

Hinweise zu den FDZ-Daten für die Erhebung PISA 2012 – Fünfte Erhebungsrunde

1. Allgemeines

Bereitgestellt werden hier die Daten aus der Erhebung der fünften Erhebungsrunde des „**Programme for International Student Assessment**“ (**PISA**) im Jahr 2012. Die im Einzelnen bereitgestellten Daten beziehen sich lediglich auf die beiden deutschen Teilstichproben, welche im Rahmen des internationalen Programms gezogen wurden. Bei den fünfzehnjährigen Schülerinnen und Schülern (PISA-Sample) handelt es sich um eine repräsentative Stichprobe für die deutsche Population. Die Stichprobe der Schülerinnen und Schüler der neunten Klassen besteht aus jeweils zwei vollständigen Klassen der für das „PISA-Sample“ gezogenen Schulen. Darüber hinaus werden Daten für die Lehrer, Eltern und Schulleiter für den Erhebungszeitraum 2012 bereitgestellt.

2. Bereitgestellte Datensätze

Im Einzelnen werden die folgenden Datensätze im SPSS-Format über das Forschungsdatenzentrum (**FDZ**) zur Verfügung gestellt:

- 01_Schülerdaten_PBA_pisa_Sample.sav
- 02_Schülerdaten_PBA_grade_Sample.sav
- 03_Elterndaten_PISA_2012.sav
- 04_Lehrerdaten_PISA_2012.sav
- 05_Schulleiterdaten_PISA_2012.sav

Neben so genannten „ID-Variablen“ und Konstrukt bezogenen Variablen enthalten die bereitgestellten Datensätze auch einige Variablen zur Gewichtung und zur Durchführung der Replikationsverfahren zur korrekten Schätzung von Standardfehlern. Die Notwendigkeit der Verwendung dieser Variablen ergibt sich im weitesten Sinne aus dem Umstand, dass es sich bei den hier gezogenen Stichproben nicht um einfache Zufallsstichproben handelt. Für eine vertiefende Darstellung der Stichprobenziehung sei einerseits auf das technische Kapitel im deutschen Berichtsband zu PISA 2012 verwiesen (Heine, Sälzer, Borchert, Sibberns & Mang, 2013) und andererseits, in Bezug auf die Datenauswertungsprozeduren, auf das „Data-Analysis“ Handbuch (OECD, 2009) sowie den technischen Report (OECD, 2014).

3. Codebücher und Kodierung der Variablen

Ergänzend zu den oben gelisteten Datensätzen werden jeweils dazugehörige Codebücher als Excel-Dokument bereitgestellt.

An dieser Stelle wird ausdrücklich darauf hingewiesen, dass es sich bei den vorliegenden Datensätzen um „Rohdatensätze“ handelt – in dem Sinne, dass die einzelnen Variablen „so wie erhoben“ zur Verfügung gestellt werden.

Dies bedeutet, dass sich der Anwender vor jeglichen Auswertungen anhand der Codebücher und durch Inspektion der Daten über das Werteschema und die Kodierung der jeweils interessierenden Variablen Klarheit verschaffen muss. So sind beispielsweise nicht für alle Variablen im Datensatz bereits allgemeingültige „missing-value“-Definitionen im Rahmen des „SPSS-Werteschemas“ vorweggenommen. In den bereitgestellten Datensätzen sind daher auch SPSS-Wertelabels (soweit vergeben) in englischer wie auch in deutscher Sprache enthalten – dies spiegelt im Wesentlichen die „Herkunft“ der Variablen im Rahmen der internationalen oder aber rein nationalen Erhebung (Fragen, die den internationalen PISA-Fragebögen hinzugefügt wurden) wieder.

4. Beschreibung der Inhalte und Variablengruppen der einzelnen Datensätze

4.1. Der Datensatz „01_Schülerdaten_PBA_pisa_Sample.sav“.

Dieser Datensatz bezieht sich auf die deutsche „PISA-Stichprobe“ der fünfzehnjährigen Schülerinnen und Schüler ($n = 5001$). Der Datensatz lässt sich über die Variablen CNT, OECD, SCHOOLID und StIDStd mit den unter <http://www.oecd.org/pisa/pisaproducts/> frei verfügbaren internationalen Daten der Erhebung im Jahr 2012 verbinden.

Neben diesen vier „ID“-Variablen sind folgende Variablen (Gruppen) enthalten:

- 4.1.1. Sämtliche Variablen aus der *internationalen* Erhebung sowie die *nationalen Ergänzungsfragen*, wie sie im Schülerfragebogen in Deutschland erhoben wurden (SPSS-Variablenansicht: Zeilen 8 - 510).
- 4.1.2. Vier Variablen zum Taschenrechner-Gebrauch und zum Anstrengungs-Barometer (vgl. Butler & Adams, 2007; Kunter, Schümer, Artelt, Baumert, Klieme, Neubrand, et al., 2002); (SPSS-Variablenansicht: Zeilen 511 - 514).
- 4.1.3. Drei Variablen zur Formatversion des jeweils verwendeten Schülerfragebogen und des Testheftes für die drei PISA Kompetenzen - *Mathematik, Lesen* und *Naturwissenschaften*; (SPSS-Variablenansicht: Zeilen 515 - 517).
- 4.1.4. Verschiedene, im Rahmen der internationalen Auswertungsvorgaben abgeleitete Skalen aus den Einzelvariablen des Schülerfragebogens. Für eine detaillierte Beschreibung der angewendeten Skalierungsprozeduren und der jeweiligen Inhalte der Skalen sei hier auf die OECD-Literatur zu PISA 2012 verwiesen (OECD 2013; OECD, 2014); (SPSS-Variablenansicht: Zeilen 518 - 608).
- 4.1.5. Plausible Values für die drei Kompetenzbereiche - *Mathematik, Lesen* und *Naturwissenschaften* – (für Mathematik mit Subskalen) (SPSS-Variablenansicht: Zeilen 609 - 658)
- 4.1.6. Das finale Fallgewicht (vgl. OECD, 2014) (SPSS-Variablenansicht: Zeilen 659).
- 4.1.7. 80 Replikationsgewichte; (SPSS-Variablenansicht: Zeilen 660 - 739).
- 4.1.8. Sowie drei weitere Variablen zur Gewichtung und Varianzschätzung, welche im Rahmen der Internationalen Auswertung verwendet wurden; (SPSS-Variablenansicht: Zeilen 740 - 742).

4.2. Der Datensatz „02_Schülerdaten_PBA_grade_Sample.sav“.

Dieser Datensatz bezieht sich auf die klassenbasierte Ergänzungsstichprobe derjenigen Schülerinnen und Schüler in Deutschland, die sich zum Zeitpunkt der Erhebung in einer 9. Klasse befunden haben ($n = 9998$). Da dieser Datensatz keine internationale Entsprechung hat, lassen sich bei diesem Datensatz die vier Variablen CNT, OECD, SCHOOLID und StIDStd **nicht** zum Verbinden mit anderen Datensätzen verwenden. Die Beschreibung der anderen Variablen (Gruppen) entspricht der unter Abschnitt 4.1 aufgeführten Beschreibung.

4.3. Der Datensatz „03_Elterndaten_PISA_2012.sav“.

Dieser Datensatz beinhaltet die Aussagen der Eltern der Schülerinnen und Schüler, welche an der PISA 2012-Erhebung teilgenommen haben und welche ggf. einer solchen Befragung zugestimmt haben (die Teilnahme an der Elternbefragung war für die Eltern freiwillig; $n = 2885$). In einigen Bundesländern durften die Schülerinnen und Schüler den Schülerfragebogen nur dann ausfüllen, wenn eine schriftliche Einverständniserklärung der Erziehungsberechtigten vorlag. Die Teilnahme am Leistungstest war bundesweit verpflichtend. Dieser Datensatz lässt sich über die Variablen CNT, OECD, SCHOOLID und StIDStd mit den unter <http://www.oecd.org/pisa/pisaproducts/> frei verfügbaren internationalen Daten der Eltern, der Schülerinnen und Schüler sowie über die Variable SCHOOLID mit den Daten der Schulleiter verbinden.

Neben den „ID“-Variablen sind folgende Variablen (Gruppen) enthalten:

- 4.3.1. Sämtliche Variablen aus der *internationalen* Erhebung sowie die *nationalen Ergänzungsfragen*, wie sie im Elternfragebogen in Deutschland erhoben wurden (SPSS-Variablenansicht: Zeilen 8 - 167).
- 4.3.2. Verschiedene, im Rahmen der internationalen Auswertungsvorgaben abgeleitete Skalen aus den Einzelvariablen des Elternfragebogens. Für eine detaillierte Beschreibung der angewendeten Skalierungsprozeduren und der jeweiligen Inhalte der Skalen sei hier auf die OECD-Literatur zu PISA 2012 verwiesen (OECD 2013; OECD, 2014);
(SPSS-Variablenansicht: Zeilen 168 - 195).
- 4.3.3. Das sog. Senat-Gewicht der Eltern (vgl. OECD, 2014)
(SPSS-Variablenansicht: Zeile 196).
- 4.3.4. Das Datum der Erstellung des Datensatzes durch das internationale PISA Konsortium 2012
(SPSS-Variablenansicht: Zeile 197).

4.4. Der Datensatz „04_Lehrerdaten_PISA_2012.sav“.

Dieser Datensatz bezieht sich auf die nationale Erhebung der Mathematik-Lehrkräfte in PISA 2012 ($n = 2084$). Alle Mathematik-Lehrkräfte an PISA-Schulen, die eine Lehrbefähigung für die Jahrgangsstufe 9 hatten, erhielten einen Lehrerfragebogen. Da für die Lehrkräfte keine Stichprobe gezogen wurde, erfolgte vorab auch keine Listung aller in Frage kommenden Lehrkräfte. Eine Gewichtung bzw. Angaben darüber, welcher Anteil aller möglichen Mathematik-Lehrkräfte an der Befragung teilnahm, sind daher nicht möglich. Da dieser Datensatz keine internationale Entsprechung hat, lässt sich dieser Datensatz **nicht** mit anderen internationalen Datensätzen verbinden. Es besteht allerdings die Möglichkeit diesen Datensatz mit den deutschen Angaben der Schulleiter über die Variable SCHOOLID zu verbinden.

Darüber hinaus entspricht die Kodierung der Variablen SCHOOLID im Datensatz ,04_Lehrerdaten_PISA_2012.sav` der Kodierung der Variablen mit derselben Bezeichnung im Datensatz ,01_Schülerdaten_PBA_pisa_Sample.`.

Auf diese Weise lassen sich Schüler, Lehrer und Schulleiter welche sich an einer Schule befinden, identifizieren.

Der Datensatz enthält ferner folgende Variablen (Gruppen):

4.4.1. Sämtliche Variablen aus der *nationalen* Erhebung, wie sie im Lehrerfragebogen in Deutschland erhoben wurden

(SPSS-Variablenansicht: Zeilen 5 - 267).

4.4.2. Ausführliche sowie kurze Bezeichnung der Klassenzugehörigkeit der Lehrerin bzw. des Lehrer, falls diese Angaben vorhanden waren

(SPSS-Variablenansicht: Zeilen 268-269).

4.5. Der Datensatz „05_Schulleiterdaten_PISA_2012.sav“.

Dieser Datensatz beinhaltet die Aussagen der Schulleiterinnen und Schulleiter der 230 befragten Schulen der PISA 2012 Stichprobe ($n = 230$). Analog zum Datensatz der Schülerinnen und Schüler lässt sich auch dieser Datensatz über die Variablen CNT, OECD und SCHOOLID mit den unter <http://www.oecd.org/pisa/pisaproducts/> frei verfügbaren internationalen Daten der Schulleiter, Schülerinnen und Schüler sowie der Eltern verbinden. Neben diesen drei „ID“-Variablen sind folgende Variablen (Gruppen) enthalten:

- 4.5.1. Sämtliche Variablen aus der *internationalen* Erhebung sowie die *nationalen Ergänzungsfragen*, wie sie im Schulleiterfragebogen in Deutschland erhoben wurden (SPSS-Variablenansicht: Zeilen 8 - 461).
- 4.5.2. Verschiedene, im Rahmen der internationalen Auswertungsvorgaben abgeleitete Skalen aus den Einzelvariablen des Schulleiterfragebogens. Für eine detaillierte Beschreibung der angewendeten Skalierungsprozeduren und der jeweiligen Inhalte der Skalen sei hier auf die OECD-Literatur zu PISA 2012 verwiesen (OECD 2013; OECD, 2014); (SPSS-Variablenansicht: Zeilen 462 - 494).
- 4.5.3. Das finale Schul-Fallgewicht sowie das sog. Senat-Gewicht der Schulleiter (vgl. OECD, 2014) (SPSS-Variablenansicht: Zeilen 495-496).
- 4.5.4. Das Datum der Erstellung des Datensatzes durch das internationale PISA Konsortium 2012 (SPSS-Variablenansicht: Zeile 497).

4.6. Verbinden der Datensätze untereinander mit der Datei „00_matching_datei.sav“.

Zusätzlich zu den oben genannten Datensätzen wird der SPSS Datensatz 00_matching_datei.sav bereitgestellt.

Dieser Datensatz beinhaltet die für die deutsche Teilstichprobe der internationalen Erhebung verwendeten Verbindungsvariablen idstud, idsch, idteach sowie weitere Variablen zur Klassenbezeichnung.

Über diese Variablen können die Datensätze der unterschiedlichen Agregatebenen (Schulen / Schüler / Lehrer) miteinander verbunden werden. Aufgrund von (teilweise) fehlenden Angaben zu Klassenbezeichnungen bzw. allgemein fehlenden Werten, können bei einem solchen zusammenfügen der Datensätze nicht immer sämtliche Fälle verbunden werden.

5. Literatur

Butler, J., & Adams, R. J. (2007). The impact of differential investment of student effort on the outcomes of international studies. *Journal of Applied Measurement*, 8(3), 279.

Heine, J.-H., Sälzer, C., Borchert, L., Siberns, H., & Mang, J. (2013). Technische Grundlagen des fünften internationalen Vergleichs. In M. Prenzel, C. Sälzer, E. Klieme, & O. Köller (Hrsg.), *PISA 2012 - Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann.

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., et al. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente*. Berlin: Max-Planck-Inst. für Bildungsforschung.

OECD. (2009). *PISA data analysis manual*. Paris: OECD Publishing.

OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.

OECD. (2014). *PISA 2012 Technical Report*. OECD Publishing. Abgerufen von http://www.oecd-ilibrary.org/education/pisa-2009-technical-report_9789264167872-en