

Hinweise zu den FDZ-Daten für die Erhebung PISA 2018 – Siebte Erhebungsrunde

1. Allgemeines

Bereitgestellt werden hier die Daten aus der Erhebung der siebten Erhebungsrunde des „**Programme for International Student Assessment**“ (**PISA**) im Jahr 2018. 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 Klasse besteht aus einer zufälligen Auswahl von 15 Schülerinnen und Schülern einer 9. Klasse der für das „PISA-Sample“ gezogenen Schulen.

Darüber hinaus werden Daten für die Lehrer, Eltern und Schulleiter für den Erhebungszeitraum 2018 bereitgestellt.

Es wird darauf hingewiesen, dass diese Daten für Analysen auf Bundeslandebene nicht geeignet sind. Eine geeignete Stichprobengröße findet sich dafür in den Erhebungen des Ländervergleiches des Instituts zur Qualitätsentwicklung im Bildungswesen (**IQB**).

2. Bereitgestellte Datensätze

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

- PISA2018_Schuelerdaten_Elterndaten_CBA_PISA_Sample.sav
- PISA2018_Schuelerdaten_Elterndaten_CBA_GRADE_Sample.sav
- PISA2018_Schulleiterdaten_CBA.sav
- PISA2018_Lehrerdaten_CBA.sav
- PISA2018_Schuelerdaten_Timing_CBA_PISA_Sample.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 2018 verwiesen (Mang et al., 2019) und andererseits, in Bezug auf die Datenauswertungsprozeduren, auf das „Data-Analysis“ Handbuch (OECD, 2009) sowie den technischen Report (OECD, in Vorbereitung).

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) wider.

4. Beschreibung der Inhalte und Variablengruppen der einzelnen Datensätze

4.1. Der Datensatz

„PISA2018_Schuelerdaten_Elterndaten_CBA_PISA_Sample.sav“.

Dieser Datensatz bezieht sich auf die deutsche „PISA-Stichprobe“ der fünfzehnjährigen Schülerinnen und Schüler ($n = 5451$). Es gilt ein Embargo für Angaben zur Zusatzdomäne *Global Competence* bis **22.10.2020**, sowie für Angaben zu Cronotypes und zur Selbstregulation bis **31.12.2022**. Der Datensatz lässt sich über die Variablen `CNTRYID` bzw. `CNT`, `CNTSCHID` und `CNTSTUID` mit den unter <http://www.oecd.org/pisa/data/2018database/> frei verfügbaren internationalen Daten der Erhebung im Jahr 2018 verbinden.

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

- 4.1.1. Weitere organisatorische und administrative Variablen
(SPSS-Variablenansicht: Zeilen 5 - 24)
- 4.1.2. Sämtliche Variablen aus der *internationalen* Erhebung sowie die *nationalen Ergänzungsfragen* (plus die nationalen Angaben der Schulart), wie sie im Schülerfragebogen in Deutschland erhoben wurden (SPSS-Variablenansicht: Zeilen 25 - 529).
- 4.1.3. Sämtliche Variablen aus der *internationalen* Erhebung sowie die *nationalen Ergänzungsfragen*, wie sie im Elternfragebogen in Deutschland erhoben wurden (SPSS-Variablenansicht: Zeilen 530 - 719).
- 4.1.4. Verschiedene, im Rahmen der internationalen Auswertungsvorgaben abgeleitete Skalen aus den Einzelvariablen des Schülerfragebogens und des Elternfragebogens. Für eine detaillierte Beschreibung der angewendeten Skalierungsprozeduren und der jeweiligen Inhalte der Skalen sei hier auf die OECD-Literatur zu PISA 2018 verwiesen (OECD 2019; OECD, in Vorbereitung);
(SPSS-Variablenansicht: Zeilen 720 - 831).
- 4.1.5. Das finale Fallgewicht (vgl. OECD, in Vorbereitung)
(SPSS-Variablenansicht: Zeile 832).
- 4.1.6. 80 Replikationsgewichte;
(SPSS-Variablenansicht: Zeilen 833 - 912).
- 4.1.7. Zwei weitere Variablen zur Gewichtung und Varianzschätzung, welche im Rahmen der internationalen Auswertung verwendet wurden;
(SPSS-Variablenansicht: Zeilen 913 - 914).
- 4.1.8. Plausible Values für die drei Kompetenzbereiche - *Mathematik, Lesen und Naturwissenschaften*
(SPSS-Variablenansicht: Zeilen 915 - 944)
- 4.1.9. Plausible Values für die weitere erhobene Domäne *Global Competence* (nicht in Deutschland; vgl. OECD, in Vorbereitung)
(SPSS-Variablenansicht: Zeilen 945 - 954)
- 4.1.10. Plausible Values für die Subskalen der Hauptdomäne *Lesen*
(SPSS-Variablenansicht: Zeilen 955 - 1004)
- 4.1.11. Das sog. Senat-Gewicht der Eltern (vgl. OECD, in Vorbereitung)
(SPSS-Variablenansicht: Zeile 1005).

4.1.12. Das Datum der Erstellung des Datensatzes durch das internationale PISA Konsortium 2018 (SPSS-Variablenansicht: Zeile 1006).

4.2. Der Datensatz

„PISA2018_Schuelerdaten_Elterndaten_CBA_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 ($n = 3567$). Da dieser Datensatz keine internationale Entsprechung hat, kann man diesen Datensatz über die Variablen `CNTRYID` bzw. `CNT`, `CNTSCHID` und `CNTSTUID` **nur** mit den deutschen SchulleiterInnen und LehrerInnen Datensätzen verbinden, ein internationales Pendant der OECD gibt es nicht. Die Beschreibung der anderen Variablen (Gruppen) entspricht der unter Abschnitt 4.1 aufgeführten Beschreibung. Es gilt ein Embargo für Angaben zur Zusatzdomäne *Global Competence* bis **22.10.2020**, sowie für Angaben zu Cronotypes und zur Selbstregulation bis **31.12.2022**.

4.3. Der Datensatz „PISA2018_Lehrerdaten_CBA.sav“.

Dieser Datensatz bezieht sich auf die nationale Erhebung der Deutsch-Lehrkräfte in PISA 2018 ($n = 4950$; $n = 2077$ Deutsch-Lehrkräfte, $n = 2873$ sonstige Lehrkräfte). Im Mittelpunkt des Forschungsinteresses standen dabei Lehrkräfte, die in Klassenstufen mit einem hohen Anteil von fünfzehnjährigen Schülerinnen und Schülern unterrichten. Alle Lehrkräfte (inkl. Referendarinnen und Referendare), die eine Lehrbefähigung besitzen, um Schülerinnen und Schüler der 9. Jahrgangsstufe und/oder der 10. Jahrgangsstufe zu unterrichten, erhielten einen Lehrerfragebogen. Von allen Lehrkräften einer Schule, die unter die oben beschriebene Definition fielen, wurden zwei Gruppen per Zufall ausgewählt: Eine Gruppe von 15 Lehrkräften, die gegenwärtig Deutsch unterrichten, sowie eine Gruppe von 20 sonstigen Lehrkräften. Um sicherzustellen, dass alle Lehrkräfte befragt wurden, die in der gezogenen 9. Klasse einer Schule unterrichteten und um insbesondere Analysen zum Klassenkontext durchführen zu können, wurden diese Lehrkräfte zusätzlich in die Lehrerstichprobe aufgenommen, auch wenn sie bei der vorherigen Stichprobenziehung zufällig nicht gezogen wurden (vgl. Mang et al., 2019). Der Datensatz lässt sich über die Variablen CENTRYID bzw. CNT, CNTSCHID und CNTTCHID mit den unter <http://www.oecd.org/pisa/data/2018database/> frei verfügbaren internationalen Daten der Erhebung im Jahr 2018 verbinden.

Der Datensatz enthält ferner folgende Variablen (Gruppen):

- 4.3.1. Weitere organisatorische und administrative Variablen (SPSS-Variablenansicht: Zeilen 5 - 12)
- 4.3.2. Identifikationsvariable für Lehrerinnen und Lehrer, welche die gezogene 9. Klasse unterrichten (**Embargo bis 31.12.2022**; SPSS-Variablenansicht: Zeile 13)
- 4.3.3. Identifikationsvariable für Lehrerinnen und Lehrer, ob sie zu der Gruppe der DeutschlehrerInnen oder sonstigen LehrerInnen gehören (SPSS-Variablenansicht: Zeile 14)
- 4.3.4. Sämtliche Variablen aus der *internationalen* Erhebung sowie die *nationalen Ergänzungsfragen*, wie sie im Lehrerfragebogen in Deutschland erhoben wurden (SPSS-Variablenansicht: Zeilen 15 - 311).
- 4.3.5. Verschiedene, im Rahmen der internationalen Auswertungsvorgaben abgeleitete Skalen aus den Einzelvariablen des Lehrerfragebogens. Für eine detaillierte Beschreibung der angewendeten Skalierungsprozeduren und der jeweiligen Inhalte der Skalen sei hier auf die OECD-Literatur zu PISA 2018 verwiesen (OECD 2019; OECD, in Vorbereitung); (SPSS-Variablenansicht: Zeilen 312 - 359).
- 4.3.6. Zwei weitere Variablen für mögliche Gewichtungen der Lehrer (vgl. OECD, in Vorbereitung) (SPSS-Variablenansicht: Zeile 360-361).
- 4.3.7. Das Datum der Erstellung des Datensatzes durch das internationale PISA Konsortium 2018 (SPSS-Variablenansicht: Zeile 362).

4.4. Der Datensatz „PISA2018_Schulleiterdaten_CBA.sav“.

Dieser Datensatz beinhaltet die Aussagen der Schulleiterinnen und Schulleiter der befragten und an diesem Fragebogen teilgenommenen Schulen der PISA 2018 Stichprobe ($n = 191$). Analog zum Datensatz der Schülerinnen und Schüler lässt sich auch dieser Datensatz über die Variablen CNTRYID bzw. CNT, CNTSCHID mit den unter <http://www.oecd.org/pisa/data/2018database/> frei verfügbaren internationalen Daten der Erhebung im Jahr 2018 verbinden.

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

4.4.1. Weitere organisatorische und administrative Variablen

(SPSS-Variablenansicht: Zeilen 4 - 11)

4.4.2. Sämtliche Variablen aus der *internationalen* Erhebung sowie die *nationalen*

Ergänzungsfragen, wie sie im Schulleiterfragebogen in Deutschland erhoben wurden (SPSS-Variablenansicht: Zeilen 12 - 177).

4.4.3. 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 2018 verwiesen (OECD 2019; OECD, in Vorbereitung);

(SPSS-Variablenansicht: Zeilen 178 - 195).

4.4.4. Das finale Schul-Fallgewicht, eine Summengewichtung sowie das sog. Senat-Gewicht der Schulleiter (vgl. OECD, in Vorbereitung)

(SPSS-Variablenansicht: Zeilen 196 - 198).

4.4.5. Das Datum der Erstellung des Datensatzes durch das internationale PISA Konsortium 2018

(SPSS-Variablenansicht: Zeile 199).

4.5. Der Datensatz

„PISA2018_Schuelerdaten_Timing_CBA_PISA_Sample.sav“.

Dieser Datensatz beinhaltet die Zeit-Angaben der fünfzehnjährigen Schülerinnen und Schüler der PISA 2018 Stichprobe ($n = 5451$). Die Bearbeitungszeit-Angaben sind immer in Millisekunden angegeben. Der Datensatz lässt sich über die Variablen `CNTRYID` bzw. `CNT`, `CNTSCHID` und `CNTSTUID` mit den unter <http://www.oecd.org/pisa/data/2018database/> frei verfügbaren internationalen Daten der Erhebung im Jahr 2018 verbinden.

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

4.5.1. Weitere organisatorische und administrative Variablen

(SPSS-Variablenansicht: Zeilen 5 - 13)

4.5.2. Bearbeitungszeiten in Millisekunden für die Erklärungs- und Einführungsseiten sowie

Variablen aus der *internationalen* Erhebung, wie sie im Schülerfragebogen in Deutschland erhoben wurden (SPSS-Variablenansicht: Zeilen 14 - 132).

4.5.3. Das Datum der Erstellung des Datensatzes durch das internationale PISA Konsortium 2018

(SPSS-Variablenansicht: Zeile 133).

5. Literatur

Mang, J., Wagner, S., Gomolka, J., Schäfer, A., Meinck, S., & Reiss, K. (2019). *Technische Hintergrundinformationen PISA 2018*. München: Technische Universität München.

<http://doi.org/10.14459/2019md1518258>

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

OECD (2019). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing: Paris.

<https://doi.org/10.1787/b25efab8-en>

OECD (in Vorbereitung). *PISA 2018 Technical Report*. OECD Publishing. Abgerufen von

<http://www.oecd.org/pisa/data/2018-technical-report/>