

BEARBEITUNGSBERICHT

PROTOKOLL DER VON SIDOS DURCHGEFÜHRTEN KONTROLLEN DER DOKUMENTATION UND DES DATENSATZES

6705 PISA.ch - Messung der Kompetenzen von 15jährigen

Bundesamt für Statistik - BfS / Abteilung Gesellschaft und Bildung, Espace de l'Europe 10, 2010 Neuchâtel (a)

Schweizerische Konferenz der kant. Erziehungsdirektoren - EDK, Postfach 5975, 3001 Bern (b)

Huguette McCluskey-Cavin (Leitung) (a)

2003AHc PISA.ch - National. Zusatzstichprobe der französischen Schweiz - Umfrage zur Grundkompetenz der SchülerInnen der 9. Klasse - 2000

GENERELLE BEMERKUNGEN

Der Datensatz, auf welchen sich die folgende Kontrolle bezieht, entspricht der Zusatzstichprobe der französischen Schweiz des **nationalen** PISA Projekts, d.h. basiert auf der Stichprobe der SchülerInnen der 9. Klasse und nicht jener der 15-Jährigen. Es existiert zusätzlich zum hier kontrollierten Datensatz ein separater Datensatz (FS Items Grade 9) mit den kodierten Testresultaten. Jener Datensatz ist Teil des zugestellten Materials.

SIDOS hat auch die anderen Teile des nationalen Projektes PISA archiviert. Diese können auch bezogen werden. Es geht um folgende Datensätze: der Datensatz mit den Daten der SchülerInnen der 9.Klasse der gesamten Schweiz (Student Grade 9), derjenige mit den kodierten Testresultaten (Items Grade 9), sowie der Datensatz mit den Daten aus der Umfrage zuhanden der Schuladministration (School Grade 9).

Zusätzlich kann auch der Datensatz mit den Einträgen der 15-jährigen SchülerInnen zu den nationalen Variablen bezogen werden.

Sowohl in der französischen als auch in der deutschen Schweiz wurden in gewissen Kantonen Zusatzstichproben erhoben, um innerkantonale Vergleiche durchführen zu können.

Die Daten der Zusatzstichprobe der Deutschschweiz wurden in den nationalen Datensatz eingefügt, da das Stichprobeverfahren in beiden Fällen dasselbe war. Dies ist im Falle der welschen Schweiz anders. Deshalb sind diese Daten in einer separaten Datei gespeichert. Der Aufbau des Datensatzes, bestehend aus den Variablen des Kontextfragebogens für die SCHÜLER/INNEN und den Variablen – konstruiert nach den Testresultaten – der Leistungsmessung im Lesen, in Mathematik und Naturwissenschaften, entspricht demjenigen des nationalen Datensatzes. Deshalb kann für die Hauptsache auf die Dokumentation des nationalen Datensatzes zurückgegriffen werden. Für die spezifischen Aspekte siehe auch Dossier « Dokumentation Cosortium romand ».

KONTROLLE I : MINI-KONTROLLE

A) Dokumentation

Das PISA 2000 data Manual ist auch für die Daten der welschen Zusatzstichprobe Referenz-Dokumentation.

Dokumentation des Forschungsprojekts

Es gibt ein Dokument, welches einige deskriptive Analysen der Zusatzstichprobe der französisch sprachigen Schweiz vorstellt. Es heisst « Rapport du SRED » und ist im Ordner « Documentation Consortium romand » enthalten.

Dokumentation der Datenerhebungsmethode

Die spezifischen Aspekte dieses Datensatzes im Vergleich zum nationalen Datensatz der SchülerInnen der 9.Klasse betreffen die Stichprobenziehung. Die allgemeine Methode, gültig für alle Datensätze, ist im Nationalen Bericht beschrieben, enthalten im pdf-Dokument im Dossier « Dokumentation BFS_French_Subsample».

Die Stichprobenziehung für dieses Teilprojekt ist beschrieben im rapport de méthode, auf Seiten 34ff (Echantillonnage spécial des classes de 9ème de PISA 2000). Dieser Rapport ist im Dossier « Dokumentation BFS_French_Subsample» im Ordner « Manuals » enthalten.

Die Befragung wird in der französischen Schweiz (BE-f,FR_f,GE,JU,NE,VD,VS_f) durchgeführt vom Département de l'instruction publique du Canton de Genève - DIP, Service de la recherche en éducation - SRED, Quai du Rhône 12, 1205 Genève.

Externe Dokumentation

Das nationale (sowie das internationale) Forschungsprojekt ist in verschiedenen thematisch gegliederten Publikationen (D/F/I/E) des Bundesamtes für Statistik beschrieben, u.a « Für das Leben gerüstet ? Die Grundkompetenzen der Jugendlichen – Nationaler Bericht der Erhebung PISA 200 » und « Das Projekt PISA und die Durchführung in der Schweiz ». Alle Publikationen sind zu beziehen oder einzusehen via homepage : <http://www.pisa.admin.ch>

Kontextuelle Dokumentation

In Bezug auf die Daten, welche zum nationalen Projekt gehören (einschliesslich der Zusatzbefragung in gewissen deutschschweizer Kantonen, um kantonale Vergleiche machen zu können), ist vom Bundesamt für Statistik eine ausführliche Dokumentation zusammengestellt worden, welche SIDOS mit den Daten mitliefert als notwendiger Bestandteil der nötigen Information für eine Sekundäranalyse. Es handelt sich dabei um die sogenannte « **Dokumentation BFS_French_Subsample** ».

Zum Thema internationales Forschungsprojekt PISA (Befragung der 15-Jährigen) sind Informationen via Homepage der OECD zu beziehen : <http://www.pisa.oecd.org/>

Bewertung eventueller Schwierigkeiten in den Darstellungen der Dokumentationsdateien

Die Dokumentationsdateien liefern auf gut präsentierte Weise viel Detailinformation. Die Schwierigkeit liegt nicht an der Darstellung, sondern an der Komplexität der Stichprobe, da drei Projekte gleichzeitig (nationales, kantonales und internationales Projekt) auf z.T. unterschiedliche Weise (dies gilt für die Zusatzstichproben in der deutschen und in der französischen Schweiz) erhoben wurden. In manchen Dokumenten werden alle Projekte gemeinsam beschrieben, andere Dokumentationsunterlagen beziehen sich nur auf dieses Teilprojekt.

Evaluation der Gesamtdokumentation

Die Gesamtdokumentation ist ausführlich.

Fragebögen in allen verwendeten Sprachen

Der SchülerInnen-Fragebogen ist vorhanden. Es ist derselbe, wie derjenige des gesamt nationalen Projektes.

Die Testhefte können nicht bei SIDOS bezogen werden. Es ist eine begründete Anfrage an die Direktion PISA.ch nötig, um sie zu erhalten.

Kontaktadressen : http://www.statistik.admin.ch/stat_ch/ber15/pisa/pisa_d_n002.htm oder Anfrage via vorbereitetes Formular SIDOS Webpage <http://www.sidos/data/projects/pisa/permission.doc>

Anweisungen/Schulungen für die Interviewer

Die Testadministratoren, welche in den Schulen die Tests durchgeführt haben und den Fragebogen ausfüllen liessen, erhielten eine Wegleitung mit den Anweisungen und den Texten, die obligatorisch vorgelesen werden müssen. Diese Manuals sind in allen drei Landessprachen der Dokumentation beigefügt.

Karten, Item-Liste, zusätzliches Material zum Fragebogen

Nicht relevant

Kodierlisten

Für die Variablen des Fragebogens: Eine Liste mit postkodierte Variablen, resp. Schweiz spezifischen Optionen der Erhebung sowie Schweiz spezifischen Klassierungen stellt das Dokument «Student Grade9 codebook Variable Info.pdf.» (in der « **Dokumentation BFS_French_Subsample**») dar. Sie wurde für den nationalen Datensatz erstellt. Doch der Datensatz der welschen Zusatzbefragung enthält die selben Variablen des Fragebogens.

Lieferung eines Codebuchs durch die Forscher

Zwei Arten von Codebooks sind in der « **Dokumentation BFS_French_Subsample**» enthalten:

- 1) Liste mit allen im nationalen Datensatz enthaltenen Variablen und ihren definierten Werte und Labels – kann auch benutzt werden für diesen Datensatz.

Student Grade9 codebook File Info.pdf

- 2) Beschreibung der Postkodierung von bestimmten Variablen des Fragebogens:

French Sample Codebook Variable Info.pdf.

B) Verbindung Erhebungsinstrument-Datensatz

Der Datensatz setzt sich aus Variablen zwei verschiedener Erhebungsinstrumente zusammen; aus dem Fragebogen und dem Test. Jene Variablen, welche sich auf die Testergebnisse beziehen, sind Konstruktionen. Die Test-Items selber stellen einen separaten Datensatz (FS Items Grade 9) dar.

Die Verbindung zwischen Erhebungsinstrument und Datensatz stellt für die Variablen des Fragebogen kein Problem dar, da die Kodierungen im Fragebogen eingetragen sind. Für die Verbindung zwischen Test und Datensatz muss auf das « PISA 200 DataManual.pdf » zurückgegriffen werden.

Zur Hilfe der Orientierung folgt hier ein grober Überblick über die Struktur des Datensatzes mit Verweis auf Beschreibung der Konstruktion der Variablen im « **PISA 2000 DataManualFrench** »

Folgend auf vier Identifikationsvariablen und einer Kontextvariablen stellen die ersten rund 250 Variablen des Datensatzes die codierten Antworten aus dem Fragebogen dar. Die Namen stehen in direktem Zusammenhang mit dem Fragebogen. Die Logik der Namensgebung der Variablen des Fragebogens ist auf Seite 10 beschrieben.

Es folgen Indikatoren, welche aus direkten Kombinationen aus den Antworten des Fragebogens gebildet wurden. Diese Variablen werden ab Seite 10 beschrieben.

Gewichte, werden im separaten File « Pondération_SR » erklärt.

Die weiteren Variablen sind nach den Antworten der Testfragen gebildet. Ihre Konstruktion ist ab Seite 18 beschrieben. Zu der Definition der Leistungsniveaus : Was die Levels inhaltlich bedeuten (wie gut man sein muss, um einem bestimmten Niveau zugeordnet zu werden) ist im Nationalen Bericht p25 beschrieben.

Reihenfolge der gruppierten Variablen des Datensatzes mit der Beschreibung aus dem DataManual :

Auf Seiten 4/5 des « PISA 2000 DataManualFrench» ist der Aufbau beschrieben, ohne dabei die Namen der Variablen zu nennen. Zur besseren Verständlichkeit und Übersichtlichkeit wird hier der Aufbau mit Hinweis auf Nummer und Anzahl der Variablen ergänzt (in Klammern : Seitenzahl im DataManualFrench.pdf , wo die Konstruktion wird) :

Identification variables:

V1 to V4: Identification variables for the school and student.

Contextual variables:

V5: Canton

Variables related to the questionnaire:

V6 to V253: **The student responses** on the three questionnaires, i.e., the student questionnaire and the two international options: computer familiarity questionnaire and cross curriculum competencies questionnaire (p10)

The student questionnaire indices: (p10)

V254 to V264: Indices directly included in the student questionnaire file (p11)

V265 to V296: Weighted likelihood estimates: Indices derived through a direct combination of the answers from the student questionnaire

V265 to V279: Indices from the student questionnaire (p14)

V280 to V282: Indices from the computer familiarity questionnaire (p16)

V283 to V296: Indices from the cross-curriculum competencies questionnaire (p16)

Weights:

V297 to V299: The student weights (p18)

Student performance scores:

V300 to V335: Two kinds of estimate derived from the test items:

Weighted likelihood estimate:

V300 to V313 (p18)

The plausible values: (p19)

V300 to V325: The students' performance scores in reading

V326 to V330: The students' performance scores in mathematics for each student who was assessed with one of the booklets that contain mathematics material

V331 to V335: The students' performance scores in science for each student who was assessed with one of the booklets that contain science material

Replicates: (p27: the construction of the Fay's replicates is explained, not the one of the Jackknife's replicates, because in the national file there are the 80 Fay's replicates integrated)

V336 to V731: The 132 reading, 132 mathematics and 132 science Jackknife's replicates for the computation of the sampling variance estimates.

Vorhandensein von Variablen- und Werte-Labels in der SPSS-Datei

Variablen- und Werte-Labels sind in der SPSS-Datei für die Variablen des Fragebogens integriert. Für die übrigen Variablen sind nur die Variablen-Labels eingefügt.

C) Datendatei(en)

Konvertierung der Datei(en)

Die Daten-Files wurden von SPSS.sav in SPSS.por konvertiert.

Probleme mit der Lesbarkeit der Dateien

Keine.

Globale Diagnosen

Missings sind nicht definiert, ebenso wenig die Filtersprünge.

Zahl der Fälle

4833

Zahl der Variablen

731

Konstruierte Variablen

Die Replikations-Variablen zur Schätzung der Stichprobenvarianz wurden in dieser Stichprobe nach einem anderen Verfahren gebildet als jenes, welches für den nationalen Datensatz verwendet wurde.

Die Variablen jkf001 bis jkf132 sind Replikationen nach dem Jackknife-Methode. Sie entsprechen den Variablen des nationalen Datensatzes w_fstr1 à w_fstr80 und beziehen sich auf die Lesekompetenz.

Die Variablen mrepl1 bis mrepl32 entsprechen den Variablen mrepl1 bis mrepl80 ("Math BRR Replicate" im nationalen Datensatz)

Die Variablen srepl1 bis srepl132 entsprechen den Variablen srepl1 bis srepl80 ("Science BRR Replicate" im nationalen Datensatz)

Kontextuelle Variablen

Die ersten vier Variablen `stidstrt`, `stidsch`, `stidstd`, `schoolid` sind Identifikationsvariablen und die Variable `canton` stellt eine Klassifizierung nach politischen oder sprachlichen Kriterien dar. Die Wertelabels sind im SPSS-File angegeben.

Gewichtungsvariable(n)

Es sind drei Gewichtungsvariablen (je eine für Lesen, Mathematik und Naturwissenschaften) im Datensatz enthalten. Ihre Konstruktion ist beschrieben im Dokument « Pondération_SR » im Ordner « Dokumentation Consortium romand ». Es handelt sich dabei um eine andere Konstruktion als jene der Gewichte des nationalen Datensatzes.

Bewertung des Zustands der Daten

Die Daten mussten nach einer ersten Kontrolle überarbeitet werden. Die Werte waren zu einem grossen Teil nicht analog der Angaben im Fragebogen kodiert. Dies wurde von SIDOS korrigiert.

Um eine möglichst grosse Übereinstimmung dieses Datensatzes mit jenem der nationalen Stichprobe zu erhalten, mussten verschiedene Variablen umcodiert oder zum Teil auch weggelassen werden. Letzteres auch aus dem Grund, weil wir für jene zusätzlichen Variablen über keine detaillierte Dokumentation verfügen.

KONTROLLE II: VARIABLEN

Systematische Kontrolle von Variable zu Variable

wurde nicht durchgeführt.

Fehlende Variablen?

Bei den Replicates fehlte im Originaldatensatz eine Variable für Mathematik und eine für Naturwissenschaften. Diese zwei Variablen wurden noch nachgeliefert.

`st41n01` bis `st41n06` : Diese Variablen wurden aus dem Datensatz herausgenommen, weil nicht mehr zu rekonstruieren war, in welchen Fällen die Schulnoten nach dem französischen System gegeben werden. Im Falle des nationalen Datensatzes existiert dieses Problem nicht, weil alle Privatschulen ausgeschlossen sind.

KONTROLLE III: WERTE

Eine systematische Kontrolle wurde nicht durchgeführt. Im Falle von 116 Variablen musste eine Recodierung vorgenommen werden, da die Werte gegenüber derjenigen im Fragebogen um eins verschoben eingegeben waren.

SIDOS hat von der Stelle, die für die Datenerhebung in der welschen Schweiz verantwortlich war, eine Datei erhalten, die offenbar für die Berechnung von Indizes benutzt wurde; die Skalen wurden in gewissen Variablen umgekehrt (S. den Technischen Bericht, S. 225 (PDF-Dokument) oder 229 (Seitenangabe im Dokument)), die Variablen- und Wertelabels aber nicht angepasst. Der ursprüngliche Zustand wurde wieder hergestellt.

Betroffene Variablen: `st31q01` `st31q04` `st31q06` `st35q01` `st35q04` `st35q06` `st35q08` `st35q09`.

Syntax:

```
recode st31q01 st31q04 st31q06 st35q01 st35q04 st35q06 st35q08 st35q09
```

(1=4) (2=3) (3=2) (4=1).

Vervollständigung der Wertelabels

Wurde in gewissen Fällen vorgenommen.

Rekodierungen von INAP

Die Missing Values wurden nicht rekodiert. Es sind jene Fälle, welche den Fragebogen nicht ausgefüllt haben, jedoch den Test mitgemacht haben.

Korrektion oder Rekodierungen von Wildcards

Wurde nicht durchgeführt.

Kohärenz mit veröffentlichten Resultaten

Wurde nicht kontrolliert.

Plausibilisierungstests

Wurde nicht durchgeführt.