PV) für die Gesamt- und Subskalen enthalten. <i>Besonderheiten:</i> Die Lehrerdaten wurden disaggregiert und mit den Schülerdaten gematcht; sofern mehrere Lehrerfragebögen für eine Klasse vorliegen, so sind die jeweiligen Schüler mehrfach im Datensatz enthalten. Die Plausible Values (PV) der Subdomänen entsprechen der Berichtslegung zu TIMSS 2011 (Foy et al., 2011; Mullis et al., 2012; Martin et al., 2012). Der Datensatz enthält Fälle für die Mongolei, die in der Berichtslegung zu TIMSS 2007 gesondert aufgeführt wurden.	252 153 / 386
T07_ACG.sav Schulleitungsdatensatz Internationale Daten aus dem Schulleiterfragebogen <i>Besonderheiten:</i> Der Datensatz enthält Fälle für die Mongolei, die in der Berichtslegung zu TIMSS 2007 gesondert aufgeführt wurden.	6 902 / 117
T07_Report07.sav Leistungswerte zum Bericht TIMSS 2007 Internationale Daten für das ursprüngliche Skalierungsmodell (vgl. Technical Report 2007, International Report 2007) <i>Besonderheiten:</i> Die Plausible Values (PV) der Subdomänen entsprechen der Berichtslegung zu TIMSS 2007 (Olson et al., 2008; Martin et al., 2008; Mullis et al., 2008). Der Datensatz enthält Fälle für die Mongolei, die in der Berichtslegung zu TIMSS 2007 gesondert aufgeführt wurden.	187 673 / 103
T07_ATG_Report07.sav Lehrerdatsatz Internationale Daten aus dem Lehrerfragebogen, zusätzlich sind die Plausible Values (PV) für die Gesamt- und Subskalen enthalten. <i>Besonderheiten:</i> Die Lehrerdaten wurden disaggregiert und mit den Schülerdaten gematcht; sofern mehrere Lehrerfragebögen für eine Klasse vorliegen, so sind die jeweiligen Schüler mehrfach im Datensatz enthalten. Die Plausible Values (PV) der Subdomänen entsprechen der Berichtslegung zu TIMSS 2007 (Olson et al., 2008; Martin et al., 2008; Mullis et al., 2008). Der Datensatz enthält Fälle für die Mongolei, die in der Berichtslegung zu TIMSS 2007 gesondert aufgeführt wurden.	252 153 / 386

Tabelle 4 Fortsetzung: Übersicht zu den Datensätzen

(2) Gesamtdatensätze für Deutschland Datensatz & Beschreibung	N (Fälle/ Variablen)
T07_SEFB.sav Schüler-Eltern-Datensatz Daten für Deutschland aus dem Schüler- und Elternfragebogen sowie der Schülerteilnahmeliste, zusätzlich sind die Plausible Values (PV) für die Gesamt- und Subskalen sowie der T-Wert für den KFT enthalten. <i>Besonderheiten:</i> Die Plausible Values (PV) der Subdomänen entsprechen der Berichtslegung zu TIMSS 2011 (Foy et al., 2011; Mullis et al., 2012; Martin et al., 2012). Variablen, die im Label mit einem „- a“ gekennzeichnet sind, wurden im Sinne von Konsistenz- und Plausibilitätschecks aufbereitet.	5 200 / 590
T07_Leistung.sav Leistungsdatensatz Leistungsdaten und Daten für eine Subskala des kognitiven Fähigkeitstests auf Itemebene für Deutschland, zusätzlich sind die Plausible Values (PV) für die Gesamt- und Subskalen sowie der T-Wert für den KFT enthalten. <i>Besonderheiten:</i> Die Plausible Values (PV) der Subdomänen entsprechen der Berichtslegung zu TIMSS 2011 (Foy et al., 2011; Mullis et al., 2012; Martin et al., 2012).	5 200 / 503
T07_LSFB.sav Lehrerdatsatz Daten für Deutschland aus dem Lehrerfragebogen, zusätzlich sind die Plausible Values (PV) für die Gesamt- und Subskalen sowie der T-Wert für eine Subskala des kognitiven Fähigkeitstests enthalten. <i>Besonderheiten:</i> Die Lehrerdaten wurden disaggregiert und mit den Schülerdaten gematcht; sofern mehrere Lehrerfragebögen für eine Klasse vorliegen, so sind die jeweiligen Schüler mehrfach im Datensatz enthalten. Die Plausible Values (PV) der Subdomänen entsprechen der Berichtslegung zu TIMSS 2011 (Foy et al., 2011; Mullis et al., 2012; Martin et al., 2012). Variablen, die im Label mit einem „- a“ gekennzeichnet sind, wurden im Sinne von Konsistenz- und Plausibilitätschecks aufbereitet.	7 851 / 608
T07_SLFB.sav Schulleitungsdatensatz Daten für Deutschland aus dem Schulleiterfragebogen <i>Besonderheiten:</i> Variablen, die im Label mit einem „- a“ gekennzeichnet sind, wurden im Sinne von Konsistenz- und Plausibilitätschecks aufbereitet.	246 / 399
T07_DEU_Report07.sav Leistungswerte zum Bericht TIMSS 2007 Daten für Deutschland für das ursprüngliche Skalierungsmodell <i>Besonderheiten:</i> Die Plausible Values (PV) der Subdomänen entsprechen der Berichtslegung zu TIMSS 2007 (Olson et al., 2008; Martin et al., 2008; Mullis et al., 2008).	5 200 / 103
T07_LSFB_Report07.sav Lehrerdatsatz Daten für Deutschland aus dem Lehrerfragebogen, zusätzlich sind die Plausible Values (PV) für die Gesamt- und Subskalen sowie der T-Wert für eine Subskala des kognitiven Fähigkeitstests enthalten. <i>Besonderheiten:</i> Die Lehrerdaten wurden disaggregiert und mit den Schülerdaten gematcht; sofern mehrere Lehrerfragebögen für eine Klasse vorliegen, so sind die jeweiligen Schüler mehrfach im Datensatz enthalten. Die Plausible Values (PV) der Subdomänen entsprechen der Berichtslegung zu TIMSS 2007 (Olson et al., 2008; Martin et al., 2008; Mullis et al., 2008). Variablen, die im Label mit einem „- a“ gekennzeichnet sind, wurden im Sinne von Konsistenz- und Plausibilitätschecks aufbereitet.	7 851 / 608

2.2 Besonderheiten der Leistungsdatensätze

Die Schülerdatensätze und die kombinierten Lehrer-Schülerdatensätze aus dem nationalen sowie internationalen Datenpaket enthalten zusätzlich zu Variablen die sich auf die Fragebogenitems beziehen, unterschiedliche Variablen für die Leistungen der Schülerinnen und Schüler.

Da für Sekundäranalysen in der Regel nicht einzelne Testaufgaben sondern aggregierte Leistungsscores verglichen werden, sind die Leistungsdaten auf Itemebene jeweils in einem separaten Datensatz enthalten.

Seit TIMSS 2011 wird für die Ermittlung von Schätzern für die Schülerleistungen in den einzelnen Inhaltsbereichen und kognitiven Anforderungsprozessen ein neuer Skalierungsansatz genutzt. Um auch Trends berichten zu können, wurden ebenfalls die Leistungsdaten aller vorangegangenen Studienzyklen neu skaliert und ersetzt (Foy, Brossman & Galia, 2011). Die vorliegenden Datensätze enthalten die durch den neuen Skalierungsansatz ermittelten Werte, wie sie in der Berichtslegung von TIMSS 2011 für TIMSS 2007 berichtet wurden. Sie erlauben damit Trendanalysen für die Subbereiche.

Um auch die Ergebnisse aus der Berichtslegung zu TIMSS 2007 vollständig replizieren zu können, sind in gesonderten Datensätzen auch jene auf die Subdomänen bezogenen Leistungsvariablen bereitgestellt, die für die Berichtslegung zu TIMSS 2007 genutzt wurden. Zu beachten ist dabei, dass in den Printversionen der internationalen sowie nationalen Berichtslegung zu TIMSS 2007 die Subskalen *Knowing* (Replizieren) und *Applying* (Anwenden) vertauscht waren. Die Datensätze und die Onlineversionen der internationalen Ergebnisberichte wurden nachträglich von der internationalen Studienleitung korrigiert. Alle Datensätze enthalten die korrigierten Werte für die Subskalen.

Zu beachten ist zudem, dass die Leistungswerte sowohl für die Gesamt- als auch für die Subskalen in den jüngeren Datensatzversionen mit zwei Nachkommastellen vorliegen, während in den „Report07“-Datensätzen mehr Nachkommastellen enthalten sind. Für Berechnungen mit den aktualisierten Datensätzen kann es dadurch zu geringfügigen Abweichungen im Vergleich zu den Ergebnissen aus der Berichtslegung von TIMSS 2007 kommen.

2.3 Besonderheiten der Fragebogendatensätze

Der Schüler-Elterndatensatz für Deutschland umfasst alle Daten zu Angaben aus der Schülerteilnahmeliste, den Schülerfragebögen und dem Elternfragebogen.

Zu beachten ist, dass die ausführlichen Aufbereitungs- und Bereinigungsschritte der nationalen Datensätze geringfügige Abweichungen im Vergleich zu der nationalen Berichterstattung zu TIMSS 2007 bedingen können: Betroffen sind insbesondere die Abbildungen 6.4-6.6 (Bonsen, Frey & Bos, 2008).⁴

⁴ Die Abweichungen, die in Folge ausführlicher Bereinigungs- und Aufbereitungsschritte entstanden sind und 5 Fälle betreffen, begründen sich durch zweierlei: a) es konnte ein

Die Daten zur Befragung der Lehrkräfte sowie Schulleitungen sind in gesonderten Dateien zu finden. Die Daten für Lehrkräfte wurden disaggregiert und mit den von ihnen unterrichteten Schülerinnen und Schülern kombiniert.

2.4 Identifikationsvariablen

Um die Daten aus den unterschiedlichen Befragungs- und Testinstrumenten miteinander zu verknüpfen und gemeinsam zu analysieren, gibt es eine Reihe von Identifikationsvariablen (ID-Variablen). Diese ordnen beispielsweise Schülerinnen und Schüler einer Klasse zu. In der folgenden Übersicht sind die fünf wichtigsten ID-Variablen aufgeführt. Weitere Schlüssel- und Identifikationsvariablen sind in der Dokumentation der Erhebungsinstrumente zu TIMSS 2007 gelistet (Bos et al., 2009).

Tabelle 5: Identifikationsvariablen für die Zuordnung und Verknüpfung der Befragten in den Datensätzen

Identifikationsvariable	Beschreibung
IDSTUD	Identifikationsvariable für Schülerinnen und Schüler
IDCLASS	Identifikationsvariable für Klassen
IDSCHOOL	Identifikationsvariable für Schulen
IDTEACH	Identifikationsvariable für Lehrerinnen und Lehrer
IDLINK	Lehrer-Link

Die ID-Variablen werden für das Matchen, Aggregieren und Disaggregieren der Daten benötigt. So können beispielsweise die Schulleitungs- und Schülerangaben über die IDSCHOOL, die in beiden Datensätzen vorliegt, verknüpft werden. Die *International Database* (Foy & Olson, 2009), die online verfügbar ist, enthält neben den internationalen Einzeldatensätzen aus TIMSS 2007 für jeden Teilnehmer auch jeweils eine Linkdatei, die es ermöglicht Schüler- und Lehrerdatensätze (mit Hilfe des IEA *International Database* [IDB] Analyzer) zu kombinieren. Für das Zusammenspielen von Datensätzen mit dem IDB-Analyzer ist es wichtig zu beachten, dass die Datensätze in ihren Originalbezeichnungen genutzt werden.

Bei den Datensätzen, die über das FDZ bereitgestellt werden, handelt es sich, wie oben beschrieben, um kombinierte Datensätze, in denen die Informationen aus verschiedenen Datensätzen verknüpft vorliegen und/oder die Informationen für alle internationalen Teilnehmer in einem Datensatz

zusätzlicher Elternfragebogen zugeordnet werden und b) in Fällen in denen für Schülerinnen und Schüler zwei Elternfragebögen vorlagen (dies konnte auftreten, wenn ein fremdsprachiger und zusätzlich ein deutschsprachiger Elternfragebogen ausgefüllt wurde) wurde vom ursprünglichen Vorgehen abweichend ein Verfahren genutzt, um Informationen aus beiden Fragebögen nutzbar zu machen.

enthalten sind.

Für die kombinierten Lehrer-Schülerdatensätze gilt es zu beachten, dass nicht alle Schülerinnen und Schüler genau einer Lehrerin oder einem Lehrer zugeordnet werden können, da einige Klassen von mehreren Lehrkräften unterrichtet werden. Um diese Komponente des Studiendesigns angemessen bei der Analyse der Daten berücksichtigen zu können, sind die GewichtungsvARIABLEN (hier: Lehrergewichte), die weiter unten beschrieben werden, zu nutzen. Da im Rahmen von TIMSS sowohl Mathematik- als auch Sachunterrichtslehrkräfte befragt wurden, gibt es über entsprechende Gewichtungen der Daten zudem die Möglichkeit fachspezifische Analysen vorzunehmen.

Die kombinierten Lehrer-Schülerdatensätze liegen bereits in zwei verschiedenen Versionen vor, so dass auf die Subbereiche bezogene Analysen entweder analog zur Berichterstattung zu TIMSS 2011 oder TIMSS 2007 vorgenommen werden können. Für Trendanalysen ist jedoch in jedem Fall mit den 2011 ermittelten Variablen zu arbeiten. Um die Leistungswerte zum Bericht zu TIMSS 2007 für den Schüler- oder Leistungsdatensatz zu nutzen, sind a) die enthaltenen *Plausible Values* der Subdomänen zu entfernen oder umzubenennen und b) die *Plausible Values* der Subdomänen aus dem „Report07“-Datensatz über die IDSTUD anzuspielen.

2.5 Fehlende Werte

Fehlende Werte werden in den Datensätzen in der Regel mit den Codes 7, 8 und 9 gekennzeichnet. Allerdings ist es bei einigen Variablen möglich, dass einstellige Zahlen valide, interpretierbare Antworten darstellen. In diesen Fällen werden die Codes für fehlende Werte aus mehreren Ziffern gebildet, indem vor die Codes eine oder mehrere Neunen gesetzt werden (z.B. 97, 98, 99, 999, 9999).

Die verschiedenen Codes bieten Informationen über die Gründe des Fehlens. Die Bedeutung der einzelnen Codes ist allerdings nicht über alle Instrumente und Items hinweg identisch, jedoch in der Regel im Datensatz genauer spezifiziert. Generell ist darauf zu achten, dass auch andere Werte auf fehlende Werte hinweisen, so werden auch systembedingt fehlende Werte (sysmis) in den SPSS-Datensätzen verwendet.

Während die einzelnen Items in der Regel einen gewissen Anteil fehlender Werte haben, weisen die Leistungsscores keine fehlenden Werte auf, da Schülerinnen oder Schüler, die nicht am Test teilgenommen haben (z.B. aus Krankheitsgründen), nicht in dem Datensatz enthalten sind.

Für Items aus Frageblöcken, in denen lediglich zwischen den beiden Antwortmöglichkeiten „Ja“ und „Nein“ unterschieden wird, lässt sich häufig eine hohe Anzahl an fehlenden Werten feststellen. Eine Aufbereitung dieser Variablen in dem Sinne, dass fehlende Werte bei bestimmten Antwortmustern als „Nein“ umkodiert werden, wurde unterlassen, um einer Fehlinterpretation vorzubeugen. Die Interpretation der fehlenden Werte sowie ggf. entsprechende Aufbereitungsschritte obliegen dem jeweiligen Anwender.

2.6 Zusätzliche Items

Die Dokumentation der Erhebungsinstrumente zu TIMSS 2007 enthält alle Fragen, die Schülerinnen und Schülern, Eltern, Lehrerinnen und Lehrern sowie Schulleitungen in den Fragebögen aus TIMSS 2007 vorgelegt wurden. Neben diesen Variablen enthalten die Datensätze eine Reihe weiterer Variablen, die nicht direkt Teil der Datenerhebung waren. Dabei handelt es sich beispielsweise um Designvariablen, wie die Identifikations-, Gewichtung- und *Jackknife*-Variablen. Zudem wurden in den Schülerteilnahmelisten Informationen wie das Alter oder Noten erhoben. Des Weiteren wurde eine Reihe von Indizes und neuen Variablen gebildet, welche die Befragungsangaben aus mehreren Items zusammenfassen.

Die folgenden Tabellen listen die wichtigsten zusätzlichen Items auf. Variablen, die in der Regel für Analysen zu substantziellen Forschungsfragen ungeeignet sind (z.B. Variablen zur Berechnung von Gewichten), werden nicht angeführt, sondern können im *Technical Report* und dem *User Guide* zu TIMSS 2007 nachgeschlagen werden.

Tabelle 6: Nicht in der Dokumentation der Erhebungsinstrumente aufgeführte Items – Schüler-Elterndatensatz

Variable	Beschreibung
SFB2101	Aussagen zum Fach Deutsch: mit meinen Leistungen zufrieden
SFB2102	Aussagen zum Fach Deutsch: muss mehr lernen als andere
SFB2103	Aussagen zum Fach Deutsch: verstehe das meiste
SFB2104	Aussagen zum Fach Deutsch: kann Aufgaben meistens lösen
TRIKind	TR/ Integrationskind
EF29	Ganztagsangebote

Tabelle 7: Nicht im Datensatz vorliegende Items – Schulleitungsdatensatz

Variable	Beschreibung
AC4GFTVY	Erhebungsinstrument wurde in Deutschland nicht eingesetzt
AC4GBONS	Erhebungsinstrument wurde in Deutschland nicht eingesetzt
AC4GDE5C	Diese String-Variable zu AC4GDSOI wurde zwar in der Dokumentation der Erhebungsinstrumente zu TIMSS 2007 aufgeführt, ist aber nicht im Datensatz enthalten, da für sie keine Angaben gemacht wurden

III Besondere Hinweise für die Datenanalyse

In den folgenden Abschnitten werden in Ergänzung Informationen und Hinweise für den Umgang mit den TIMSS-Daten und ihre Analyse gegeben. Hier wird die Verwendung von GewichtungsvARIABLEN erläutert, auf die Besonderheiten bezüglich der Berechnung von Standardfehlern hingewiesen sowie der Umgang mit Plausible Values und Benchmark-VARIABLEN beschrieben.

3.1 Verwendung der GewichtungsvARIABLEN

In TIMSS wurde eine sogenannte disproportionale, stratifizierte Klumpenstichprobe gezogen (Olson et al., 2008). Anders als bei einer einfachen Zufallsstichprobe hatten die Schülerinnen und Schüler bei der Stichprobenziehung in TIMSS nicht dieselbe Ziehungswahrscheinlichkeit – dasselbe gilt für deren Eltern, Lehrerinnen, Lehrer und Schulleitungen –, das heißt, die Stichprobe ist nicht proportional zur Grundgesamtheit aller Viertklässlerinnen und Viertklässler in Deutschland gezogen worden. Beispielsweise haben Schülerinnen und Schüler aus Schulen mit wenigen Jahrgangsklassen eine größere Wahrscheinlichkeit in die Stichprobe zu gelangen als Kinder aus Schulen mit vielen Jahrgangsklassen. Zudem kann auch der Stichprobenausfall eine Verzerrung bedeuten, da Ausfälle einen Zusammenhang mit studienrelevanten Personenmerkmalen aufweisen können. Vertiefende Informationen zum Stichprobendesign und zur korrekten Analyse finden sich im *Technical Report* und *User Guide* zu TIMSS 2007 (Foy & Olson, 2009; Olson et al., 2008).

Ohne Berücksichtigung des komplexen Stichprobendesigns führen Analysen mit TIMSS-Daten zu verzerrten Ergebnissen. Für die Analysen sind daher die im Datensatz enthaltenen GewichtungsvARIABLEN zu nutzen. Bei den meisten statistischen Analyseprogrammen kann die entsprechende Gewichtung einfach vorgenommen werden. Um den unterschiedlichen Umgang der Statistikprogramme mit der Fallzahl bei der Berechnung der Standardfehler

berücksichtigen zu können, ist neben dem *Total Student Weight* (TOTWGT: Summe der Gewichte entspricht dem Populationsumfang) auch das *Student House Weight* (HOUWGT: Summe der Gewichte entspricht dem Stichprobenumfang) in den entsprechenden Datensätzen zu finden.

In Tabelle 8 sind die Gewichtungsvariablen sortiert nach Datensätzen aufgeführt. In der Tabelle ist zudem ersichtlich, für welche Untersuchungspopulation repräsentative Aussagen auf Grundlage von gemachten Angaben getroffen werden können. Zusätzlich ist für jede Gewichtung angegeben, auf welche Größe sich die Summe der einzelnen Gewichte bezieht.

Für die Analyse der Lehrerdaten ist als Besonderheit Folgendes zu beachten: Da es sich bei TIMSS um eine nach Schülerinnen und Schülern sowie Schulen stratifizierte Stichprobe handelt, werden Aussagen für Lehrkräfte nach dem Schema „X Prozent der Schülerinnen und Schüler haben Lehrkräfte, die angeben...“ getroffen.

Tabelle 8: Die Gewichtungswvariablen im Überblick

Datensatz	Variable	Repräsentative Aussagen für die...	...zu Angaben von...	...mit Fallzahl gewichtet auf
Schülerdatensatz, (Schüler-Elterndatensatz), Leistungsdatensatz	HOUWGT (Student House Weight)	Schülerpopulation	Schülerinnen und Schülern, (Eltern)	Stichprobenebene
	TOTWGT (Total Student Weight)	Schülerpopulation	Schülerinnen und Schülern, (Eltern)	Populations-ebene
	SENWGT (Senate Weight)	Schülerpopulation	Schülerinnen und Schülern	eine int. Konstante
Lehrer-Schülerdatensatz	TCHWGT (Overall Teacher Weight)	Schülerpopulation	Lehrkräften	Populations-ebene
	MATWGT (Mathematics Teacher Weight)	Schülerpopulation	Mathematik-lehrkräften	Populations-ebene
	SCIWGT (Science Teacher Weight)	Schülerpopulation	Sachunter-richtslehrkräften	Populations-ebene
Schulleitungsdatensatz	SCHWGT (School Weight)	Schulpopulation	Schulleitungen	Populations-ebene
	STOTWGTU (Sum of Student Weights)	Schülerpopulation	Schulleitungen	Populations-ebene

3.2 Berechnung von Standardfehlern

In TIMSS wurde eine Cluster- bzw. Klumpenstichprobe gezogen, das heißt, die Ziehung der Personen ist nicht unabhängig, sondern es werden gesamte Klassen befragt und getestet. Die wichtigsten Gründe hierfür sind die Analysierbarkeit von Klassen- und Kompositionseffekten sowie die vereinfachte Testadministration im Klassenkontext (Bonsen et al., 2008). Dabei gilt es zu beachten, dass bei Klumpenstichproben in der Regel die Standardfehler unterschätzt werden, wodurch Signifikanztests zu optimistisch ausfallen. Des Weiteren handelt es sich bei TIMSS um eine stratifizierte Stichprobe, das heißt, die Grundgesamtheit an Viertklässlerinnen und Viertklässler kann in unterschiedliche Schichten unterteilt werden (z.B. Bundesländer, Schulen, Schulklassen). Die Stratifikation bewirkt, dass Standardfehler in der Regel überschätzt werden, wodurch Signifikanztests zu konservativ ausfallen. Ob der Effekt von Klumpenstichprobe oder Stratifikation überwiegt, kann nicht allgemein entschieden werden, sondern hängt von der konkreten Forschungsfrage ab. Allerdings überwiegt in der Regel der Effekt der Klumpenstichprobe, sodass die Standardfehler trotz Stratifikation unterschätzt werden. Zur korrekten Schätzung finden sich in den Datensätzen Informationen über die Zuweisung der Schülerinnen und Schüler zu den Clustern und Strata. Mithilfe dieser Information kann die sogenannte *Jackknife Repeated Replication Technique* (JRR) angewandt werden, die annähernd unverzerzte Schätzer für den Standardfehler von Häufigkeiten, Mittelwerten, Prozentwerten und anderen Statistiken liefert. In der folgenden Tabelle sind die beiden Variablen aufgeführt, die für die Anwendung des *Jackknife*-Verfahren benötigt werden.

Tabelle 9: Die Variablen zur Anwendung der *Jackknife Repeated Replication Technique*

Variable	Beschreibung
JKZONE	Jackknife Zone
JKREP	Jackknife Replicate Code

Die JRR-Technik ist nicht standardmäßig in SPSS implementiert. Ein benutzerfreundliches, einfach handhabbares *Add-On*-Programm zu SPSS ist der IDB Analyzer. Bei diesem Programm sind für die TIMSS-Daten bereits alle Voreinstellungen zur Anwendung der JRR-Technik vorprogrammiert.⁵ Zudem stehen eine Reihe anderer kostenloser sowie kommerzieller Programme (z.B. AM, Stata, WesVar) zur Anwendung der JRR-Technik zur Verfügung.

⁵ Der IDB Analyzer steht kostenfrei zum Download zur Verfügung: <http://www.iea.nl/data.html>

3.3 Berechnungen mit Plausible Values

Mit der in TIMSS eingesetzten Leistungstestung soll es gelingen, von den in den Testheften gezeigten Schülerantworten auf Kompetenzausprägungen zu schließen. In TIMSS wird für die Bestimmung individueller Leistungswerte auf statistische Modelle zurückgegriffen, die auf der *Item Response Theory* basieren (IRT, Boomsma, van Duijn & Snijders, 2001; van der Linden & Hambleton, 1997). Die IRT geht von der Annahme aus, dass die beobachtbaren Antworten einer Person in einem Test durch eine nicht beobachtbare, testbezogene Fähigkeit erklärbar sind. Die beobachteten Antworten werden verwendet, um die Höhe der nicht beobachtbaren (*latenten*) Fähigkeit zu schätzen. Ein Vorzug der IRT besteht darin, dass sich die Schwierigkeit einer Aufgabe und die Fähigkeit einer Person auf derselben Skala abbilden lassen. In TIMSS wird die IRT-Skalierung nicht zuletzt aufgrund des implementierten Testheftdesigns (*Multi-Matrix-Design*) gewählt. Dieses Untersuchungsdesign sieht vor, dass die einzelnen teilnehmenden Schülerinnen und Schüler immer nur eine kleine Auswahl aller Testaufgaben des TIMSS-Aufgabenpools bearbeiten. Da diese Ausschnitte nicht immer in ihrer Schwierigkeit übereinstimmen, sind auch die darauf basierenden Testleistungen nicht ohne weiteres vergleichbar. Durch die Skalierung wird dies ermöglicht – eine ausreichend große Stichprobe und Überlappungen der Testheftinhalte vorausgesetzt (vgl. z.B. Kolen, 1981; Kolen & Brennan, 2004; Lord, 1980). International wurde für die Modellierung der mathematischen und naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler ein sogenanntes dreiparametrisches logistisches Modell (3PL-Modell) genutzt. Für internationale Vergleichsstudien hat es den Vorteil, dass neben der Schwierigkeit und Trennschärfe auch eine spezifische Ratewahrscheinlichkeit für jede Testaufgabe berücksichtigt wird (Foy, Galia & Li, 2008).⁶

Allerdings hat die IRT auch Grenzen. Die eingeschränkte Aufgabenanzahl pro Testheft sowie Schwierigkeitsunterschiede zwischen den Testheften führen mitunter zu ungenauen Fähigkeitsschätzungen für einzelne Personen (Bos & Voss, 2008). Da TIMSS aber nicht auf Individualdiagnostik abzielt, sondern Aussagen über Gruppen treffen möchte, sind diese Ungenauigkeiten zu vernachlässigen. Um sicherzustellen, dass sich auf der Grundlage individueller Testleistungen akkurate Rückschlüsse auf Populationsmerkmale ziehen lassen, wird in TIMSS zur Ermittlung der Schülerleistung (*Personenparameter*) der *Plausible-Values*-Ansatz gewählt (Mislevy, 1991; Mislevy, Beaton, Kaplan & Sheehan, 1992), ein Verfahren, dem die Theorie der Multiplen Imputation zugrunde liegt (Rubin, 1987). Die Grundidee dieses Ansatzes ist es, die nicht beobachtbare Fähigkeit einer Person als fehlenden Wert zu betrachten, der durch einen ‚plausiblen Wert‘ ersetzt werden kann.

⁶ Zu beachten ist, dass für die Berichtslegung zu TIMSS 2007 zur Ermittlung der Leistungsscores auf den Subskalen separate Skalierungen durchgeführt wurden (Foy et al., 2008). Da in TIMSS 2011 ein neuer Skalierungsansatz verfolgt wurde, der eine mehrdimensionale Modellierung zur Ermittlung der Leistungsscores für die Subskalen (Inhaltsbereiche und kognitive Anforderungsprozesse) beinhaltet, wurden für Trendanalysen die Subskalen aus TIMSS 2007 nach diesem Modell neu skaliert (Foy et al., 2011).

Dazu werden alle verfügbaren und als relevant erachteten Informationen über eine Person herangezogen. Neben der Information über die Testleistung der Person wird dabei eine Vielzahl von Informationen aus den Hintergrundfragebögen berücksichtigt. Aus den Informationen wird eine bedingte Wahrscheinlichkeitsverteilung für die Fähigkeit der Person erstellt. Je mehr Informationen über eine Person vorliegen, desto geringer wird die Streuung dieser Verteilung. *Plausible Values* sind also Fähigkeitswerte, die für jede Testperson zufällig aus diesen auf Grundlage von Hintergrundinformationen bedingten Verteilungen gezogen werden. Durch die Ziehung mehrerer *Plausible Values* wird berücksichtigt, dass die Ersetzung eines fehlenden Werts immer auch mit Unsicherheit behaftet ist. Als Konvention hat sich die Ziehung von fünf Werten etabliert.

Entsprechend liegen pro Schülerin oder Schüler nicht ein, sondern mehrere Leistungsschätzer pro Gesamt- bzw. Subskala (vgl. Tabelle 10) vor, die in den Datensätzen jeweils in fünf *Plausible-Value*-Variablen vorzufinden sind. Für eine korrekte Auswertung müssen Analysen mit den Leistungsdaten immer fünfmal einzeln durchgeführt werden und die Ergebnisse dann zusammengefasst werden. Beispielsweise wird die Korrelation zwischen Mathematikkompetenz und Mathematikmotivation berechnet, indem zunächst die fünf Korrelationskoeffizienten der fünf *Plausible Values* mit der Variable zur Mathematikmotivation berechnet werden. Im zweiten Schritt wird der Mittelwert dieser Korrelationskoeffizienten berechnet. Der Standardfehler hierzu lässt sich in zwei additive Komponenten unterteilen. Der erste Teil ist der mittlere Standardfehler der fünf Korrelationskoeffizienten. Hinzu kommt ein zweiter Teil, der sich aus der Varianz der einzelnen Standardfehler ergibt. Die Verrechnungsvorschriften hierzu finden sich bei Little und Rubin (2002).

Für weniger komplexe Analysen bietet sich auch hier die Nutzung des IDB Analyzers an. Bei diesem Programm ist die Zusammenführung von Analysen mit den unterschiedlichen *Plausible-Value*-Variablen als Standard implementiert. Zudem stehen eine Reihe anderer kostenloser sowie kommerzieller Programme zur Anwendung zur Verfügung.

3.4 Berechnungen mit Benchmarks/Kompetenzstufen

Um eine inhaltliche Interpretation der von den getesteten Schülerinnen und Schülern erreichten Kompetenzwerte zu ermöglichen, werden in TIMSS sogenannte Kompetenzstufen gebildet. Die Entwicklung und Beschreibung der Kompetenzstufen wird von einem international zusammengesetzten Expertengremium, dem *Science and Mathematics Item Review Committee* (SMIRC), unter Leitung der internationalen Studienleitung vorgenommen. Ausgangspunkt der Entwicklung der Kompetenzstufen ist zunächst die Festlegung von zentralen Bezugspunkten, sogenannten Benchmarks, auf den Leistungsskalen (Martin et al., 2008; Mullis, Martin & Foy et al., 2008). Durch die Setzung von vier Benchmarks (400, 475, 550, 625) wird die Leistungsskala in fünf Abschnitte eingeteilt, die in Deutschland als Kompetenzstufen bezeichnet werden (Bos et al., 2008). Die niedrige Benchmark liegt bei 400 Punkten und bildet die Grenze zwischen

Kompetenzstufe I und Kompetenzstufe II. Die durchschnittliche Benchmark liegt bei 475 Punkten und markiert den Beginn von Kompetenzstufe III. Die hohe und die fortgeschrittene Benchmark liegen bei 550 beziehungsweise 625 Punkten. Mit ihrer Überschreitung beginnen Kompetenzstufe IV beziehungsweise Kompetenzstufe V.

Neben den Leistungsscores in Form von fünf *Plausible Values* für die Gesamt- und Subskalen, liegen für die Gesamtskalen zur Mathematik- und Naturwissenschaftskompetenz auch sogenannte Benchmark-Variablen vor (vgl. Tabelle 8). Analysen erfolgen analog zum in Abschnitt 3.3 beschriebenen Verfahren.

Tabelle 10: Die *Plausible-Value-Variablen* zur Messung der mathematischen und naturwissenschaftlichen Kompetenz für die Gesamtskalen und die Subskalen sowie die *Plausible-Value-Variablen* für die Kompetenzstufen (Benchmarks)

Plausible Values/Variablen Mathematik	Beschreibung
ASMMAT01, ASMMAT02, ASMMAT03, ASMMAT04, ASMMAT05	Mathematics (Overall)
ASMIBM01, ASMIBM02, ASMIBM03, ASMIBM04, ASMIBM05	Mathematics (Overall) International Benchmark Reached
ASMDAT01, ASMDAT02, ASMDAT03, ASMDAT04, ASMDAT05	Data Display
ASMGEO01, ASMGEO02, ASMGEO03, ASMGEO04, ASMGEO05	Geometric Shapes & Measures
ASMINUM01, ASMINUM02, ASMINUM03, ASMINUM04, ASMINUM05	Number
ASMKNO01, ASMKNO02, ASMKNO03, ASMKNO04, ASMKNO05	Math Knowing
ASMAPP01, ASMAPP02, ASMAPP03, ASMAPP04, ASMAPP05	Math Applying
ASMREA01, ASMREA02, ASMREA03, ASMREA04, ASMREA05	Math Reasoning
Plausible Values/Variablen Naturwissenschaften	Beschreibung
ASSSCI01, ASSSCI02, ASSSCI03, ASSSCI04, ASSSCI05	Science (Overall)
ASSIBM01, ASSIBM02, ASSIBM03, ASSIBM04, ASSIBM05	Science (Overall) International Benchmark Reached
ASSEAR01, ASSEAR02, ASSEAR03, ASSEAR04, ASSEAR05	Earth Science
ASSLIF01, ASSLIF02, ASSLIF03, ASSLIF04, ASSLIF05	Life Science
ASSPHY01, ASSPHY02, ASSPHY03, ASSPHY04, ASSPHY05	Physical Science
ASSKNO01, ASSKNO02, ASSKNO03, ASSKNO04, ASSKNO05	Science Knowing
ASSAPP01, ASSAPP02, ASSAPP03, ASSAPP04, ASSAPP05	Science Applying
ASSREA01, ASSREA02, ASSREA03, ASSREA04, ASSREA05	Science Reasoning

Literatur

- Bonsen, M., Frey, K. A. & Bos, W. (2008). Soziale Herkunft. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 141–156). Münster: Waxmann.
- Bonsen, M., Lintorf, K., Bos, W. & Frey, K. A. (2008). TIMSS 2007 Grundschule - Eine Einführung in die Studie. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 19–48). Münster: Waxmann.
- Boomsma, A., van Duijn, M. A. J. & Snijders, T. A. B. (Hrsg.). (2001). *Essays on item response theory*. New York: Springer.
- Bos, W. & Voss, A. (2008). Empirische Schulentwicklung auf Grundlage von Lernstandserhebung. *Die Deutsche Schule*, 100 (4), 449–458.
- Bos, W., Bonsen, M., Baumert, J., Prenzel, M., Selter, C. & Walther, G. (Hrsg.). (2008). *TIMSS 2007. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.
- Bos, W., Bonsen, M., Kummer, N., Lintorf, K. & Frey, K. (2009). *TIMSS 2007. Dokumentation der Erhebungsinstrumente zur Trends in International Mathematics and Science Study*. Münster: Waxmann.
- Foy, P. & Olson, J. F. (2009). *TIMSS 2007 International Database and User Guide: TIMSS & PIRLS International Study Center*, Boston College.
- Foy, P., Brossman, B. & Galia, J. (2011). Scaling TIMSS and PIRLS 2011 Achievement Data. In M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS and PIRLS methods and procedures*. Zugriff am 29.04.2013. Verfügbar unter http://timssandpirls.bc.edu/methods/pdf/TP11_Scaling_Achievement.pdf.
- Foy, P., Galia, J. & Li, I. (2008). Scaling the Data from the TIMSS 2007 Mathematics and Science Assessments. In J. F. Olson, M. O. Martin & I. V. S. Mullis (Hrsg.), *TIMSS 2007 Technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision - KFT 4-12+R : Manual*. Göttingen: Beltz Test.
- IQB - Institut zur Qualitätsentwicklung im Bildungswesen. *Überprüfen von Bildungsstandards. Ländervergleich*. Zugriff am 29.04.2013. Verfügbar unter <http://www.iqb.hu-berlin.de/bista/control>.
- Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement*, 18 (1), 1–11.
- Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking. Methods and practices* (2. Aufl.). New York: Springer.
- Little, R. J. A. & Rubin, D. B. (2002). *Statistical analysis with missing data* (2. Aufl.). New York: John Wiley & Sons.

- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Maaz, K., Baumert, J., Gresch, C. & McElvany, N. (2010). *Der Übergang von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Martin, M. O. & Mullis, I. V. S. (2004). Overview of TIMSS 2003. In M. O. Martin, I. V. S. Mullis & S. J. Chrostowski (Hrsg.), *TIMSS 2003. Technical Report* (S. 2–21). Chestnut Hill, Mass: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Foy, P., Olson, J. F., Erberer, E., Preuschoff, C. et al. (2008). *TIMSS 2007. International science report : findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56, 177–196.
- Mislevy, R. J., Beaton, A. E., Kaplan, B. A. & Sheehan, K. M. (1992). Estimation population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29 (2), 133–161.
- Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1998). *Mathematics Achievement in the Primary School Years. IEA's Third International Mathematics and Science Report*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J. F., Preuschoff, C., Erberer, E. et al. (2008). *TIMSS 2007. TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Olson, J. F., Berger, D. R., Milne, D. & Stanco, G. M. (2008). *TIMSS 2007 Encyclopedia. A Guide to Mathematics and Science Education Around the World, Volumes 1 and 2*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A. & Eberber, E. (2005). *TIMSS 2007 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Olson, J. F., Martin, M. O. & Mullis, I. V. S. (Hrsg.). (2008). *TIMSS 2007 Technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley + Sons.
- TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. (2007). *TIMSS 2007 Contextual Questionnaires*. Zugriff am 29.04.2013. Verfügbar unter <http://timss.bc.edu/timss2007/context.html>.
- van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.

Wendt, H., Tarelli, I., Bos, W., Frey, K. & Vennemann, M. (2012). Ziele, Anlage und Durchführung der Trends in International Mathematics and Science Study (TIMSS 2011). In W. Bos, H. Wendt, O. Köller & C. Selzer (Hrsg.), *TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 27–68). Münster: Waxmann.