

# Datensatzdokumentation zu PISA-E 2003 Deutschland

## Inhaltsverzeichnis

1	Einleitung .....	2
2	Skalierung der PISA-E 2003 Daten .....	3
2.1	<i>Modell 1 Länder in 4 Dimensionen</i> .....	3
2.2	Ergänztes „neues“ Hintergrundmodell (ESCS und Migration).....	3
2.3	<i>Modell 2 Länder in 7D (Mathe Subskalen)</i> .....	4
2.4	<i>Modell 3 Länder in 5 D (NaWi Subskalen)</i> .....	4
2.5	Oversampling nach Migrationshintergrund .....	4
2.6	Postadjustierung der Gewichte nach PISA-I 2003 .....	4
2.7	Sonder- und Förderschulen .....	5
3	Der Datensatz von PISA-E 2003.....	7
3.1	Übersicht über den PISA-E 2003 Datensatz und diese Dokumentation .....	7
4	Leistungsvariablen - Ergänzungen zum Skalenhandbuch.....	9
4.1	Dateienübersicht.....	9
4.2	Gewichte.....	9
4.3	Leistungsskalen .....	10
4.4	Leistungsskalen, neues Hintergrundmodell .....	11
5	Kontextvariablen - Ergänzungen zum Skalenhandbuch .....	13
5.1	Schülerfragebogen.....	13
5.2	Noten .....	16
5.3	Migrationshintergrund.....	16
5.4	Index ESCS .....	17
5.5	Soziale Indikatoren.....	17
6	„Ergänzungen zum Skalenhandbuch“: Stichprobenlisten .....	20
6.1	Varianzstrata.....	20
6.2	Replikationsgewichte EOM und ADJ .....	20
6.3	Schulleiterfragebogen.....	21
6.4	Skalenwerte zum Schulleiterfragebogen .....	25
7	Kognitive Items .....	27

## 1 Einleitung

Die vorliegende Datendokumentation erläutert den Datensatz zur PISA-E 2003 Studie des PISA Konsortiums Deutschland. Der internationale Vergleich wurde um einen Vergleich zwischen den Ländern der Bundesrepublik in Deutschland erweitert.- Für diese Studie wurde die Stichprobe um ein Mehrfaches vergrößert und weiter Schülerrinnen und Schülern mit Migrationshintergrund besonders berücksichtigt. Der Vergleich der Länder Deutschlands basiert fünfzehnjährigen Schülerinnen und Schülern, die am internationalen PISA-Testprogramm teilnehmen. Detaillierte Informationen zu dieser Erweiterungsstudie sind nachzulesen bei Prenzel et al. (2005). PISA 2003 Der zweite Vergleich der Länder in Deutschland – Was wissen und können und Jugendliche? Münster: Waxmann Verlag GmbH.

Die vorliegende Dokumentation des PISA-E 2003 Datensatzes wird dem Forschungsdatenzentrum des Instituts für Qualitätsentwicklung (IQB) vom Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN) zur Verfügung gestellt, um Sekundäranalysen durch Wissenschaftlerinnen und Wissenschaftler ausserhalb des PISA Konsortium Deutschlands zu unterstützen.

Das PISA 2003 Data Analysis Manual (OECD, 2005) gibt einen Überblick über die Datenbank der PISA 2003 Erhebung. Dabei vermittelt es sowohl Informationen über die Variablen der 4 Kompetenzbereiche – Mathematik, Lesen, Naturwissenschaften und Problemlösen – als auch über den Schülerfragebogen und den Schulleiterfragebogen. Zur Ergänzung des Manuals kann der Technical Report (OECD, 2005) herangezogen werden. Es erläutert die verwendeten Methoden. Zur Beschreibung der Daten wird jedoch in erster Linie auf das Skalenhandbuch (Ramm et al., 2006) Bezug genommen, welches die messtheoretischen Eigenschaften der verwendeten Aufgaben und Fragebogenitems dargestellt (Im Folgenden wird das Data Analysis Manual mit DAM und das Skalenhandbuch mit SHB abgekürzt).

In dieser Dokumentation wird zunächst auf die Skalierung der Kompetenzdaten zu PISA-E 2003 eingegangen (Kapitel 2). Die Aufbereitung der Kontextdaten ist dem Skalenhandbuch zu entnehmen. In den Kapiteln 3 bis 7 werden die Teile des Datensatzes näher erläutert. Diese Dokumentation kann nicht den Anspruch erheben, den Umgang mit den PISA-Daten allgemein zu erläutern, hier wird auf die Technischen Berichte der OECD zu PISA sowie auf die Einführungen in Technische Grundlagen in den Berichten des PISA Konsortiums Deutschland verwiesen.

Für den Fall, dass sich Fragen zum Datensatz nicht anhand der Literatur zu PISA und mit der vorliegenden Dokumentation beantworten lassen, steht das PISA 2003 Konsortium Deutschland gerne für weiterführende Auskünfte zur Verfügung.

## 2 Skalierung der PISA-E 2003 Daten

Im folgenden Kapitel wird die Skalierung der Kompetenzskalen zu PISA-E 2003 Deutschland skizziert. Die Analyse von PISA-Daten erfolgt üblicherweise anhand von so genannten Plausible Values, die als multiple Imputationen aus den der gemeinsamen Verteilung der untersuchten Kompetenzen und aller zu relevanten Kontextmerkmalen zufällig gezogen werden. Soweit die zu analysierenden Variablen in die Skalierung einbezogen wurden, erlauben die Plausible Values es, latente Zusammenhänge zwischen Kompetenzskalen und kontextmerkmalen zu analysieren. Die Kenntnis der verwendeten Skalierungsmodelle und der zugrundegeriebenen Hintergrundmodelle ist daher notwendig, um die Plausible Values für die Kompetenzskalen vor dem zugrunde liegenden Skalierungsmodell angemessen für Analysen verwenden zu können.

### 2.1 Modell 1 Länder in 4 Dimensionen

In PISA werden im Allgemeinen messfehlerbereinigte, latente Korrelationen berichtet. Bei der Berechnung von Personenmesswerten müssen dazu alle Variablen, die mit den Kompetenzen der Fünfzehnjährigen in Verbindung gebracht werden sollen, mit einbezogen werden. Das Hintergrundmodell für PISA-E enthält diese Informationen. In den vom internationalen Konsortium bereitgestellten Daten sind dies alle Information aus den Kontextfragebögen in Form von Einzelantworten oder Skalenwerten. Dazu kommen Kompetenzmittelwerte auf Schulebene, um mögliche Einflüsse auf die Kompetenzen in der Schulebene abzubilden (vgl. OECD 2005).

Ergänzend dazu wurden für den Ländervergleich Antworten auf weitere Fragen mit aufgenommen, die in der deutschen Version der Kontextfragebögen zusätzlich vor-gegeben wurden. Eine zweite Ergänzung besteht in zusätzlichen Variablen, die Interaktionen zwischen einigen ausgewählten, zentralen Schülermerkmalen in das Hintergrundmodell einbringen. Das betrifft die Interaktion zwischen Bildungsgang, Migrationsstatus und Geschlecht sowie Bildungsgang, Migrationsstatus und ESCS einschließlich der zugehörigen Zweifach-Interaktionen. Der Vorteil liegt darin, dass in Darstellungen von Kompetenzergebnissen, die nach diesen Merkmalen aufgeschlüsselt sind, latente Gruppenmittelwerte berichtet werden können.

Als Kompetenzdimensionen wurden in Modell 1 die vier Kompetenzskalen Mathematik, Lesen, Naturwissenschaften und Problemlösen aus den internationalen Testinstrumenten verwendet. Insgesamt ist das Modell 1 damit mit dem Skalierungsmodell für die von der OECD zur Verfügung gestellten Kompetenzwerte vergleichbar, im Unterschied dazu wurde das Hintergrundmodell für PISA-E ergänzt.

Die Plausible Values für dieses Modell sind in der Datei „P2003-EOM\_03\_Leistungsskalen\_v05.sav“ gespeichert.

### 2.2 Ergänztes „neues“ Hintergrundmodell (ESCS und Migration)

Das Hintergrundmodell für PISA-E 2003 wurde nach der Bereitstellung eines Datensatzes vom Plausible Values und nach einer Vorabpublikation der Ländermittelwerte ergänzt bzw. geändert und es wurde ein zweiter Satz Plausible Values für das Modell 1 erstellt. Zum einen wurde die Berechnung des ESCS bei der Behandlung im Ausland erworbenen Schulabschlüssen verbessert. Zum anderen wurden in der Variable zur Herkunft mehr Fälle zugeordnet (durch „Handarbeit“) und eine Variable zu Akkulturationsgruppen aufgenommen. Dieser zweite Satz Plausible Values ist in der Datei „P2003-EOM\_Leistung\_neues\_HG\_v11.sav“ gespeichert.

Es können sich in bestimmten Analysen geringe Unterschiede zwischen den beiden Datensätzen zeigen. Dem Bericht zum Ländervergleich liegen im Allgemeinen die Plausible Values aus dem ersten Datensatz zugrunde; Analysen mit den veränderten

Kontextvariablen müssen jedoch mit Plausible Values nach dem „neuen“ Hintergrundmodell vorgenommen werden.

### **2.3 Modell 2 Länder in 7D (Mathe Subskalen)**

In einem weiteren Modell wurden die Subskalen für Mathematik skaliert. Insgesamt wurden dabei sieben Dimensionen geschätzt, nämlich für die Kompetenzbereiche Lesen, Naturwissenschaften, Problemlösen und die vier Subskalen in der Mathematik. Als Hintergrundmodell wurde das ursprüngliche Modell aus Modell 1 verwendet; die Plausible Values wurden nicht mit dem „neuen“ Hintergrundmodell neu erzeugt.

In der „P2003-EOM\_03\_Leistungsskalen\_v05.sav“ sind die Plausible Values für die Subskalen der Mathematik aus diesem Modell gespeichert; die Skalenwerte für die vier Hauptdomänen in dieser Datei stammen aus Modell 1.

### **2.4 Modell 3 Länder in 5 D (NaWi Subskalen)**

In einem dritten Modell wurden die Fächer in den Naturwissenschaften getrennt skaliert. Aufgrund der Zahl von Aufgaben, die sich den Fächern zuordnen lassen, wurde ein Score für Biologie einem Score für Physik und Chemie gemeinsam gegenübergestellt. In diesem Modell 3 wurden fünf Dimensionen geschätzt, für die Kompetenzbereiche Mathematik, Lesen, Problemlösen und die beiden Subskalen in den Naturwissenschaften. Als Hintergrundmodell wurde das ursprüngliche Modell aus Modell 1 verwendet; die Plausible Values wurden nicht mit dem „neuen“ Hintergrundmodell neu erzeugt.

In der „P2003-EOM\_03\_Leistungsskalen\_v05.sav“ sind die Plausible Values für die Subskalen der Mathematik aus diesem Modell gespeichert; die Skalenwerte für die vier Hauptdomänen in dieser Datei stammen aus Modell 1.

## **2.5 Oversampling nach Migrationshintergrund**

Mit dem Oversampling von Schülerinnen und Schülern mit Migrationshintergrund wurde die Stichprobe für den Ländervergleich um eine ausgewählte Gruppe von Jugendlichen erweitert. Um eine Stichprobe zu erhalten, die das Oversampling enthält und gleichzeitig repräsentativ für die Population der Fünfzehnjährigen ist, wurde ein entsprechendes Gewicht berechnet, in dem die Gewichte der Jugendlichen mit Migrationshintergrund so adjustiert werden, dass ihr Anteil an allen Jugendlichen dem Anteil entspricht, der in der Stichprobe aller gezogenen Fünfzehnjährigen zu beobachten ist.

Damit ergeben sich zwei mögliche Stichproben für den Ländervergleich: die gezogenen Fünfzehnjährigen (PISA-E, für Ergänzungsstichprobe) und die gezogenen Fünfzehnjährigen einschließlich der Schülerinnen und Schüler des Oversampling mit angepassten Gewichten (PISA-EOM, E inkl. Oversampling nach Migrationshintergrund). Beide Stichproben sind in gleicher Weise repräsentativ für die Fünfzehnjährigen, jedoch umfasst die PISA-E Stichprobe 38487 Schülerinnen und Schüler und die PISA-EOM Stichprobe 44580 Fünfzehnjährige. In Vorabanalysen wurde bestätigt, dass beide Stichproben zu praktisch gleichen Kompetenzergebnissen führen. Zur Einheitlichkeit der Darstellung und der Analysen wurde für den Bericht zum Ländervergleich ausschließlich auf die PISA-EOM Stichprobe zurückgegriffen. Im vorliegenden Datensatz sind entsprechend nur die Daten der Stichprobe PISA-EOM enthalten.

## **2.6 Postadjustierung der Gewichte nach PISA-I 2003**

Da die Ergänzungsstichprobe im Vergleich zur Stichprobe des internationalen Vergleichs eine neue Stichprobe darstellt, können sich im Rahmen der Stichprobenfehler unterschiedliche Schätzungen für Populationsparameter ergeben. So unterscheiden sich die Stichprobe für den Ländervergleich PISA-EOM und die Stichprobe für den

internationalen Vergleich PISA-I auch geringfügig in den Anteilen der Jugendlichen nach Schularten, woraus u.a. auch andere Mittelwerte in den Kompetenzen resultieren. Um in der Darstellung der Ergebnisse von PISA-EOM an die Ergebnisse des internationalen Vergleichs anschließen zu können, wurden die Gewichte für die Ergänzungsstichprobe nachträglich so adjustiert, dass die Anteile von Jugendlichen nach Schularten genau denen aus der Gewichtung für die internationale Stichprobe entsprechen. In der folgenden Tabelle (Tabelle 14.4 aus dem Bericht zum Ländervergleich) sind die Schülerzahlen und die Anteile in Prozent in den Schulformen nach den Gewichten für die PISA-EOM Stichprobe und nach den postadjustierten Gewichten angegeben. Insbesondere können für Berufsschulen Unterschiede beobachtet werden.

Tabelle: Anzahl und Anteil von Schülerinnen und Schülern in Schulformen, nach originalen und adjustierten Gewichten berechnet

Schulform	Schülerinnen und Schüler der Population (PISA EOM)	Anteile in Prozent	
		Populationsgewicht WP_03EOM	postadjustiertes Populationsgewicht WP_03ADJ
<b>Hauptschule</b>	186150	20.5	19.6
<b>Schule mit mehreren Bildungsgängen</b>	97514	10.7	11.6
<b>Realschule</b>	202270	22.3	23.3
<b>Integrierte Gesamtschule</b>	86942	9.6	10.0
<b>Gymnasium</b>	259202	28.6	29.1
<b>Berufsschule</b>	43597	4.8	2.7
<b>Förder- und Sonderschule</b>	31508	3.5	3.6
<b>Gesamt</b>	907182	100	100

Für alle Analysen im Bericht zum Ländervergleich wurden zur Berechnung von Kompetenzkennwerten die postadjustierten Gewichte (wp\_03adj) verwendet, während Angaben zu Populationsgrößen auf den nicht adjustierten Gewichten (wp\_03eom) beruhen. So wurden zum einen die Kompetenzwerte konsistent mit denen bereits im Bericht zum internationalen Vergleich veröffentlichten Ergebnissen berichtet. Zum anderen basieren berichtete Populationsschätzungen für Gruppengrößen auf der viel größeren Stichprobe PISA-EOM und sind entsprechend genauer.

## 2.7 Sonder- und Förderschulen

Die Testdurchführung in Sonder- und Förderschulen stellt eine erhebliche Belastung der zu untersuchenden Schülerinnen und Schüler dar. Daher wurden nur vergleichsweise wenig Fünfzehnjährige aus Sonder- und Förderschulen getestet. So wurden in acht Ländern Deutschlands im Rahmen der internationalen Stichprobe zehn Sonder- und Förderschulen mit insgesamt 107 Schülerinnen und Schülern untersucht. Für den Ländervergleich sind damit etwa 7 Prozent einer populationsrepräsentativen Stichprobe erreicht, für den internationalen Vergleich etwa 60 Prozent. Die Repräsentativität der Sonder- und Förderschulen wird durch die verwendete Gewichtung hergestellt.

In PISA-E 2003 werden die Ergebnisse dieser Schülerinnen und Schüler verwendet, um den Stichproben in jedem Land einen angemessenen Anteil von Sonder- und Förderschulen hinzuzufügen. Das heißt, dass mit der Annahme gearbeitet wurde, die Ergebnisse der Sonder- und Förderschulen unterscheiden sich nicht zwischen den Ländern Deutschlands. Die Ergebnisse derselben 107 untersuchten Sonder- und Förder-

schülerinnen und -schüler gehen also entsprechend den Anteilen von Sonder- und Förderschulen in den sechzehn Ländern in deren Kompetenz-Ergebnisse ein. In den Leistungsdaten sind Zeilen für diese Einbeziehung von Sonderschülern enthalten. Sind kenntlich durch Einträge in der Schüler-ID (idstud), die größer als 7000000000 sind. Diese Werte stellen nicht die Ergebnisse von unterschiedlichen Jugendlichen dar, vielmehr entsprechen je 16 dieser Zeilen einem Jugendlichen in Sonder- und Förderschulen.

Für die Testdurchführung in diesen Schulen wurde ein verkürztes Testheft eingesetzt, welches die leichtesten Aufgaben des PISA-Tests enthielt. Ein Fragebogen wurde nicht vorgegeben. Daher lassen sich Kompetenzergebnisse auf eine Stichprobe beziehen, welche die Sonder- und Förderschulen einschließt, während Analysen, die Angaben aus den Fragebögen verwenden, diese Schulen hingegen ausschließt.

### 3 Der Datensatz von PISA-E 2003

In den folgenden Kapiteln wird der Daten zum Ländervergleich von PISA 2003 in seinen Bestandteilen vorgestellt und erläutert. Die Gliederung der Darstellung folgt der Gliederung des Datensatzes.

#### 3.1 Übersicht über den PISA-E 2003 Datensatz und diese Dokumentation

Der Datensatz zu PISA-E 2003 ist in 18 Datendateien unterteilt, die im Sinne von Modulen zu verstehen sind. Auf Schülerebene sind 16 Dateien enthalten, die zu einer Gesamtdatei verbunden werden können. Ebenso können die beiden auf Schulebene enthaltenen Dateien zu einer gemeinsamen Datei verbunden werden.

Die Dateien zum PISA-E 2003 Datensatz sind im Ordner *Daten\_eom* gespeichert. Im Ordner *Instrumente* sind die Instrumente, die für die Erweiterungsstudie verwendet wurden, abgelegt. Einen Überblick die Instrumente, Testbereiche und Teilstichproben gibt das SHB auf S. 23. Im Ordner *Anhang* finden sich die Manuale für die Testleiter und Schulkoordinatoren. Im Ordner *docs* werden zusätzliche Dokumente zur Erläuterung des Datensatzes zur Verfügung gestellt.

Im Folgenden werden die Dateien des PISA-E 2003 Datensatzes (im Ordner daten\_eom) aufgelistet und anschließend kurz beschrieben.

- (1) [P2003-EOM\\_01\\_match\\_files\\_v04.sps](#)
- (2) [P2003-EOM\\_02\\_gewichte\\_v04.sav](#)
- (3) [P2003-EOM\\_03\\_Leistungsskalen\\_v05.sav](#)
- (4) [P2003-EOM\\_04\\_Leistung\\_neues\\_HG\\_v11.sav](#)
- (5) [P2003-EOM\\_05\\_Leistung\\_neues\\_HG\\_länderskala\\_v02.sav](#)
- (6) [P2003-EOM\\_06\\_SFB\\_v05.sav](#)
- (7) [P2003-EOM\\_07\\_SFB\\_im\\_HGM\\_v05.sav](#)
- (8) [P2003-EOM\\_08\\_Noten.sav](#)
- (9) [P2003-EOM\\_09\\_Migrationshintergrund.sav](#)
- (10) [P2003-EOM\\_10\\_escs\\_v03.sav](#)
- (11) [P2003-EOM\\_11\\_escs-ipn.sav](#)
- (12) [P2003-EOM\\_12\\_soziale\\_Indikatoren\\_erweitert\\_v08.sav](#)
- (13) [P2003-EOM\\_13\\_varstrata\\_bs2\\_so3.sav](#)
- (14) [P2003-EOM\\_14\\_replicates\\_eom.sav](#)
- (15) [P2003-EOM\\_15\\_replicates\\_adj.sav](#)
- (16) [P2003-EOM\\_16\\_SLFB\\_items\\_Int\\_Nat\\_v01.sav](#)
- (17) [P2003-EOM\\_17\\_SLFB\\_skalen\\_Int\\_Nat\\_v01.sav](#)
- (18) P2003\_EOM\_18\_items\_conq.zip

In der Syntaxdatei (1) "P2003-EOM\_01\_match\_files\_v04.sps" können die Datendateien auf Schülerebene und auf Schulebene miteinander verbunden werden.

Die Datei (2) „P2003-EOM\_02\_gewichte\_v04.sav“ enthält auf Schülerebene die beiden Populationsgewichten für den Ländervergleich und weitere zentrale Schülermerkmale aus der Schülerteilnahmelisten.

Die Datei (3) „P2003-EOM\_03\_Leistungsskalen\_v05.sav“ beinhaltet die Plausible Values (PVs) zu den Kompetenzbereichen Mathematik, Lesen, Naturwissenschaften und Problemlösen.

Die Datei (4) „P2003-EOM\_04\_Leistung\_neues\_HG\_v11.sav“ gibt ebenfalls Plausible Values wieder, die auf Basis des „neuen“ (vgl. Abschnitt 2.2) Hintergrundmodells erstellt wurden.

Die Datei (5) „P2003-EOM\_05\_Leistung\_neues\_HG\_länderskala\_v02.sav“ enthält länderspezifische Plausible Values auf Basis des neuen Hintergrundmodells.

Die Datei (6) „P2003-EOM\_06\_SFB\_v05.sav“ enthält die mit dem Schülerfragebogen erhobenen Variablen.

Die Datei (7) „P2003-EOM\_07\_SFB\_im\_HGM\_v05.sav“ enthält die Auswahl von Variablen aus dem Schülerfragebogen, die im Hintergrundmodell der Kompetenzskalierung verwendet wurden.

Die Datei (8) „P2003-EOM\_08\_Noten.sav“ beinhaltet aufbereitete Schulnoten für die Fächer Deutsch, Mathematik, Biologie, Physik und Chemie. Die Rekodierung erfolgte nach Land und Schulform - teils sogar nach Schule - in eine „1 bis 6“-Schulnotenskala.

Die Datei (9) „P2003-EOM\_09\_Migrationshintergrund.sav“ beinhaltet die für den Bericht zum Ländervergleich verwendeten Variablen zum Migrationsstatus.

Die beiden Dateien (10) „P2003-EOM\_10\_escs\_v03.sav“ und (11) „P2003-EOM\_11\_escs-ipn.sav“ berichten zwei Varianten des Index ESCS. Die erste Variante entspricht der Bildung des ESCS laut OECD, die zweite eine Überarbeitung, die am IPN für den Ländervergleich vorgenommen wurde.

Die Datei (12) „P2003-EOM\_12\_soziale\_Indikatoren\_erweitert\_v09.sav“ enthält gebildete Variablen zur sozialen Herkunft im Ländervergleich (vgl. Ehmke, T., Siegle, T., Hohensee, F. (2005). Soziale Herkunft im Ländervergleich. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* (S. 235-268). Münster: Waxmann.)

Die Datei (13) „P2003-EOM\_13\_varstrata\_bs2\_so3.sav“ enthält vier Versionen von Zuordnungen von primary sampling units zu Schulpaaren und von Schulpaaren zu Straata für eine Varianzschätzung nach dem BRR-Verfahren.

Die Datei (14) „P2003-EOM\_14\_replicates\_eom.sav“ beinhaltet replicate weights die auf Grundlage der Originalgewichte wp\_03eom beruhen. Die Datei (15) „P2003-EOM\_15\_replicates\_adj.sav“ beinhaltet replicate weights auf Basis der postadjusted Gewichte wp\_03adj.

Die Datei (16) „P2003-EOM\_16\_SLFB\_items\_Int\_Nat\_v01.sav“ enthält die im Schulleiterfragebogen erhobenen Variablen.

In der Datei (17) „P2003-EOM\_17\_SLFB\_skalen\_Int\_Nat\_v01.sav“ sind die Skalenwerte der Skalen aus dem Schulleiterfragebogens in Form von standardisierte WLEs enthalten.

Der Ordner (18) P2003\_EOM\_18\_items\_conq.zip enthält die Antworten auf die kognitiven Items sowie weitere SPSS-Syntax-Dateien, die erläutern sollen, wie die Items aus der SPSS-Datendatei für die Skalierung durch das Programm ConQuest vorbereitet und in einer Textdatei gespeichert wurden.

## 4 Leistungsvariablen - Ergänzungen zum Skalenhandbuch

In diesem Kapitel werden die Dateien mit Variablen zu den Kompetenzen der Jugendlichen näher erläutert.

### 4.1 Dateienübersicht

Mit der Syntaxdatei (1) "P2003-EOM\_01\_match\_files\_v04.sps" können die Dateien auf Schülerebene wie auf Schulebene zu jeweils einem zusammenhängendem Datensatz verbunden werden. Zu beachten ist, dass es nicht sinnvoll ist, alle Dateien auf Schülerebenen miteinander zu verbinden. Vielmehr gibt es mehrere Dateien, die als Parallelversionen zu betrachten sind, so dass in Regel eine Auswahl der Dateien zusammengefügt werden wird.

### 4.2 Gewichte

Die Datei „P2003-EOM\_02\_gewichte\_v04.sav“ enthält auf Schülerebene die beiden Populationsgewichte für den Ländervergleich. Das Gewicht *wp\_03adj* wurde für die Analyse von Kompetenzdaten berechnet. Dieses Gewicht ermöglicht die konsistenteren Reproduktion der Leistungsergebnisse von PISA-EOM, gemessen an PISA-I. Mit dem Gewicht *wp\_03eom* können die genaueren Schätzer für Populationsverhältnisse verwendet werden. Detaillierte Erläuterungen s.o. und in Kapitel 14 des Berichts zum Ländervergleich (Carstensen, C. H., Knoll, S., Siegle, T., Rost, J. & Prenzel, M. (2005). Technische Grundlagen des Ländervergleichs. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche* (S. 386-401). Münster: Waxmann).

Im Unterschied zu PISA 2000 gibt es bei PISA 2003 nur ein Schülergewicht, welches für alle vier Leistungsdomänen verwendet wird. Das gilt auch für PISA-EOM. Bei allen so weit genannten Gewichten handelt es sich um **Populationsgewichte**, d.h. die Summe der Gewichte reproduziert die Populationszahlen. Stichprobengewichte werden von PISA international nicht verwendet und wurden auch nicht für PISA-EOM hergestellt.

Als weitere zentrale Schülermerkmale sind Klassenstufe, Geschlecht und Geburtstag aus der Schülerteilnahmeliste und Variablen für Land und Schulart enthalten. Die Variable *sfxbl* enthält die Zuordnung zu Schulartern, anhand derer im Kapitel mit Länderergebnissen im Bericht zum Ländervergleich berichtet wurde. Alle Variablen dieser Datei sind der Tabelle 6.1 aufgelistet.

Tab. 6.1 Variablen in der Datei mit den Gewichten

Nr.	PISA-E	SHB
1	idstud	vollständige Schüler-ID
5	wp_03adj	postadjustiertes Populationsgewicht PISA EOM 2003
6	wp_03eom	Populationsgewicht PISA EOM 2003
9	trc1q06m	Tracking Variable Geburtsmonat
10	trc1q06y	Tracking Variable Geburtsjahr
11	trc2q07	Tracking Variable Study Programme
12	land	Variable, die die 16 Bundesländer kodiert

14	sfxbl	besondere Schulform-Variable; sie kam dadurch zustande, dass in den einzelnen Länderkapiteln des E-Bandes nicht über alle Schulformen berichtet wurde
15	idsch	Schul-ID

#### 4.3 Leistungsskalen

Die Datei (3) „P2003-EOM\_03\_Leistungsskalen\_v05.sav“ beinhaltet die Plausible Values (PVs) zu den Kompetenzbereichen Mathematik, Lesen, Naturwissenschaften und Problemlösen.

Die so genannten Plausible Values stellen keine Personenfähigkeitsparameter im Sinne einer Individualdiagnostik dar; vielmehr handelt es sich um Zufallsziehungen aus der gemeinsamen a posteriori Verteilung aller Informationen, die über eine Schülerin bzw. einen Schüler in der Skalierung berücksichtigt wurden. Zur Berechnung wird die Zahl der richtigen Lösungen, die Kompetenz in den anderen Domänen und die Variablen aus dem Hintergrundmodell zusammengetragen, um eine Wahrscheinlichkeitsverteilung für die Kompetenz jeder Person zu berechnen. Die Plausible Values werden bei PISA verwendet, da mit ihnen varianzabhängige Populationsparameter (Mittelwertunterschiede, Varianzen und Korrelationen) latent geschätzt werden können.

Zu Analysen mit Leistungsskalen werden in PISA ausschließlich die Plausible Values verwendet; jede Analyse wird mit allen 5 Plausible Values für eine Leistungsskala durchgeführt und die Ergebnisse anschließend gemittelt. Zur Verdeutlichung folgen zwei Beispiele für solche Berechnungen, allgemein ist dies in den Technical Reports zu PISA nachzulesen.

##### **Beispiel 1:**

Angenommen, Sie möchten den Mittelwert für Mathematik berechnen. Dazu berechnen Sie erst den Mittelwert für jeden PV in Mathematik und mitteln anschließend diese Mittelwerte. Dies gelingt problemlos, wenn Sie die Mittelwerte in SPSS zuerst über "Descriptives" berechnen und anschließend diese Tabelle kopieren, in Excel einfügen und dort die Mittelwerte per Excel-Befehl "MITTELWERT" berechnen. Das Ergebnis ist der gesuchte Mittelwert.

### Descriptive Statistics

	N	Mini-mum	Maxi-mum	Mean	Std. Devia-tion	Va-riance
Plausible value in math	884358	132,216 9	808,412 5	503,078875	103,1219860	10634,14 4
Plausible value in math	884358	161,427 0	783,486 5	503,098570	102,0695444	10418,19 2
Plausible value in math	884358	177,083 7	785,667 5	502,721923	102,3380352	10473,07 3
Plausible value in math	884358	168,359 6	804,517 8	503,029750	103,0277798	10614,72 3
Plausible value in math	884358	97,4762	823,212 3	502,998543	102,4069074	10487,17 5
Valid N (listwise)	884358					

#### **Beispiel 2:**

Angenommen, Sie möchten die Korrelation zwischen Mathematik und Naturwissenschaften berechnen. Dazu berechnen Sie erst die Korrelationen zwischen den 5 PVs von Mathematik und Naturwissenschaften: PV1 Math mit PV1 Science, PV2 Math mit PV2 Science, usw. Anschließend mitteln Sie diese 5 Korrelationen. Das Ergebnis ist die gesuchte Korrelation.

**Tipp:** Da sich die Tabellen im Ergebnisausdruck von SPSS komplett kopieren lassen, kann man auch ganze Ergebnistabellen fünfmal berechnen lassen und anschließend in Excel „halb-automatisch“ mitteln lassen. Dazu liegt die Excel-Datei „5PVs\_mitteln\_leer.xls“ anbei, die genau diese Funktion erfüllen soll; diese Datei kann/muss an konkrete Analysen angepasst werden.

#### **4.4 Leistungsskalen, neues Hintergrundmodell**

Die Datei (4) „P2003-EOM\_04\_Leistung\_neues\_HG\_v11.sav“ gibt ebenfalls Plausible Values wieder, die auf Basis eines „neuen“ Hintergrundmodells (vgl. Abschnitt 2.2) erstellt wurden.

(Text vom Abschnitt 2.2:) Das Hintergrundmodell für PISA-E 2003 wurde nach der Bereitstellung eines Datensatzes vom Plausible Values und nach einer Vorabpublikation der Ländermittelwerte ergänzt bzw. geändert und es wurde ein zweiter Satz Plausible Values für das Modell 1 erstellt. Zum einen wurde die Berechnung des ESCS bei der Behandlung im Ausland erworbenen Schulabschlüssen verbessert. Zum anderen wurden in der Variable zur Herkunft mehr Fälle zugeordnet (durch „Handarbeit“) und eine Variable zu Akkulturationsgruppen aufgenommen. Dieser zweite Satz Plausible Values ist in der Datei „P2003-EOM\_Leistung\_neues\_HG\_v11.sav“ gespeichert.

Es können sich in bestimmten Analysen geringe Unterschiede zwischen den beiden Datensätzen zeigen. Dem Bericht zum Ländervergleich liegen im Allgemeinen die Plausible Values aus dem ersten Datensatz zugrunde; Analysen mit den veränderten Kontextvariablen müssen jedoch mit Plausible Values nach dem „neuen“ Hintergrundmodell vorgenommen werden.

Die Datei (5) „P2003-EOM\_05\_Leistung\_neues\_HG\_länderskala\_v02.sav“ enthält länderspezifische Plausible Values auf Basis des neuen Hintergrundmodells. Pro Schüler und Kompetenzbereich werden wiederum fünf PVs angegeben. Hier wurde zusätzlich eine Lineartransformation der Messwerte durchgeführt, das Resultat ist die Län-

derskala, mit einem Mittelwert von 100 und einer Standardabweichung von 30 in jedem Land. Für den Bericht wurde diese Skalierung nicht verwendet.

## 5 Kontextvariablen - Ergänzungen zum Skalenhandbuch

### 5.1 Schülerfragebogen

Die Datei (6) „P2003-EOM\_06\_SFB\_v05.sav“ enthält die mit dem Schülerfragebogen erhobenen Variablen. Neben den Items sind zu allen gebildeten Skalen standardisierte individuelle Personenparameterschätzwerte (Weighted Likelihood Estimates, WLE) nach Warm (1989) enthalten.

Die Datei (7) „P2003-EOM\_07\_SFB\_im\_HGM\_v05.sav“ enthält die Auswahl von Variablen aus dem Schülerfragebogen, die im Hintergrundmodell der Kompetenzskalierungen verwendet wurden. Während die Datei 6 den vollständigen Variablenatz des Schülerfragebogens enthält, sind in Datei 7 lediglich die im Hintergrundmodell tatsächlich verwendeten Variablen enthalten.

Tab. 7.1: Variablen des Schülerfragebogens

Nr.	PISA-E	SHB	Seite
1	idstud	vollständige Schüler-ID	
2	idsch	Schul-ID	
10	stdyprog	Schulformen in unterschiedlichen Kodierungen an.	135
11	schulart		
12	posta1e	posta1	125
13	posta2e	posta2	126
14	st01q01	Klassenstufe (Grade)	132
15	st01q02	Schulart (Programme)	133
16	st02q02	Im SHB findet sich unter diesem Variablenlabel auf S. 131 das Geburtsjahr und auf S. 208 der Geburtsmonat. In diesem Datensatz beinhaltet diese Variable den Geburtsmonat.	131, 208
17	st02q03	Geburtsjahr	
18	st03q01	st03q01	132
19-30	st04q01d	st04q01d	153
31	st05q01	st05q01	155
32	st06q01	st06q01	155
33	st07q01	st07q01	138
34	st09q01	st09q01	138
35-46	st11q01	st11q01	151
47-54	st12q01	st12q01	152
55-66	st13q01	st13q01	151
67-74	st14q01	st14q01	152
75-77	st15q01e	st15q01d	145

78	immig4e	Migrationsstatus inkl. offene Antworten gecleaned	146
79	akkult2	akkult: Sprachliche Akkulturation	150
80-86	st15q04e	st15q04d	146.148
87-129	st16q01	st16q01d	147
130-145	st17q01	st17q01d	142ff.
146-150	st18q01	st18q01	138, 144
151	st19q01	st19q01	145
152	st20q01	st20q01d	208
153	st21q01	st21q01d	208
154-155	st22q01	st22q01d	208
156	st22q04d	st22q04d	210
157-167	st23q01d	st23q01d	211
168-171	st24q01	st24q01	211f.
172-178	st25q01	st25q01	212
179-183	st26q01	st26q01	159f.
184-189	st27q01	st27q01	160
190	st28q01	st28q01	181
191-196	st29q01	st29q01	203
197-204	st30q01	st30q01	265
205-212	st31q01	st31q01	243
213-222	st32q01	st32q01	244f.
223-228	st33q01	st33q01	203f.
229-242	st34q01	st34q01	277
243-245	st35q01	st35q01	158.162
246	st36q01	st36q01	161
247-256	st37q01	st37q01	184
257-267	st38q01	st38q01	196
268-280	ec01q01	ec01q01	209f.
281-329	ic01q01	ic01q01	97f., 206
330	age	Alter	
331	mmins	MMINS	162
332	tmins	TMINS	159
333	pcmath	pcmath = mmins/tmins	162
334-335	isco_m	Die Variablen isco_m und isco_v zum Beruf der Mutter (m) und des Vaters (v) wurden in vierstellige ISCO-Codes kodiert.	137

336-338	isei_m	Die Berufe der Eltern wurden anhand des ISCO-88 klassifiziert und in eine Rangskala (ISEI) nach Ganzeboom & Treiman (1996) überführt.	137
339-340	isei_v2	Falls der ISEI der Mutter oder des Vaters fehlt, wird dieser fehlende Wert mit den Angaben des jeweils anderen Elternteils "aufgefüllt".	
341	hisei	Höchster ISEI in der Familie	136
342-344	siops_m	Standard Index of Occupational Prestige Scale (SIOPS) (nähere Erläuterungen dazu im vorliegenden Kapitel)	
345-347	isco10m	Angabe der ISCO-Hauptgruppen für Mutter, Vater und Schüler mit 30	137f.
356	schabm	Höchster Schulabschluss in der Familie	
359	berfabv	Höchste Berufsausbildung in der Familie	
362	bildabv	Diese Variablen vereint den höchsten Schulabschluss und die höchste Berufsausbildung.	
365-367	fisced	Diese Variablen vereint den höchsten Schulabschluss und die höchste Berufsausbildung und wird nach ISCED-Angaben kodiert.	151f.
368	pared	pared	136
369	st15q01d	st15q01d	145
370	st16q01d	st16q01d	147
371	escs	Index of Economic, Social and Cultural Status	136
372	anxmat_w	ANXMAT	244
373	atsch_w	ATSCHL	211
374	attcom_w	ATTCOMP	104
375	belong_w	BELONG	160
376	compen_w	COMPENV	141
377	complr_w	COMPLRN	184
378	coopl_r_w	COOPLRN	185
379	cstrat_w	CSTRAT	276
380-382	cultpb_w	CULTPOSS	140
383	discli_w	DISCLIM	180
384	elab_w	ELAB	275
385	hedres_w	HEDRES	136
386	highco_w	HIGHCONF	103
387	instmo_w	INSTMOT	266
388	intcon_w	INTCONF	102
389	intmat_w	INTMAT	265

390	intuse_w	INTUSE	98
391	mathef_w	MATHEFF	243
392	memor_w	MEMOR	274
393	pcself_w	?	
394	pcuse_w	?	
395	progus_w	PRGUSE	99
396	routco_w	ROUTCONF	100
397	scmat_w	SCMAT	245
398	sturel_w	STUREL	159
399	suphom_w	?	
400	suppar_w	?	
401	teachs_w	TEACHSUP	196
402	wealth_w	WEALTH	138

Kommentar [uk1]: Was passiert mit damit? Nicht auffindbar

## 5.2 Noten

Die Datei (8) „P2003-EOM\_08\_Noten.sav“ beinhaltet aufbereitete Schulnoten für die Fächer Deutsch, Mathematik, Biologie, Physik und Chemie (trcnde\_r trcnma\_r trcnbi\_r trcnph\_r rcnch\_r). Die Rekodierung erfolgte nach Land und Schulform - teils auch nach Schule - in eine „1 bis 6“-Schulnotenskala.

Für die Länder Schleswig-Holstein und Mecklenburg-Vorpommern wurde für die Schulform Integrierte Gesamtschule (IG) die Übertragungsskala (trcnde\_ü trcnma\_ü trcnbi\_ü trcnph\_ü trcnch\_ü) gebildet. Diese kommt zum Einsatz, wenn Noten sich nicht in das 6er-System übertragen lassen. Weiter sind Kursvariablen für Deutsch, Mathematik, Biologie, Physik und Chemie (kurs\_de kurs\_ma kurs\_bi kurs\_ph kurs\_ch) angegeben, die anhand von Notenangaben gebildet wurden. Die Kodierungen A, B, C, E, F, G, H, Z, indizieren, dass die berichteten Noten aus einem Kurssystem stammen, während „N“ bedeutet, dass die berichteten Noten auf einem Notensystem basieren. Details zur Aufbereitung der Noten finden sich bei Kienzle (2005).

## 5.3 Migrationshintergrund

Die Datei (9) „P2003-EOM\_09\_Migrationshintergrund.sav“ beinhaltet die für den Bericht zum Ländervergleich verwendeten Variablen zum Migrationsstatus.

Tab. 7.3: Variablen zum Migrationshintergrund

Nr.	PISA-E	SHB	Seite
1	idstud		
2-5	st15q01e	st15q01d	145
6	migkid	Diese Variable wird mit 1 kodiert, wenn die Mutter (st15q02 S. 145 SHB), der Vater aus dem Ausland kommt (st15q02 S. 145 SHB) oder zu Hause überwiegend nicht deutsch gesprochen wird (st16q01 S.147 SHB).	145ff.

7	nichtmig	Ohne Migrationshintergrund	
8	immig4e		146
9	akkult3	Akkulturationsgruppen: Jugendliche ohne Migrationshintergrund weiter differenziert	150
10	akkult4	Akkulturationsgruppen: Jugendliche ohne Migrationshintergrund nicht weiter differenziert	150
12	m_herke	Variable st15q02e wurde rekodiert, bezeichnet Herkunft Mutter	145
13	v_herke	Variable st15q03e wurde rekodiert, bezeichnet Herkunft Vater	145
14	herke	Herkunft Eltern; wenn Vater und Mutter beide aus gleichem Land, dann herke=0	
15	herkunft	Herkunft der Eltern	
16	s_vweil	Kategorisierung der Variable st15q04 in 4 Kategorien; dies ist Einreisevariable	
17-51	st16q01	st16q01	147-150

#### 5.4 Index ESCS

Die beiden Dateien (10) „P2003-EOM\_10\_escs\_v03.sav“ und (11) „P2003-EOM\_11\_escs-ipn.sav“ berichten zwei Varianten des Index ESCS. Die erste Variante entspricht der Bildung des ESCS laut OECD, die zweite eine Überarbeitung, die am IPN für den Ländervergleich vorgenommen wurde (geringfügige Änderung in der Rekodierungssyntax der PARED-Variable).

Ziel des ESCS-Indexes ist es, unterschiedliche Merkmale der sozialen Herkunft in einem Indikator zusammenzufassen. Der ESCS-Index wurde in PISA 2003 aus drei Indikatoren durch eine Hauptkomponentenanalyse abgeleitet: dem höchsten Bildungsgrad der Eltern (umgerechnet in Ausbildungsjahre, PARED), dem höchsten sozioökonomischen Status der Eltern (HISEI) und einem Index für häusliche und lernrelevante Besitztümer in der Familie (HOMEPOS).

#### 5.5 Soziale Indikatoren

Die Datei (12) „P2003-EOM\_12\_soziale\_Indikatoren\_erweitert\_v09.sav“ enthält gebil-dete Variablen zur sozialen Herkunft im Ländervergleich (vgl. Ehmke, T., Siegle, T., Hohensee, F. (2005). Soziale Herkunft im Ländervergleich. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?* (S. 235-268). Münster: Waxmann.)

Tab. 7.5: Variablen zu Sozialen Indikatoren

Nr.	PISA-E	SHB	Seite
-----	--------	-----	-------

15-19	risc1ma	Für die 5 Leistungswerte wird jeweils ermittelt, ob der Schüler zur Risikogruppe gehört (codiert mit 1) oder nicht (0). Als Risiko gelten Werte unter 420.40.	
20-24	best1ma	Für die 5 Leistungswerte wird jeweils ermittelt, ob der Schüler zur Spitzengruppe gehört (codiert mit 1) oder nicht (0). Zur Spitzengruppe gehören Leistungswerte über 668.70.	
25	riscmma	Kodierung der Risikogruppe nach über die 5 PVs gemittelten Kompetenzwerten	
26	bestmma	Spitzengruppe Kodierung der Spitzengruppe nach über die 5 PVs gemittelten Kompetenzwerten	
27	immig2e	Der nationale Migrationsstatus immig4e wird rekodiert in die dichotome Variable immig2e mit (1 und 2=1) und (3 und 4=0).	146
28	escs		136
29-32	escs4l	Mit den Variablen escs4l und escs4e werden die Quartile des ESCS-Index für beide Stichproben angegeben.	
33-38	misei4l	Ebenso wird der ISEI-Index in Quartilen für die Mutter (misi4l, misi4e), den Vater (fisei4l, fisei4e) und dem höchsten ISEI für die Eltern (hisei4l, hsei4e) getrennt für die beiden Stichproben aufgeführt.	137
39-41	mempfull	Diese 3 Variablen geben an, ob die Mutter, der Vater oder beide vollzeitbeschäftigt sind.	
42-44	bildabm	Höchster Bildungsabschluss	
45-50	misced	Die 3 Variablen zum Schul- und Berufsbildungsabschluss der Eltern können im SHB detailliert nachgelesen werden.	151f.
51-52	myred, fyred	Diese Variablen geben die Anzahl der Schuljahre der Eltern wieder.	
53	pared	Mit dieser Variable wurde im internationalen Schülerfragebogen erhoben und gibt den Bildungsgrad der Eltern respektive die Anzahl ihrer Schuljahre wieder.	136
54	booklw	Diese Variable ist dichotom kodiert und gibt an, ob mehr als 100 Bücher im Elternhaus vorhanden sind oder bis zu 100 Büchern. Im internationalen Schülerfragebogen wurde diese Variable diffenenzierter erhoben. Detaillierte Informationen sind dem SHB zu entnehmen.	145
55-56	cultpo4l	Für beide Stichproben wird der Besitz an Kulturgütern erhoben. Weitere Informationen zur skala CULTPOSS sind dem SHB zu entnehmen.	140
57-59	homeposb	Die Variabel HOMEPOSS bezieht sich auf die Anzahl der Besitztümer und den Büchern zu Hause. Auch für diese Variable werden jeweils Quartile für die beiden Stichproben angegeben.	137

60	famstruc	Dem SHB ist detailliert zu entnehmen, welche Angaben in den Index Familienstruktur (FAMSTRUC) eingehen.	153f.
61	famstrue	Mit dieser Variable wird die Variable famstruc durch zusätzliche Informationen erweitert.	
62	nkids	Numerische Angabe der Kinderzahl in der Familie	154
63	nkids4	Einteilung der Kinderzahl in 4 Kategorien.	
64	kigalw	Zur Erläuterung können die Kategorien der Variable st20q01 zum Besuch des Kindergartens nachgelesen werden.	208

## 6 „Ergänzungen zum Skalenhandbuch“: Stichprobenlisten

### 6.1 Varianzstrata

Die Datei (13) „P2003-EOM\_13\_varstrata\_bs2\_so3.sav“ enthält vier Versionen von Zuordnungen von primary sampling units zu Schulpaaren und von Schulpaaren zu Stra-  
ta für eine Varianzschätzung nach dem BRR-Verfahren.

Zur Reduzierung der Zahl von so genannten Varianzstrata werden diese zusammen-  
gelegt. Die Variablen sind als „pseustrt“ (800 Varianzstrata), „pseu400“ (400 Vari-  
anzstrata), „pseu200“ (200 Varianzstrata) oder „pseu100“ (100 Varianzstrata) be-  
nannt. Im Bericht sind die Varianzstrata in der Variable „pseu200“ verwendet worden,  
die primary sampling unit sind in der Variablen PSU enthalten.

### 6.2 Replikationsgewichte EOM und ADJ

Die Datei (14) „P2003-EOM\_14\_replicates\_eom.sav“ beinhaltet replicate weights die auf Grundlage der Originalgewichte wp\_03eom beruhen. Die Datei (15) „P2003-  
EOM\_15\_replicates\_adj.sav“ beinhaltet replicate weights auf Basis der postadjustierte  
Gewichte wp\_03adj.

In PISA wird eine stratifizierte Klumpenstichprobe erhoben, d.h. es werden zunächst Schulen aus definierten Strata gezogen und anschließend werden Schüler aus diesen Schulen zufällig gezogen. Die Schüler sind damit nicht unabhängig voneinander gezogen und die sogenannte *effektive Stichprobengröße* (sinngemäß die Zahl der unabhängigen Beobachtungen) ist kleiner als die Zahl der untersuchten Schüler. Damit gelten die Formeln zur Berechnung von Standardschätzfehlern in einfachen Zufalls-  
stichproben (wie in SPSS verwendet) nicht. Mit so genannten Replikationsmethoden lassen sich Standardfehler errechnen. Dafür werden die Replikationsgewichte verwendet.

Um die Größenordnungen von Standardschätzfehlern näherungsweise berechnen zu können, können die Designeffekte verwendet werden. Verwendet man ein Stich-  
probengewicht (-> Gewichte), erhält man Standardfehler, die nicht die Homogenität der Schüler in den Schulen berücksichtigen, sondern auf der Annahme beruhen,  
dass die Schüler unabhängige Beobachtungen darstellen. Berücksichtigt man den Grad der Abhängigkeiten, ergeben sich kleinere „effektive“ Stichproben. Korrigiert man das N in der Berechnung der Standardschätzfehler um den Designeffekt, erhält man näherungsweise korrekte Standardschätzfehler. Der Designeffekt kann in Sub-  
gruppen variieren. Daher ist nur die Berechnung mit geeigneten Verfahren wie z.B.  
der Software WesVar zuverlässig.

In der folgenden Tabelle sind Designeffekte für die Gesamtstichprobe PISA-I 2003 in Deutschland angegeben. Die Verwendung bei PISA-EOM kann nur den Charakter einer vorläufigen Abschätzung der Standardschätzfehler haben.

Skala	DEFF
Mathe int.	5.0
Mathe 1 int.	4.1
Mathe 2 int.	5.0
Mathe 3 int.	5.3
Mathe 4 int.	4.5

Reading int.	2.9
Science int.	2.8
Problem Solving int.	3.1

Die Standardschätzfehler, die man mit SPSS unter Verwendung eines Stichproben-gewichtes berechnet hat, müssen mit  $\sqrt{DEFF}$  multipliziert werden, um die nähe-rungsweise korrekten Standardschätzfehler zu erhalten.

### 6.3 Schulleiterfragebogen

Die Datei (16) "P2003-EOM\_16\_SLFB\_items\_Int\_Nat\_v01.sav" enthält die im Schulleiter-fragebogen erhobenen Variablen.

Tab. 8.3: Skalen zum Schulleiterfragebogen

Nr.	PISA-E	SHB	Seite
12	absen01r	absen01	304
15	agmather	agmath	322
16	agmaseitr	agmseit	322
17	alterslr	alters	284
18	anmerkur	anmerku	Kap. 10
19	aufsichr	aufsich	338
20	delevISR	delevis	320
21	deputatr	deputat	286
22-26	eltern1r	eltern1	346
27	eltfachr	eltfach	347
28	eltfoerr	eltfoer	347
29	eltfreir	eltfrei	347
30	elthausr	elthaus	347
31-34	eltinf1r	eltinf	346
35	eltlehrr	eltehr	347
36	eltprogr	eltprog	347
37	eltunter	eltunte	347
38	erscheir	erschei	338
39-45	eva0001r	eva0001	335
46-49	fach001r	fach001	286
50	fehlendr	fehlend	338
51	fkallefr	fkallef	322
52	fkanderr	fkander	322
53	fkbiolor	fkbiolo	323
54	fkchemir	fkchemi	322

55	fkinforr	fkinfor	322
56	fkmather	fkmathe	322
57	fkmfreqr	fkmfreq	322
58	fkphysir	fkphysi	322
59	fkzieler	fkziele	322
60	foeinfor	foeinfo	327
61	foemathr	foemath	327
62	foescier	foescie	327
63	freqvisr	frequvis	320
64	geschler	gschle	285
65	geskonfr	geskonf	322
66	ggnkorrr	ggnkorr	331
67-74	human01r	human01	289
75-79	impko01r	impko01	345
80-89	inko001r	inko001	344
90-95	instr01r	instr01	287
96-104	jg05matr	jg05mat	286
105-110	kons001r	kons001	343
111	lbbeaufr	lbbeauf	284
112	lbelterr	lbelter	284
113	lbfkonfr	lbfkonf	284
114	lblehrer	lblehre	284
115	lbschuer	lbschue	284
116	lbsleitr	lbsleit	284
117-121	leist01r	leist01	284
122-131	monito01r	monit01	333
132-139	polit01r	polit01	291
140	qualifir	qualifi	285
141	quatimer	quatime	285
142	reflvisr	reflvis	320
143	refzielr	refziel	320
144	rulegemr	rulegem	336
145-151	rules01r	rules01	336
152	sanktior	sanktio	338
153	sauberer	saubere	338
154-156	seit001r	seit001	285
157	seprojier	seproje	328
158-159	setime1r	setime1	328

160	setitelr	setiteld	328
161-162	spbel01r	spbel01	318
163	spbrosr	spbrosr	318
164-180	spcnt01r	spcnt01	315
181	spdiensr	spdiens	318
182	spelterr	spelter	318
183	spimpulr	spimpul	318
184	spinterr	spinter	318
185	sppaperr	sppaper	318
186-194	sppers1r	sppers1	318
195	sppublir	sppubli	318
196	sprograr	sprogra	314
197-200	spsl001r	spsl001	319
201-203	spunt01r	spunt01	318
204	standarr	standar	331
205	stdtestr	stdtest	331
206-213	struc01r	struc01	288
214-221	symb001r	symb001	292
222-226	time01r	time01	303
227	trefextr	trefext	322
228	trefmatr	trefmat	322
229-232	umwelt1r	umwelt1	297
233	vergleir	verglei	331
234-238	vertr01r	vertr01	305
239	visanger	visange	320
240-245	visit01r	visit01	321
246	wetander	wetande	327
247	wetmathr	wetmath	327
248	wetscier	wetscie	327
249	zusinfor	zusinfo	327
250	zusmathr	zusmath	327
251	zusscier	zusscie	327
277	instru_m	INSTRUC	287
278	frstr_m	FR_STRUC	288
279	frhum_m	FR_HUMAN	289
280	frpol_m	FR_POLIT	290
281	frsym_m	FR_SYMB	292
284	spunt_m	SPCONFD	314

285	sporg_m	SPCONTOE	315
286	sppaed_m	SPCONTPA	317
288	fachko_m	ANWESKF	323
289	monito_m	MONITOR	333
290	selev_m	SELFEVA	335
292	dizpl_m	DISZIPLN	338
294	eltinf_m	ELTINFO	346
299	instru_n	INSTRUC	287
300	frstr_n	FR_STRUC	288
301	frhum_n	FR_HUMAN	289
302	frpol_n	FR_POLIT	290
303	frsym_n	FR_SYMB	292
306	spunt_n	SPCONFD	314
307	sporg_n	SPCONTOE	315
308	sppaed_n	SPCONTPA	317
310	fachko_n	ANWESKF	323
311	monito_n	MONITOR	333
312	selev_n	SELFEVA	335
314	dizpl_n	DISZIPLN	338
316	eltinf_n	ELTINFO	346
322	schsize	SCHLSIZE	282
323	pcgirls	PCGIRLS	282
324	ratcomp	RATCOMP	296
330	stratio	TSTRATIO	299
331	sc01r01	SC01Q01	280
332-333	sc02r01	SC02Q01	281
334	sc03r01	SC03Q01	280
335-338	sc04r01	SC04Q01	280
339-352	sc05r01d	SC05Q01	282
353	sc06q01	SC06Q01D	283
354	sc06q02	SC06Q02D	283
355	sc06r05d	SC06Q05D	281
356-365	sc06r06d	SC06Q06D	282
366-368	sc07r01d	SC07Q01D	302
369-388	sc08r01d	SC08Q01D	299
389-394	sc09r01	SC09Q01	296
395-400	sc10r01	SC10Q01	309
401-407	sc11r01	SC11Q01	339

408-412	sc12r01	sc12q01	328
413-420	sc13r01	sc13q01	329
421	sc14r01	sc14q01	306
422-423	sc15r01	sc15q01	326
424-427	sc16r01	sc16q01	324
428-432	sc17r01	sc17q01	324
433-435	sc18r11	sc18q01	298
436-438	sc18r21	sc18q21	298
451-472	sc19r11	sc19q11	300
473-475	sc20r01	sc20q01	330
477-479	sc21r01	sc21q01	341
480-482	sc22r01	sc22q01	341
483-485	sc23r01	sc23q01	342
486-489	sc24r01	sc24q01	340
490-502	sc25r01	sc25q01	307f.
503-562	sc26r01d	sc26q01d	310ff.
563-590	sc27r01d	sc27q01d	313f.
591-599	sc28r01d	sc28q01d	325
600-601	sc29r01d	sc29q01d	306
602-618	sc30r01d	sc30q01d	326f.

#### 6.4 Skalenwerte zum Schulleiterfragebogen

In der Datei (17) „P2003-EOM\_17\_SLFB\_skalen\_Int\_Nat\_v01.sav“ sind die Skalenwerte der Skalen aus dem Schulleiterfragebogens in Form von standardisierten WLEs enthalten.

Für die Variablen der Datei (17) „P2003-EOM\_17\_SLFB\_skalen\_Int\_Nat\_v01.sav“ des Schulleiterfragebogens werden der Personenparameter (WLE) in z-standardisierter (...\_w) Form sowie die Zahl der bearbeiteten Items (...\_n) angegeben. Die Skalennamen in Datei 17 sind teilweise nicht identisch mit den Skalennamen im PISA 2003 Skalenhandbuch; Skalen mit abweichenden Namen sind folgend aufgelistet.

Tab. 8.4: Variablen des Schulleiterfragebogens

Nr.	PISA-E	SHB	Seite
2	instru	INSTRUC	286
4	frstr	FR_STRUC	288
6	frhum	FR_HUMAN	289
8	frpol	FR_POLIT	290
10	frsym	FR_SYMB	292
16	spunt	SPCONTFD	314
18	sporg	SPCONTOE	315

20	sppaed	SPCONTPA	317
24	fachko	ANWESKF	323
26	monito	MONITOR	333
28	selev	SELFEVA	335
32	dizpl	DISZIPLN	338
36	eltinf	ELTINFO	346
40	speff	spimpul, spunkt01, spunkt02	318
42	foerd	zusmath, zusscier, zusinfor, wetmath, wetscie, wetande	327
44	tccoop	inko002, inko004, inko004, inko500, inko600, inko700, inko008, inko900, inko010	344
46	scinfr	SCMATBUI	295
48	sceduc	SCMATEDU	294
52	tcmora	TCMORALE	340
54	stmora	STMORALE	339
56	tcbeha	TEACBEHA	307f.
58	stbeha	STUDBEHA	306f.
60	tccons	TCHCONS	342
62	eval	sc12q01, sc12q03, sc12q04, sc12q05	328
64	ueval	sc13q04, sc13q05, sc13q06, sc13q07, sc13q08	329

## 7 Kognitive Items

Der Ordner (18) P2003\_EOM\_18\_items\_conq.zip enthält die Antworten auf die kognitiven Items sowie ausgewählte SPSS-Syntax-Dateien, die erläutern sollen, wie die Items aus der SPSS-Datendatei zur Skalierung mit der Software ConQuest vorbereitet und in einer Textdatei gespeichert wurden.

Die Itemdaten sind in der Datei EOM\_items\_conq.sav enthalten. Das Format entspricht der Form, in der die Itemdaten durch das ACER weitergegeben werden, als zusammengefasste, aber nicht bewertete Itemvariablen. Zur Bewertung (scoring) muss ein entsprechender ConQuest-Befehl verwendet werden. Die weiteren Dateien im .zip Archiv enthalten einen solchen ConQuest-Job und weitere Hilfsdateien dazu, sowie SPSS-Syntax zum Exportieren der Itemdaten. Die Dateien stammen aus der Skalierung der PISA 2003-E Daten, geben aber keinen Überblick über alle zur Skalierung erforderlichen Schritte und Daten.