

Hinweise zu den FDZ-Daten für die Erhebung PISA 2013 – nationale Ergänzung der internationalen fünften Erhebungsrunde (Version 2)

Jörg-Henrik Heine
Zentrum für internationale Bildungsvergleichsstudien (ZIB e.V.)
Technische Universität München (TUM)

vom FDZ am IQB überarbeitete Fassung zur Version 2 der Datensätze

Inhalt

| | |
|---|---|
| Allgemeines | 1 |
| Bereitgestellte Datensätze | 2 |
| Beschreibung der Inhalte und Variablengruppen der Datensätze zum zweiten Messzeitpunkt..... | 3 |
| PISA Plus Gewichtung – Dokumentation und Verwendung der Gewichte | 4 |
| Zusammenfügen der Datensätze | 5 |
| Indikatoren zum Sozioökonomischen Status – Coding-Prozeduren ISCO08, ISEI, SIOPS..... | 8 |
| Literatur..... | 9 |

Allgemeines

Bereitgestellt werden hier die Daten aus der **nationalen Ergänzungsstudie** der internationalen fünften Erhebungsrunde „**Programme for International Student Assessment**“ (PISA) im Jahr 2013. Die im Einzelnen bereitgestellten Daten beziehen sich auf die deutsche Stichprobe der Schülerinnen und Schüler der neunten Klassen. Im Rahmen des Sampling wurden hier jeweils zwei vollständige Klassen aus den für das „PISA-Sample“ gezogenen Schulen erhoben. Darüber hinaus werden Daten für die Lehrkräfte, Eltern und Schulleitungen für den (zweiten) Erhebungszeitraum 2013 bereitgestellt.

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 (*simple random sample*) 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). Ferner ergibt sich die Notwendigkeit der Verwendung von Gewichtungsvariablen aus dem Umstand, dass zum zweiten Messzeitpunkt im Rahmen des PISA Plus Projektes ein nicht zu vernachlässigender Stichprobenausfall aufgrund der Freiwilligkeit der Teilnahme auf der Ebene der Schülerinnen und Schüler, als auch auf der Ebene der Schulen, zu verzeichnen war (vgl. Heine, Nagy, Meinck, Zühlke & Mang, 2017).

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 Anwenderinnen und Anwender vor jeglichen Auswertungen durch Inspektion der Daten über das Werteschema und die Kodierung der jeweils interessierenden Variablen Klarheit verschaffen müssen. 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.

Bereitgestellte Datensätze

Im Einzelnen wurden die folgenden Datensätze zum zweiten Messzeitpunkt im SPSS-Format dem Forschungsdatenzentrum (FDZ) als sogenannte *Scientific Use Files (SUF)* zur Verfügung gestellt (Datenversion 1):

- PISA-Plus-2012-2013_Datensatz_SchuelerInnen_COG_SUF_Antrag.sav
- PISA-Plus-2012-2013_Datensatz_SchuelerInnen_FB_SUF_Antrag.sav
- PISA-Plus-2012-2013_Datensatz_LehrerInnen_SUF_Antrag.sav
- PISA-Plus-2012-2013_Datensatz_SchulleiterInnen_SUF_Antrag.sav
- PISA-Plus-2012-2013_Datensatz_Gewichte_SUF_Antrag.sav
- PISA-Plus-2012-2013_Link_Datensatz_SUF_Antrag.sav

Im Zuge der Datenaufbereitung zur zweiten Datenversion wurden diese Teildatensätze zu zwei Datensätzen zusammengefasst:

- PISA-Plus_2012-2013_Datensatz_T2_SUF_v2_Antrag.sav
- PISA-Plus_2012-2013_LehrerInnendaten_T2_SUF_v2_Antrag.sav

Der Datensatz „PISA-Plus_2012-2013_Datensatz_T2_SUF_v2_Antrag.sav“ umfasst Variablen aus dem Schülerinnen- und Schülerfragebogen (PISA-Plus-2012-2013_Datensatz_SchuelerInnen_FB_SUF_Antrag.sav), den Testdaten (PISA-Plus-2012-2013_Datensatz_SchuelerInnen_COG_SUF_Antrag.sav), dem Gewichte-Datensatz (PISA-Plus-2012-2013_Datensatz_Gewichte_SUF_Antrag.sav), dem Link-Datensatz (PISA-Plus-2012-2013_Link_Datensatz_SUF_Antrag.sav), und dem Schulleitungsfragebogen (PISA-Plus-2012-2013_Datensatz_SchulleiterInnen_SUF_Antrag.sav). Die unterschiedlichen Datenquellen sind in den Variablenlabeln im Gesamtdatensatz wie folgt gekennzeichnet:

- Schülerfragebogen: PISA-Plus-2012-2013_Datensatz_SchuelerInnen_FB_SUF_Antrag.sav
- Linkdaten: PISA-Plus-2012-2013_Link_Datensatz_SUF_Antrag.sav
- Testdaten: PISA-Plus-2012-2013_Datensatz_SchuelerInnen_COG_SUF_Antrag.sav
- Gewichtsdaten: PISA-Plus-2012-2013_Datensatz_Gewichte_SUF_Antrag.sav
- Schulleitungsfragebogen: PISA-Plus-2012-2013_Datensatz_SchulleiterInnen_SUF_Antrag.sav

Der Datensatz „PISA-Plus_2012-2013_LehrerInnendaten_T2_SUF_v2_Antrag.sav“ beinhaltet die Angaben aus dem Lehrkräftefragebogen zum zweiten Messzeitpunkt. Dabei wurden auch fachspezifische Angaben (Mathematik, Biologie, Chemie, Physik, Naturwissenschaften) erfasst. Da Schülerinnen und Schüler in der Regel von mehreren Lehrkräften (z. B. verschiedene Lehrkräfte in Mathematik und Biologie) unterrichtet wurden, kann der Lehrkräftedatensatz nicht direkt mit dem Gesamtdatensatz zum zweiten Messzeitpunkt verknüpft werden. Zunächst muss ausgewählt werden, welche fachspezifischen Lehrkräfteangaben herangespielt werden sollen. Eine kurze Anleitung (mit SPSS-Syntax) dazu wird im Abschnitt „Zusammenfügen der Datensätze“ gegeben.

Neben den Daten zum zweiten Messzeitpunkt wird ein Gesamtdatensatz zum ersten Messzeitpunkt bereitgestellt: „PISA-Plus_2012-2013_Datensatz_T1_SUF_v2_Antrag.sav“. Dieser Datensatz basiert auf den Teildatensätzen der Neuntklässlerinnen und Neuntklässler der PISA-2012-Studie (http://doi.org/10.5159/IQB_PISA_2012_v5). Im Datensatz enthalten sind die Variablen aus dem

Schülerinnen- und Schülerfragebogen, Schulleitungsfragebogen und Lehrkräftefragebogen. Im Vergleich zur vorherigen Datenversion wurde am FDZ des IQB nachträglich eine Klassen-ID-Variable („idclass_FDZ“) aus der Schul-ID und der Klassenbezeichnung gebildet. Die Klassen-ID war nicht Teil des übergebenen Datenpakets und könnte in Einzelfällen zu fehlerhaften Zuordnungen von Schülerinnen und Schülern zu Klassen führen.

Außerdem sind die schulischen Kompetenzen (in Form von Plausible Values) zum ersten Messzeitpunkt enthalten. Die Plausible Values (PVs) sind allerdings auf einer anderen Metrik ($M = 500$, $SD = 100$) als die korrespondierenden PVs zum zweiten Messzeitpunkt (Logit-Metrik). Zur Analyse der Leistungsentwicklung der Schülerinnen und Schüler sollten deshalb die Kompetenzschätzer im Gesamtdatensatz zum zweiten Messzeitpunkt (PISA-Plus_2012-2013_Datensatz_T2_SUF_v2_Antrag.sav) genutzt werden. In diesem Datensatz werden Kompetenzschätzer zu beiden Messzeitpunkten bereitgestellt.

Die Gesamtstichprobe in den Datensätzen umfasst $n = 9998$ Schülerinnen und Schüler. Um die Datensätze auf die Analytestichprobe von $n = 4871$ (Heine et al., 2017) zu reduzieren, können die Filtervariablen „Pplus_STD_ParticipationTH1_dicho“ (Teilnahme an den Kompetenztests zum zweiten Messzeitpunkt, $1 = ja$) und „inpp“ (Schülergesamtgewicht vergeben, $1 = ja$) im Gesamtdatensatz zum zweiten Messzeitpunkt genutzt werden.

Beschreibung der Inhalte und Variablengruppen der Datensätze zum zweiten Messzeitpunkt

Der Datensatz **PISA-Plus_2012-2013_Datensatz_T2_SUF_v2_Antrag.sav** bezieht sich auf die klassenbasierte Ergänzungsstichprobe derjenigen Schülerinnen und Schüler in Deutschland, die sich zum ersten Messzeitpunkt (2012) in einer 9. Klasse befunden haben. Der Gesamtdatensatz zum zweiten Messzeitpunkt (2013) umfassen zunächst (unter den jeweils gleichen ID-Variablen) dieselben Schülerinnen und Schüler ($n = 9998$), welche in den Daten zu Messzeitpunkt 1 (2012) enthalten sind. Die Datensätze der beiden Messzeitpunkte 2012 (PISA-Plus_2012-2013_Datensatz_T1_SUF_v2_Antrag.sav) und 2013 (PISA-Plus_2012-2013_Datensatz_T2_SUF_v2_Antrag.sav) lassen sich über die Variablen **SCHOOLID** und **StIDStd** miteinander verbinden (siehe Abschnitt „Zusammenfügen der Datensätze“). Beide Datensätze lassen sich **allerdings NICHT** über die Variablen **SCHOOLID** und **StIDStd** mit den unter <http://www.oecd.org/pisa/pisaproducts/> frei verfügbaren **internationalen Daten** der Erhebung im Jahr 2012 verbinden. Es muss an dieser Stelle darauf hingewiesen werden, dass insbesondere aufgrund der Freiwilligkeit einer Teilnahme zum zweiten Messzeitpunkt auf Schulleben, wie auch auf Ebene der Schülerinnen und Schüler ein nicht unerheblicher Datenausfall zu T2 (2013) vorliegt, sodass nicht von einem Vollständigen „data-matching“ ausgegangen werden kann. Einzelheiten zum Ausfallmechanismus werden in dem publizierten Aufsatz „Empirische Grundlage, Stichprobenausfall und Adjustierung im PISA-Längsschnitt 2012–2013“ (Heine, Nagy, Meinck, Zühlke & Mang, (2017) behandelt.

Die Testdaten im Gesamtdatensatz zum zweiten Messzeitpunkt umfassen neben den ID-Variablen die folgenden Variablengruppen:

- 1.1. Verschiedene Schul- und Schülerdemographische Variablen wie z. B. zum Teilnahmestatus der Schulen am 2. Messzeitpunkt, der Schulart, des angestrebten Schulabschlusses oder der jeweils verwendeten Fragebogen- und Testheftversion.
- 1.2. Variablen zu denjenigen PISA-Fragen (kognitive Items) zu den Domänen Mathematik, Lesen und Naturwissenschaften, welche zu beiden Messzeitpunkten (2012-2013) zur Erfassung der Kompetenzen der Schülerinnen und Schüler eingesetzt wurden. Diese Variablen enthalten die kodierten Antworten (richtige Lösung vs. Falsche Lösung) der Schülerinnen und Schüler zum Messzeitpunkt 2 (2013).
Es werden hier jeweils zwei Arten der Kodierung bereitgestellt: (1) Kodierung nicht erreichter Fragen (*not reached*) als (strukturell) fehlende Werte – z. B.

„Pplus_SCALING_PXxxxxxxx_nrEQmiss“ sowie (2) Kodierung nicht erreichter Fragen (*not reached*) als falsch beantwortet – z. B. „Pplus_SCALING_PXxxxxxxx_nrEQfalse“.

- 1.3. Drei Variablen zum Taschenrechner-Gebrauch und zum Anstrengungs-Barometer (vgl. Butler & Adams, 2007; Kunter, Schümer, Artelt, Baumert, Klieme, Neubrand, et al., 2002); („Pplus_CLCUSE1“, „Pplus_CLCUSE2“, „Pplus_CLCUSE3“).
- 1.4. Punktschätzer für die Kompetenzen der Schülerinnen und Schüler in den drei PISA Kompetenzbereichen – *Mathematik, Lesen und Naturwissenschaften* – (**WLE**) nach Warm (1989), sowie dazugehörige Standardfehler. Da die Daten über beide Messzeitpunkte einer gemeinsamen Neu-Skalierung unterzogen wurden (vgl. Nagy, Lüdtke, Heine & Köller, 2017) legen wir Kompetenzschätzer für beide Messzeitpunkte T1 und T2 (2012 und 2013) vor – z. B. die Variablen „m1_wle“ für Mathematik-Kompetenz zu T1 und „m1_ser“ für den dazugehörigen Standardfehler oder „r2_wle“ für Lese-Kompetenz zu T2 und „r2_ser“ für den dazugehörigen Standardfehler, etc.
- 1.5. Plausible Values für die drei Kompetenzbereiche – *Mathematik, Lesen und Naturwissenschaften* – für beide Messzeitpunkte T1 und T2 (2012 und 2013) z. B. die Variablen „mpv1_w1“ für Mathematik-Kompetenz Plausible Value 1 zu T1 („welle 1“), oder „pv4_w2“ für Mathematik-Kompetenz Plausible Value 4 zu T2 („welle 2“), etc.
- 1.6. Das um den Datenausfall zu T2 adjustierte finale Fallgewicht „**pptotwgt**“, welches inhaltlich mit dem finale Fallgewicht zu T1 (im Datensatz **PISA-Plus_2012-2013_Datensatz_T1_SUF_v2_Antrag.sav**) mit der Bezeichnung „**W_FSTUWT**“ korrespondiert (vgl. OECD, 2014).
- 1.7. 80 um den Datenausfall zu T2 adjustierte Replikationsgewichte „**repwgt1**“ bis „**repwgt80**“, welche inhaltlich mit den Replikationsgewichten zu T1 (im Datensatz **PISA-Plus_2012-2013_Datensatz_T1_SUF_v2_Antrag.sav**) mit der Bezeichnung „**W_FSTR1**“ bis „**W_FSTR80**“ korrespondieren (vgl. OECD, 2014).
Sowie 4 weitere Gewichtungsvariablen, die sich aus der klassenbasierten Struktur der Daten ergeben („**ppschwgt**“, „**ppwscwgt**“, „**ppwscwgt**“, „**ppwithinschwgt**“)

PISA Plus Gewichtung – Dokumentation und Verwendung der Gewichte

1. Überblick:

In empirischen Bildungsstudien repräsentieren nicht alle der teilnehmenden Schülerinnen und Schüler den gleichen Anteil der Population. Zum einen haben die Schülerinnen und Schüler jeweils unterschiedliche Ziehungswahrscheinlichkeiten aufgrund der geschichteten, mehrstufigen Stichprobe. Zum anderen bestehen für die gezogenen Schülerinnen und Schüler unterschiedliche Wahrscheinlichkeiten, tatsächlich an den Studien teilzunehmen. Um exakte Populationsschätzer zu erhalten, die eine Generalisierung von der Stichprobe auf die Zielpopulation erlauben, ist die Berechnung und Anwendung von Gewichten notwendig. Im Falle der Studie PISA Plus 2013 sind beide Punkte von Bedeutung. Hierfür wurden die an der Studie PISA 2012 teilnehmende Schülerinnen und Schüler im darauffolgenden Jahr erneut getestet und befragt. Hierbei waren starke Ausfälle zwischen erster und zweiter Erhebungswelle zu beobachten (vgl. Heine, Nagy, Meinck, Zühlke & Mang, 2017). Erste Analysen ließen Zusammenhänge mit den Analysevariablen, insbesondere mit den schulischen Leistungen, erwarten. Durch eine Nachgewichtung der Teilnehmerinnen und Teilnehmer der zweiten Erhebung sollten diese Verzerrungen durch Nichtteilnahme ausgeglichen und die Verallgemeinerbarkeit der Ergebnisse von PISA-Plus 2012-2013 sichergestellt werden. Diese Nachgewichtung wurde im Februar 2015 durch die IEA Hamburg vorgenommen.

2. Zielpopulation:

In PISA Plus 2012-2013 wurden im Frühjahr 2013 diejenigen Schülerinnen und Schüler befragt, die im Vorjahr an der Studie PISA 2012 teilgenommen hatten und dabei als Teil einer intakten Klasse gezogen wurden (nicht aber die 15-Jährigen Schülerinnen und Schüler außerhalb der gezogenen Klasse(n) einer Schule). Schülerinnen und Schüler an Berufsschulen oder an im Rahmen des für PISA 2012 vorgenommenen Förderschulen-Oversamplings gezogenen Schulen waren nicht Teil der PISA Plus-Zielpopulation. Schülerinnen und Schüler an Schulen, die 2012 an PISA teilgenommen hatten, aber seitdem die Schule verlassen haben, wurden für PISA Plus 2013 nicht nachverfolgt. Sie können daher nicht als Teil der Zielpopulation betrachtet werden. Ebenso wenig wurde die Schülerinnen- und Schülerpopulation durch Hinzufügen von Zehntklässlerinnen und Zehntklässler aufgefrischt. Somit repräsentieren die für PISA Plus getesteten Schülerinnen und Schüler die Population der Neuntklässlerinnen und Neuntklässler des Vorjahres, die sich zum zweiten Messzeitpunkt immer noch in derselben Schule befanden, sofern diese eine zehnte Klasse enthielt. Bezogen auf diese Zielpopulation wurden die Gewichte adjustiert.

3. Teilnahmestatus:

Die korrekte Bestimmung des Teilnahme-/Teilnahmeberechtigungsstatus einer Untersuchungseinheit ist bedeutsam für die Gewichtung, da nur der Teilnahmedefinition entsprechende Schülerinnen und Schüler Gewichte erhalten und nur teilnahmeberechtigte Einheiten zur Nichtteilnahmeadjustierung beitragen. Bei der Gewichtung für PISA Plus wurde (IEA-Prozeduren etwa in TIMSS oder IGLU folgend) die Teilnahme einer Schülerin oder eines Schülers am Leistungstest als ausschlaggebend für seinen Teilnahmestatus betrachtet. Dies führte zu folgenden Teilnahmekriterien für PISA-Plus 2013 auf den beiden Ebenen *Schulen* und *Schülerinnen und Schüler*.

3.1. Schulen:

- 3.1.1. Als teilnehmend wird eine Schule betrachtet, wenn sie an PISA 2012 teilgenommen hat und mindestens einer ihrer Schülerinnen und Schüler am Leistungstest in PISA Plus 2012-2013 teilgenommen hat;
- 3.1.2. Als teilnahmeberechtigt wird eine Schule betrachtet, wenn sie an PISA 2012 teilgenommen hat und in ihr eine zehnte Jahrgangsstufe vorhanden ist;
- 3.1.3. Als *nicht* teilnahmeberechtigt werden Schulen ohne zehnte Jahrgangsstufe behandelt.

3.2. Schülerinnen und Schüler:

- 3.2.1. Als teilnehmend wird eine Schülerin oder ein Schüler betrachtet, wenn sie oder er am Leistungstest teilgenommen hat (Variable „Pplus_STD_ParticipationTH1_dicho“ im Datensatz des zweiten Messzeitpunkts).
- 3.2.2. Als teilnahmeberechtigt wird eine Schülerin oder ein Schüler an einer teilnehmenden Schule betrachtet, wenn sie oder er an PISA 2012 teilgenommen hat und die Schule seitdem nicht verlassen hat;
- 3.2.3. Als nicht teilnahmeberechtigt wird eine Schülerin oder ein Schüler an einer teilnehmenden Schule betrachtet, wenn er die Schule vor der Auffrischung des (PISA 2012) *student listing forms* für PISA Plus 2013 verlassen hat (und damit nicht im als Eintrag *student tracking form* verzeichnet ist); er die Schule nach Erstellung des PISA Plus 2013 *student tracking forms* verlassen hat.

Zusammenfügen der Datensätze

1. Verknüpfung der Datensätze zum ersten und zweiten Messzeitpunkt

Die Schülerdatensätze zu Messzeitpunkt 1 (2012) und Messzeitpunkt 2 (2013) lassen sich über die Schlüsselvariable **idstud_FDZ** miteinander verbinden. Folgende Variablen kommen in beiden Datensätzen in identischer Ausprägung vor: „Version_v2_29.04.2020“, „SCHOOLID“, „StIDStd“. Folgende Variablen haben in beiden Datensätzen identische Namen und müssen vor dem Zusammenführen

umbenannt werden (z. B. mit Zusatz "_T1" und "_T2"): DEU_SC91A01, DEU_SC91A02, DEU_SC90A01. In SPSS kann zur Verknüpfung folgende Syntaxvorlage genutzt werden:

* DataSet1 = PISA-Plus_2012-2013_Datensatz_T1_SUF_v2_Antragsnummer.sav

* DataSet2 = PISA-Plus_2012-2013_Datensatz_T2_SUF_v2_Antragsnummer.sav

```
DATASET ACTIVATE DataSet1.
SORT CASES BY idstud_FDZ.
DATASET ACTIVATE DataSet2.
SORT CASES BY idstud_FDZ.
DATASET ACTIVATE DataSet1.
MATCH FILES /FILE=*
  /FILE='DataSet2'
  /RENAME (Version_v2_29.04.2020 SCHOOLID StIDStd = d0 d1 d2)
  DEU_SC91A01=DEU_SC91A01_T2
  DEU_SC91A02=DEU_SC91A02_T2
  DEU_SC90A01=DEU_SC90A01_T2 DEU_SC91A01=DEU_SC91A01_T2
  DEU_SC91A02=DEU_SC91A02_T2 DEU_SC90A01=DEU_SC90A01_T2
  /BY idstud_FDZ
  /DROP= d0 d1 d2.
EXECUTE.
```

Alternativ können auch die Variablen Variablen **SCHOOLID** und **StIDStd** (gemeinsam) genutzt werden, um die beiden Datensätze zu verknüpfen. Beide Variablen dienen dabei als Schlüsselvariablen. In SPSS kann zur Verknüpfung folgende Syntaxvorlage genutzt werden:

* DataSet1 = PISA-Plus_2012-2013_Datensatz_T1_SUF_v2_Antragsnummer.sav

* DataSet2 = PISA-Plus_2012-2013_Datensatz_T2_SUF_v2_Antragsnummer.sav

```
DATASET ACTIVATE DataSet1.
SORT CASES BY SCHOOLID StIDStd (A).
DATASET ACTIVATE DataSet2.
SORT CASES BY SCHOOLID StIDStd (A).
DATASET ACTIVATE DataSet1.
MATCH FILES /FILE=*
  /FILE='DataSet2'
  /BY SCHOOLID StIDStd.
EXECUTE.
```

Achtung: Die beiden Variablen **SCHOOLID** und **StIDStd** sind nur Namensgleich mit den **international** verwendeten Variablen, enthalten aber eine gänzlich unterschiedliche Ordnungssystematik. Diese Variablen können somit nicht genutzt werden, um die vorliegenden Datensätze mit den internationalen PISA-2012-Daten der 15-Jährigen zu verknüpfen.

2. Verknüpfung des Gesamtdatensatzes mit dem Lehrkräftedatensatz zum zweiten Messzeitpunkt

Die Datensätze **PISA-Plus_2012-2013_Datensatz_T2_SUF_v2_Antrag.sav** und **PISA-Plus_2012-2013_LehrerInnendaten_T2_SUF_v2_Antrag.sav** können über die Lehrkräfte-ID (**Pplus_IDTEACH_FDZ**) miteinander verbunden werden. Aufgrund von (teilweise) fehlenden Angaben zu Klassenbezeichnungen bzw. allgemein fehlenden Werten, können dabei nicht alle Schülerinnen und Schüler mit

korrespondierenden Lehrkräfteangaben verbunden werden. Außerdem können auf Schülerinnen- und Schülerebene die Daten mehrerer Lehrkräfte hinzugefügt werden, je nachdem, welches Fach analysiert wird. Außerdem unterrichteten Lehrkräfte häufig mehrere Fächer, wodurch die Datensätze fachspezifisch zusammengeführt werden müssen. Im Gesamtdatensatz zum zweiten Messzeitpunkt sind insgesamt neun fachspezifische Lehrkräfte-Verknüpfungsvariablen enthalten: Pplus_IDTEACH_Mat_FDZ, Pplus_IDTEACH_Mat2_FDZ, Pplus_IDTEACH_Bio_FDZ, Pplus_IDTEACH_Bio2_FDZ, Pplus_IDTEACH_Che_FDZ, Pplus_IDTEACH_Che2_FDZ, Pplus_IDTEACH_Phy_FDZ, Pplus_IDTEACH_Phy2_FDZ, Pplus_IDTEACH_Nawi_FDZ.

Die Ergänzung der Lehrkräfte-ID um ein Fächerkürzel war notwendig, da Schülerinnen und Schüler in der Regel (1) mehrere Lehrkräfte und (2) Lehrkräfte verschiedener Fächer zugeordnet werden können. Vor der Verknüpfung der beiden Datensätze muss die Lehrkräfte-ID **Pplus_IDTEACH** im Lehrkräftedatensatz entsprechend der gewünschten Fachlehrkräfte umbenannt bzw. dupliziert werden, um als Schlüsselvariable dienen zu können. Die Verknüpfung lässt sich in SPSS über folgende Syntaxvorlage durchführen:

*1. im Lehrkräfte-Datensatz zunächst die globale Lehrkräfte-ID umbenennen, in diesem Beispiel für die Auswahl von Mathematik-Lehrkräften

```
DATASET ACTIVATE DataSet2.  
RENAME VARIABLES (Pplus_IDTEACH_FDZ = Pplus_IDTEACH_Mat_FDZ).
```

*2. Gesamtdatensatz mit Lehrkräfte-Datensatz verknüpfen (One-to-many-Merging). Variablen aus dem Lehrkräftefragebogen werden dem Gesamtdatensatz hinzugefügt

* DataSet1 = PISA-Plus_2012-2013_Datensatz_T2_SUF_v2_Antragsnummer.sav

* DataSet2 = PISA-Plus_2012-2013_LehrerInnendaten_T2_SUF_v2_Antragsnummer.sav

```
DATASET ACTIVATE DataSet1.  
SORT CASES BY Pplus_IDTEACH_Mat_FDZ.  
DATASET ACTIVATE DataSet2.  
SORT CASES BY Pplus_IDTEACH_Mat_FDZ.  
DATASET ACTIVATE DataSet1.  
MATCH FILES /FILE=*  
/TABLE='DataSet2'  
/BY Pplus_IDTEACH_Mat_FDZ.  
EXECUTE.
```

In den Fällen, in denen sich der Klassenverband (von der 9. zur 10. Klasse) neu zusammengesetzt hat oder der Unterricht im Kurssystem fortgeführt wird, wurden nur die Lehrkräfte befragt, die den größten Teil der ehemaligen neunten Klassen übernommen haben. Die Schülerinnen und Schüler, die nur vereinzelt in andere Kurse oder Klassen wechselten, haben daher im Datensatz keine zugeordneten Lehrkräfte. Wenn die Schüler auf sehr viele Kurse oder Klassen aufgeteilt wurden und kein Kurs oder Klasse eine gewisse Anzahl von PISA-Schülerinnen und Schüler aufwies, wurde keine Befragung der Lehrkräfte vorgenommen. Zusätzlich bleibt die Verknüpfung unvollständig, wenn eine Schülerin bzw. ein Schüler ein Fach nicht belegt hat (häufig Naturwissenschaften oder alternativ Biologie, Chemie und Physik). Angaben der zweiten Lehrkraft pro Fach gibt es nur in seltenen Fällen. In vielen Fällen wurden die Kurse der Lehrkräfte von den Schulen in anderer Form angegeben als bei den Schülerinnen und Schülern. Wenn eine eindeutige Zuordnung möglich war, wurde diese vorgenommen. Dennoch lässt sich nicht jede Lehrkraft bestimmten Schülerinnen und Schülern zuordnen. Es kam auch vor, dass in einigen Fällen Lehrkräfte von den Schulen gemeldet wurden, die gar keine PISA-Schülerinnen und Schüler

unterrichten. Lehrkräfte, die keinen Schülerinnen und Schülern zugeordnet werden können, werden im Gesamtdatensatz zum zweiten Messzeitpunkt in den Lehrkräfte-ID-Variablen nicht aufgeführt.

3. Verknüpfung mit den Daten der Studien PISA 2012 & IQB-Ländervergleich 2012

Die vorliegenden Datensätze (Version 2) zur PISA-Plus 2012-2013-Studie lassen sich über die ID-Variablen **idstud_FDZ** mit den Teildatensätzen der PISA-2012-Studie (http://doi.org/10.5159/IQB_PISA_2012_v5) zum ersten Messzeitpunkt verbinden, z. B. mit den Elternangaben zum ersten Messzeitpunkt (aufgrund dem hohen Anteil fehlender Werte, die aus einer Verknüpfung resultieren, wurde dieser Teildatensatz nicht in den Gesamtdatensatz zum ersten Messzeitpunkt integriert). Außerdem kann der Gesamtdatensatz zum ersten Messzeitpunkt über die ID-Variablen **idstud_FDZ** mit den Daten des IQB-Ländervergleichs 2012 (http://doi.org/10.5159/IQB_LV_2012_v4) verknüpft werden. Die Lehrkräfte- und Schul-IDs sind allerdings **NICHT** für ein Linking der Datensätze geeignet.

Indikatoren zum Sozioökonomischen Status – Coding-Prozeduren ISCO08, ISEI, SIOPS

1. Überblick

Die Daten wurden im elektronischen Kodier-System der IEA Hamburg (Coding-Expert) bearbeitet. Alle Daten wurden manuell kodiert und es wurde demnach keine automatische Berufs-Kodierung eingesetzt. Den kodierenden Personen wurden dazu folgende Angaben aus den erhobenen Roh-Daten vorgelegt: Beruf, Tätigkeit, Weisungsbefugnisse (Anzahl Untergebene) der Eltern. Es wurden keine Coding Sheets verwendet. Alle Kodierenden und Kodierer wurden durch IEA Personal Anfang 2013 geschult. Während und nach der Schulung wurden die Übereinstimmungen zwischen den kodierenden Personen überprüft. Alle Kodierenden und Kodierer absolvierten einen Einstellungstest des IEA für Kodierende. Die Schulung der kodierenden Personen umfasste fünf ganztägige Schulungstermine sowie die Bearbeitung von Übungsmaterial über fünf Wochen. Jede kodierende Person vergab in der Schulungsphase ca. 10.000 ISCO08-Übungs-Codes. Alle Schulungstermine sowie die Bearbeitung des Übungsmaterials waren verpflichtend. Während der Schulungsphase wurden die Übereinstimmungen kontinuierlich geprüft und problematische Antworten in den Schulungen mit Hilfe der IEA Training Software (Coding-Expert Trainer), welche die Antwort sowie die von allen Kodierern vergebenen Codes anzeigt, diskutiert. Alle Daten wurden einfach kodiert und von der IEA stichprobenartig überprüft. Alle Ergebnisse wurden in ihrer Gesamtheit auf Vollständigkeit und Plausibilität geprüft.

2. Missing Data

Es wurden folgende Missing-Codes vergeben – aber teilweise nicht im SPSS-System deklariert:

- 9701 - Hausfrau/Hausmann
- 9702 - in Ausbildung
- 9703 - aktuell nicht berufstätig
- 9704 - Kommentare & Scherzantworten
- 9705 - zu vage Angaben bzw. nicht zuzuordnen
- 9997 - nicht bearbeitbar
- 9999 – missing
- sysmiss (blank) - keine Angaben zum Vercoden in den Daten

3. Transformationen

Die zur Berechnung der sozio-ökonomischen Indizes (ISEI, SIOPS) verwendeten Transformationsalgorithmen sind als SPSS-Syntax dokumentiert. Die ISEI- und SIOPS-Transformationen stammen von Harry Ganzeboom. Die Transformation wurde von der IEA auf Basis des Ganzeboom-Skriptes für ISCO88 erstellt. Es wird keine Gewähr übernommen.

4. Variablen im Gesamtdatensatz zum zweiten Messzeitpunkt

- "Pplus_ISEI08_PISA_Plus_ISCO08_M" – auf Basis der Angaben zur Mutter
- "Pplus_ISEI08_PISA_Plus_ISCO08_V" – auf Basis der Angaben zum Vater
- "Pplus_SIOPS08_PISA_Plus_ISCO08_M" – auf Basis der Angaben zur Mutter
- "Pplus_SIOPS08_PISA_Plus_ISCO08_V" – auf Basis der Angaben zum Vater
- "Pplus_PISA_Plus_ISCO08_M" – auf Basis der Angaben zur Mutter
- "Pplus_PISA_Plus_ISCO08_V" – auf Basis der Angaben zum Vater

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.
- Heine, J.-H., Nagy, G., Meinck, S., Zühlke, O., & Mang, J. (2017). Empirische Grundlage, Stichprobenausfall und Adjustierung im PISA-Längsschnitt 2012–2013. *Zeitschrift für Erziehungswissenschaft*, 20(2), 287–306. <https://doi.org/10.1007/s11618-017-0756-0>
- 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
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450. <https://doi.org/10.1007/BF02294627>
- Nagy, G., Lüdtke, O., & Heine, J.-H., Köller, O. (2017). IRT-Skalierung der Tests im PISA-Längsschnitt 2012/2013: Auswirkungen von Testkontexteffekten auf die Zuwachsschätzung. *Zeitschrift für Erziehungswissenschaft*, 1–30. <https://doi.org/10.1007/s11618-017-0749-z>
- Heine, J.-H., Nagy, G., Meinck, S., Zühlke, O., & Mang, J. (2017). Empirische Grundlage, Stichprobenausfall und Adjustierung im PISA-Längsschnitt 2012–2013. *Zeitschrift für Erziehungswissenschaft*, 20(2), 287–306. <https://doi.org/10.1007/s11618-017-0756-0>