

# **Quick Guide to Using the Program for International Student Assessment (PISA) 2006 Electronic Codebook (ECB) for U.S. Data**



# Contents

<b>A. Introduction to PISA 2006 .....</b>	<b>1</b>
A.1 Overview .....	1
A.2 Major Features.....	2
<b>B. Description of the PISA 2006 ECB.....</b>	<b>2</b>
B.1 Description of PISA 2006 Files.....	2
B.2 CD-ROM.....	3
<b>C. Getting Started.....</b>	<b>3</b>
C.1 Installing the PISA 2006 ECB .....	3
C.2 Using the PISA 2006 ECB .....	4
C.3 ID, Weighting, and Complex Sample Adjustment Variables .....	6
C.4 Plausible Values.....	7
<b>D. A Sample ECB Session .....</b>	<b>7</b>
D.1 Selecting Variables for Analysis From the ECB.....	7
D.2 Creating a SAS Working Data File.....	8
D.3 Creating an SPSS Working Data File.....	13
D.4 Merging Student and School Data .....	17
D.5 Merging Student and Assessment Data.....	19
<b>E. Frequently Asked Questions About PISA 2006.....</b>	<b>21</b>
E.1 General and Background Questions .....	22
E.2 Sampling.....	23
E.3 Weights.....	23
E.4 Electronic Codebooks.....	25
E.5 Privileged or Restricted-Use Data .....	26



# Quick Guide to Using the PISA 2006 Data

The purpose of this “Quick Guide” is to orient users of the Program for International Student Assessment (PISA) 2006 data to suggested techniques for working with the data files. Special attention will be paid to topics that will help users avoid the most commonly made mistakes in working with PISA 2006 data. The first section of this guide provides a brief overview of the PISA 2006 survey and available data files. The next three sections describe the PISA 2006 Electronic Codebook (ECB) and how to use it. The final section contains a series of frequently asked questions (FAQs) that NCES receives from users. The questions are followed by responses from NCES. The information found in this quick guide applies to the public-use version of the PISA 2006 ECB, unless otherwise noted.

Note that stand-alone SAS and SPSS program code is available on the NCES website (<http://nces.ed.gov/surveys/pisa>). These program code files will create SAS or SPSS data files containing *all* of the variables in the files described below. As such, the created files will be larger than a subset of variables extracted using the ECB. These code files can also serve as points of reference for questions about reading in the raw data files with the ECB-generated code.

## A. Introduction to PISA 2006

### A.1 Overview

PISA 2006 represents a major international effort to provide comparable data about 15-year-old students’ achievement in mathematics literacy, science literacy, and reading literacy across 57 countries or jurisdictions, including the United States. Each PISA data collection effort (prior collections occurred in 2000 and 2003) assesses one of the three subject areas in depth, with the other two subject areas assessed as minor subject areas during that data collection cycle. In PISA 2006, science literacy was the subject area assessed in depth. Mathematics literacy and reading literacy were the minor subjects, but because of an error in printing the test booklets in the United States, only science literacy and mathematics literacy results are reported for the United States.

PISA 2006 was a 2-hour paper-and-pencil assessment of 15-year-old students selected from nationally representative samples in participating jurisdictions. PISA 2006 was not designed to provide individual student scores, but rather national and group estimates of performance. In PISA 2006, every student answered science items. However, not every student answered reading and mathematics items, because these were distributed across different versions of the test booklets.

PISA 2006 was administered in the United States between September and November 2006. The U.S. sample included both public and private schools, randomly selected and weighted to be representative of the nation. In total, 166 schools and 5,611 students participated in PISA 2006 in the United States. The overall weighted school response rate was 69 percent before the use of replacement schools. The final weighted student response rate was 91 percent.

## A.2 Major Features

PISA 2006 provides data on science literacy, mathematics literacy, and reading literacy for internationally comparable samples of 15-year-old students. The study gathered background and contextual information from student, teacher, and school questionnaires. PISA 2006 uses plausible values as indicators of individual student performance and as the basis for national and group estimates. Individual item performance on the student assessments is reported along with overall performance in plausible values.

## B. Description of the PISA 2006 ECB

The ECB software tool enables analysts to review and extract U.S. PISA 2006 data. Specifically, with the ECB software tool, an analyst can

- search the names and labels of variables in the PISA 2006 data;
- examine the unweighted response categories, frequencies, and percentages of responses for one or more PISA 2006 variables;
- create a list of “tagged” variables (called a tag list or tag file) to be extracted from the PISA 2006 raw data files, save the tag list for later use, recall a previously saved tag list, print the tagged variables as a codebook, or print the tag list itself; and
- automatically generate SAS and SPSS programs to extract selected variables from the raw data files.

The ECB software tool works only on personal computers running a Windows-based environment (Windows 95 or higher). It will not run with other computer operating systems, such as Macintosh and Linux. The ECB software tool is included on this CD-ROM, along with the national data files (in ASCII format) and other files described below.

### B.1 Description of PISA 2006 Files

The PISA 2006 data files and ECB are available in both public- and restricted-use versions. For either version, certain naming conventions are often used. For example, the following identifies the file content:

STUD06	Student questionnaire and assessment outcomes
SCHL06	School questionnaire responses
ASSESM06	Assessment item-level responses

There are also several file types used with the ECB, identified by the file extension (the letters at the end of the file name, after a period):

.dat	Raw (text) data file
.sps	SPSS control file (reads raw data into SPSS)
.sas	SAS control file (reads raw data into SAS)
.txt	Codebook text file (in pdf format)
.tag	Tag file

In using the ECB, you may encounter or create files with such extensions and names. Note that the assessment files exist only as public-use versions.

## **B.2 CD-ROM**

The public-use datasets, ECB, and supporting documentation for PISA 2006 are located on one CD-ROM, NCES 2009-057. This data product contains

- ECB software (discussed above);
- an ECB help file (HELP.PDF) which explains the ECB program features in detail;
- PISA 2006 student, school, and assessment data;
- Quick Guide to Using PISA 2006 ECB; and
- an electronic copy of the User's Guide for the Program for International Student Assessment (PISA) (NCES 2009-055), which provides details about PISA 2006 design, data collection procedures, and final data files.

The restricted-use CD-ROM (NCES 2009-011) contains the ECB software, the ECB help file, and a restricted-use supplement to the PISA 2006 Data Analysis User's Guide. The data files on the restricted-use CD-ROM have only the restricted-use variables, which are not available in the public-use school data files. Public-use variables are not found on the restricted-use CD.

## **C. Getting Started**

This section addresses

1. what you need to know to get started using the PISA 2006 data;
2. how to navigate through the ECB; and
3. how to generate program code to transfer data into SAS or SPSS format.

### **C.1 Installing the PISA 2006 ECB**

#### ***Files on PISA 2006 CD-ROM***

The PISA 2006 CD-ROM contains three folders and two files in the root folder that are described below:

- The root folder includes this document (PISA\_QuickGuide.PDF); HELP.PDF; and folders ECBW, Report, and Tag.
- The "ECBW" folder includes three data files (student, school, and assessment), documentation for data files, and an installation program (Setup.exe) for the PISA 2006 ECB.
- The "Report" folder includes an electronic copy of the PISA 2006 Data Analysis User's Guide.
- The "Tag" folder includes four tag files (one for each file type listed in B.1, and an "All Files" tag file) that provide the user with preselected lists of critical variables (e.g.,

identification [ID] variables and weights), which can be imported into an ECB session and used as the basis for producing SPSS or SAS program code. Tag files are simply a subset of variables from the entire set of variables available on the ECB. Given the large number of variables on the ECB, tag files allow users to focus on those variables they select instead of having to sift through all the variables each time. The four tag files with preselected variables allow users to easily select critical identification and sampling variables before selecting other variables of interest.

### ***Installing and Loading the ECB***

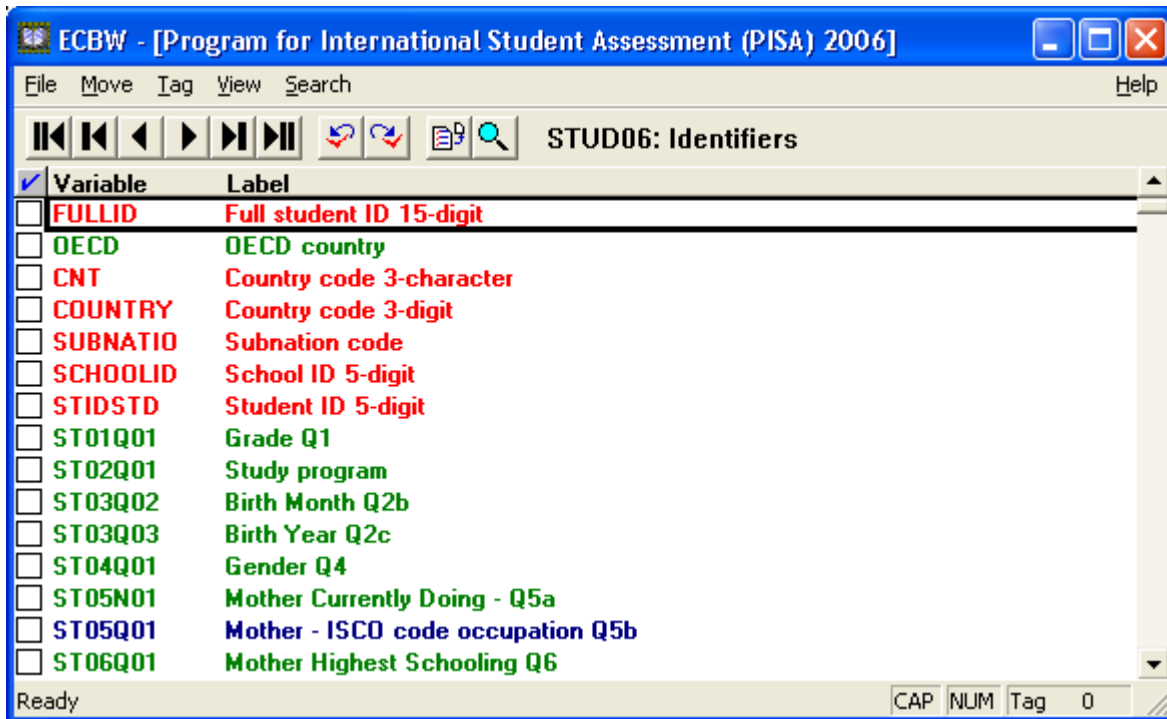
The PISA 2006 ECB will automatically install and launch on your computer when you run the ECB's file "Setup.exe." The following steps describe how to do this:

1. Insert the PISA 2006 CD into your computer's CD-ROM drive.
2. Click on the Windows START menu button, and select RUN.
3. Type D:\ECBW\Setup.exe (if your CD-ROM drive is not D, enter the appropriate drive letter). A setup program will launch and open a new window.
4. Click on OK to run the setup program and follow the directions on your screen. The ECB is ready to use once it is installed; the computer does not need to be rebooted upon installation.
5. Use the START menu in Windows to navigate to the Electronic Codebook | PISA 2006 (named "PS6" for the public-use version, "P6R" for the restricted-use version) program, or double-click on the PISA 2006 icon in the installed folder.
6. Read the "Warning" screen and respond. If you "agree," the main ECB window will appear.

## **C.2 Using the PISA 2006 ECB**

When the PISA 2006 ECB is correctly installed and operating, you should see an ECB window that looks like the screenshot on the following page, with a list of variable names and labels in red, blue, and green:





### *Navigating the List of Variables*

You can scroll through the list of variables using the scroll bar or arrow keys on your keyboard, or you can use the buttons on the left side of the toolbar (these appear as triangular arrows below the menu bar). These left-hand toolbar buttons take you to the top or bottom of different sections of the variable list. The outside buttons (triangle and two vertical bars) navigate to the top and bottom of the entire variable list. The next two buttons to the inside (triangle with one bar) take you to different files—that is, variables that are included on the student, school, or assessment files. The two center buttons (triangles) take you to sections within the student, school, or assessment files. You can check which file and section you are currently in by reading the location on the right-hand side of the toolbar.

You can also search for variables by clicking on “Search” on the menu bar at the top of the window. Using the Search function, you can search variable names, labels, descriptions, or category labels, as well as any combination of these. Double-clicking on a variable name or label will open a descriptive window for that variable, showing details about the item’s question wording, frequencies or ranges for that variable, and missing data codes.

You can “tag” variables by clicking on the box to the left of the variable name. You can then save selected lists of tagged variables, produce codebooks for variables in those lists, or produce SAS and SPSS code that will create SAS and SPSS data files for further analysis of your chosen variables.

### *Saving SAS or SPSS Program Files, Tag Files, or Codebook Text*

1. After tagging the variables of interest, go to “File” and then “Output.”
2. Select “SAS-PC,” “SPSS,” “Codebook Text,” or “Tag File.”

3. Specify the directory where you want to save the output file, and type a name for the file.
4. Click “Save” to save the file.
5. To view program code created with the “SAS-PC” or “SPSS” option, select “File” and then “View Output.”
6. The SAS or SPSS program code can then be opened in the appropriate software to generate a working data file in the respective software. It may be necessary to modify the program slightly (check for “execute” statements, period locations, and file names and locations). The code should identify the ASCII data file location on the CD-ROM.

Special steps may be necessary for working with preexisting tag files for ID and weighting variables (found in the folder “Tag” on the CD) because importing tag files will clear existing tags. Preexisting tag files can be included in user-generated tag files in two ways:

- Importing the preexisting tag file first, then selecting other variables of interest. This method is easiest if a user will be working in a single file (e.g., STUD06).
- Selecting other variables first, saving those variables as a tag file, and then using the Import Tag Files dialogue box to import both user-generated and preexisting tag files. After selecting both tag files, choose “OR” under “Combine Files With” in the dialogue box (choosing “AND” only imports variables tagged in both files).

These steps (saving a tag file and then combining the saved tag file with another preexisting tag file) may be repeated if multiple preexisting tag files (e.g., for both student background and school files) are to be combined with user-generated tag lists.

For further detailed information on using the ECB, see the ECB help file (HELP.PDF) located in the root folder of the CD. This file contains screenshots and specific directions for using all of the ECB’s functions.

### **C.3 ID, Weighting, and Complex Sample Adjustment Variables**

In analysis of PISA 2006 data, identification (ID) variables and weighting variables are of critical importance. ID variables enable the user to merge data with other files (such as merging student data with item-level assessment data) or to merge U.S. data with other countries’ or jurisdictions’ datasets. In the PISA 2006 ECB, selection of variables from different files (e.g., student-level and school-level variables) will produce SAS or SPSS program code that creates separate datasets for each file. The user must manually merge data from separate files, either using SCHOOLID to merge school and student data or using STIDSTD (full 15-digit student ID) to merge student- and item-level assessment data. See below for more information on merging data.

PISA 2006 includes two sets of weights: first, final weights for the student or school; and second, replicate weights for student-level estimates. Final weights (W\_FSTUWT for the student, W\_FSCHWT for the school) are necessary to produce correct estimates of student- or school-level continuous or categorical variables. Failing to use these weights will render results that reflect only the specific PISA 2006 sample and not the U.S. 15-year-old population that PISA 2006 was designed to represent.

There are 80 student replicate weights (W\_FSTR1–W\_FSTR80). These weights are necessary to provide correct estimates of the sampling variance (standard error) associated with any given point estimate (e.g., mean or frequency). Failing to use replicate weights may lead to incorrect inferences about differences among students or student subgroups.

SAS and SPSS software, as well as other statistical packages, have options for using final (total) weights and replicate weights.

Preselected tag lists of ID and weighting variables for the student file, the school file, the assessment file, and for all files combined are available in the “Tag” folder of the CD-ROM.

## **C.4 Plausible Values**

PISA 2006 contains plausible values for achievement scores. Plausible values are multiple estimates of individual student performance that enable group-level estimates of performance. Plausible values are used because PISA sample members did not take the full battery of assessment items (each student was given a subset of items). In the case of PISA 2006, five plausible values were computed for each student respondent, indicating possible “true” values of the student’s score on the underlying conceptual dimension (mathematics, science, or one of the science subscales). The plausible values are combined in producing estimates of group achievement.

To correctly produce estimates using plausible values, specific procedures must be followed. These procedures are part of program code written for SAS and SPSS and are available as macros for download from the NCES website. Examples of the initial program code (which calls the macro that performs the proper analysis) and the macros are available.

## **D. A Sample ECB Session**

### **D.1 Selecting Variables for Analysis From the ECB**

The script below provides an example of a session with the PISA 2006 ECB software. In this example, a user wants to create a file of variables associated with students’ science achievement and time spent doing science homework. The actions the user would perform are presented step by step.

Open the PISA 2006 ECB by clicking the Start button, going to Programs, selecting “Electronic Codebook for Windows,” and clicking on “PS6,” or by double-clicking on the PISA 2006 ECB icon on the computer desktop. Read and respond to the warning; if you click “Agree,” the main window will then appear.

Click the word “Search” on the right-hand side of the menu bar. Type HOMEWORK into the box, click “Description” in addition to the already selected “Variable,” and then click the “Tag All Match” button. Then click “Close.”

On the menu bar, select “View | Tagged Items.” A list of four variables will appear, identifying variables that record student responses to questions about homework in science, math, language, and other subjects. Click OK to close this box. Then click the “Next Tag” button on the toolbar (third from right). This will bring you to the first of the four tagged variables. The other three

will appear shortly below the first. Untag all but ST31Q03, since science homework is the main independent variable in which we are interested.

We know that the student file also contains the plausible values variables for the mathematics achievement scores, but we might not know the name of the variables. Search for this variable by opening up the Search window, typing PLAUSIBLE, and adding “Label” to the already selected “Variable” and “Description.” Then click “Search Next” and “Close.” This will bring you to the first of the five plausible values in math. Tag all five by selecting the check-box for each, or hold down the Shift key, select PV1MATH, then tag PV5MATH, which will tag all of them.

These six variables, in addition to automatically added ID and weighting variables (see next paragraph), will comprise the working dataset we will create from the raw PISA 2006 data. Since this set of variables may be needed in future analyses, it can be prudent to save the current set of tagged variables to a user-defined tag file called SCI\_HMWK. To do so, click “File | Output | Tag File” on the menu bar, and choose the location to which to save. Type SCI\_HMWK as the name and click Save.

Finally, before creating SAS or SPSS program code to generate the working datafiles, we need to import preselected tag lists of ID and weighting variables. Choose “File | Import Tag Files” from the menu bar (a warning will come up about clearing existing tags; since SCI\_HMWK has already been saved, you do not need to worry about this. Click “Yes.”). Browse for or type in the directory path for SCI\_HMWK in the File #1 box, then browse for or type in the directory path for the STUD06.tag file found in the “Tag” folder of the CD. Select option “OR” under “Combine Files With,” then click “Import.” Click “File | Output | Tag File” and save SCI\_HMWK.tag to the same location as the previous one, overwriting the existing file (you may give it a different name if you wish).

## **D.2 Creating a SAS Working Data File**

With the tag file saved, we now want to create program code to create a working data file. Here we provide an example in SAS; in the next section, we provide an example with SPSS. The SAS code will contain SAS statements that will extract only the currently tagged variables from the raw (text) PISA 2006 data file. We want to save these SAS program statements in a file named SCI\_HMWK.sas. To do so, click on “File | Output | SAS-PC” on the menu bar. Choose the directory location and type SCI\_HMWK to save. In the next dialogue box, a user typically will choose to label “Categorical” or “None” reserved code. Choosing “All” will label reserved code (e.g. codes for legitimate skips or other types of missing data) for all types of variables; choosing “Categorical” will label the reserved codes only for categorical variables; and choosing “None” will prevent reserved code labeling entirely.

SCI\_HMWK.sas can now be opened and run from within SAS to produce the SAS working data file. Quit the ECB and open the SAS statistical software package. Find the saved .sas file and open it in SAS.

The code that is produced in SCI\_HMWK.sas is reproduced below. A few lines of the code may require editing before the program can be successfully run. The possible edits and explanations are indicated below the code lines (note that these annotations will not appear in the SAS program itself).

```
LIBNAME PS6 'C:\ECBW\PS6';
```

The libname is assigned automatically by SAS with the directory path in which the program is saved. The libname PS6 may be changed to a different name, if desired.

```
DATA X1;
```

The DATA statement indicates that a temporary SAS working file (called X1) will be created; it will not be available after exiting the SAS program. If you wish to save a permanent copy of the SAS working file, type “PS6.” prior to X1. This will save a permanent copy to the directory identified in the earlier LIBNAME statement. If you want to save to a different directory, change the LIBNAME statement. You may wish to change the automatically assigned X1 name to something more memorable. You may also wish to change it because X1 is the first name always assigned by the ECB and may cause complications later with other extract files.

```
INFILE 'D:\ECBW\PS6\DATA\STUD06.DAT' LRECL=1024 PAD;
```

The INFILE statement indicates the location of the PISA 2006 raw data file from which the chosen variables will be taken. ECB-generated code will automatically choose the correct file and indicate the location of the CD drive. If your CD drive is not “D:\”, then you will need to change it.

If your tagged variables come from multiple files (student, school, or assessment), the ECB code will repeat these steps for each raw data file. For cases involving variables from the student and school raw data files, the SAS code, at the end of the program, will merge the two created files into a single working file. For cases involving variables from the student and assessment raw data files, you will have to tell the SAS program to merge the two output files (an example is provided after this program code).

```
INPUT
```

```
FULLID $ 1-15 COUNTRY $ 20-22 SUBNATIO $ 23-27  
SCHOOLID $ 28-32 STIDSTD $ 33-37 ST31Q03 258-258  
PV1MATH 661-669 .4 PV2MATH 670-678 .4 PV3MATH 679-687 .4  
PV4MATH 688-696 .4 PV5MATH 697-705 .4 / W_FSTUWT 28-36 .4  
W_FSTR1 37-45 .4 W_FSTR2 46-54 .4 W_FSTR3 55-63 .4  
W_FSTR4 64-72 .4 W_FSTR5 73-81 .4 W_FSTR6 82-90 .4  
W_FSTR7 91-99 .4 W_FSTR8 100-108 .4 W_FSTR9 109-117 .4  
W_FSTR10 118-126 .4 W_FSTR11 127-135 .4 W_FSTR12 136-144 .4  
W_FSTR13 145-153 .4 W_FSTR14 154-162 .4 W_FSTR15 163-171 .4  
W_FSTR16 172-180 .4 W_FSTR17 181-189 .4 W_FSTR18 190-198 .4  
W_FSTR19 199-207 .4 W_FSTR20 208-216 .4 W_FSTR21 217-225 .4  
W_FSTR22 226-234 .4 W_FSTR23 235-243 .4 W_FSTR24 244-252 .4  
W_FSTR25 253-261 .4 W_FSTR26 262-270 .4 W_FSTR27 271-279 .4  
W_FSTR28 280-288 .4 W_FSTR29 289-297 .4 W_FSTR30 298-306 .4  
W_FSTR31 307-315 .4 W_FSTR32 316-324 .4 W_FSTR33 325-333 .4  
W_FSTR34 334-342 .4 W_FSTR35 343-351 .4 W_FSTR36 352-360 .4  
W_FSTR37 361-369 .4 W_FSTR38 370-378 .4 W_FSTR39 379-387 .4  
W_FSTR40 388-396 .4 W_FSTR41 397-405 .4 W_FSTR42 406-414 .4
```

W\_FSTR43 415-423 .4 W\_FSTR44 424-432 .4 W\_FSTR45 433-441 .4  
 W\_FSTR46 442-450 .4 W\_FSTR47 451-459 .4 W\_FSTR48 460-468 .4  
 W\_FSTR49 469-477 .4 W\_FSTR50 478-486 .4 W\_FSTR51 487-495 .4  
 W\_FSTR52 496-504 .4 W\_FSTR53 505-513 .4 W\_FSTR54 514-522 .4  
 W\_FSTR55 523-531 .4 W\_FSTR56 532-540 .4 W\_FSTR57 541-549 .4  
 W\_FSTR58 550-558 .4 W\_FSTR59 559-567 .4 W\_FSTR60 568-576 .4  
 W\_FSTR61 577-585 .4 W\_FSTR62 586-594 .4 W\_FSTR63 595-603 .4  
 W\_FSTR64 604-612 .4 W\_FSTR65 613-621 .4 W\_FSTR66 622-630 .4  
 W\_FSTR67 631-639 .4 W\_FSTR68 640-648 .4 W\_FSTR69 649-657 .4  
 W\_FSTR70 658-666 .4 W\_FSTR71 667-675 .4 W\_FSTR72 676-684 .4  
 W\_FSTR73 685-693 .4 W\_FSTR74 694-702 .4 W\_FSTR75 703-711 .4  
 W\_FSTR76 712-720 .4 W\_FSTR77 721-729 .4 W\_FSTR78 730-738 .4  
 W\_FSTR79 739-747 .4 W\_FSTR80 748-756 .4 ;

LABEL

FULLID='Full student ID 15-digit'  
 COUNTRY='Country code 3-digit'  
 SUBNATIO='Subnation code'  
 SCHOOLID='School ID 5-digit'  
 STIDSTD='Student ID 5-digit'  
 ST31Q03='Self Study - Science Q31c'  
 PV1MATH='Plausible value in math'  
 PV2MATH='Plausible value in math'  
 PV3MATH='Plausible value in math'  
 PV4MATH='Plausible value in math'  
 PV5MATH='Plausible value in math'  
 W\_FSTUWT='FINAL STUDENT WEIGHT'  
 W\_FSTR1='FINAL STUDENT REPLICATE BRR-FAY WEIGHT1'  
 W\_FSTR2='FINAL STUDENT REPLICATE BRR-FAY WEIGHT2'  
 W\_FSTR3='FINAL STUDENT REPLICATE BRR-FAY WEIGHT3'  
 W\_FSTR4='FINAL STUDENT REPLICATE BRR-FAY WEIGHT4'  
 W\_FSTR5='FINAL STUDENT REPLICATE BRR-FAY WEIGHT5'  
 W\_FSTR6='FINAL STUDENT REPLICATE BRR-FAY WEIGHT6'  
 W\_FSTR7='FINAL STUDENT REPLICATE BRR-FAY WEIGHT7'  
 W\_FSTR8='FINAL STUDENT REPLICATE BRR-FAY WEIGHT8'  
 W\_FSTR9='FINAL STUDENT REPLICATE BRR-FAY WEIGHT9'  
 W\_FSTR10='FINAL STUDENT REPLICATE BRR-FAY WEIGHT10'  
 W\_FSTR11='FINAL STUDENT REPLICATE BRR-FAY WEIGHT11'  
 W\_FSTR12='FINAL STUDENT REPLICATE BRR-FAY WEIGHT12'  
 W\_FSTR13='FINAL STUDENT REPLICATE BRR-FAY WEIGHT13'  
 W\_FSTR14='FINAL STUDENT REPLICATE BRR-FAY WEIGHT14'  
 W\_FSTR15='FINAL STUDENT REPLICATE BRR-FAY WEIGHT15'  
 W\_FSTR16='FINAL STUDENT REPLICATE BRR-FAY WEIGHT16'  
 W\_FSTR17='FINAL STUDENT REPLICATE BRR-FAY WEIGHT17'  
 W\_FSTR18='FINAL STUDENT REPLICATE BRR-FAY WEIGHT18'  
 W\_FSTR19='FINAL STUDENT REPLICATE BRR-FAY WEIGHT19'

W\_FSTR20= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT20'  
W\_FSTR21= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT21'  
W\_FSTR22= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT22'  
W\_FSTR23= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT23'  
W\_FSTR24= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT24'  
W\_FSTR25= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT25'  
W\_FSTR26= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT26'  
W\_FSTR27= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT27'  
W\_FSTR28= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT28'  
W\_FSTR29= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT29'  
W\_FSTR30= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT30'  
W\_FSTR31= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT31'  
W\_FSTR32= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT32'  
W\_FSTR33= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT33'  
W\_FSTR34= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT34'  
W\_FSTR35= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT35'  
W\_FSTR36= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT36'  
W\_FSTR37= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT37'  
W\_FSTR38= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT38'  
W\_FSTR39= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT39'  
W\_FSTR40= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT40'  
W\_FSTR41= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT41'  
W\_FSTR42= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT42'  
W\_FSTR43= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT43'  
W\_FSTR44= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT44'  
W\_FSTR45= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT45'  
W\_FSTR46= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT46'  
W\_FSTR47= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT47'  
W\_FSTR48= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT48'  
W\_FSTR49= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT49'  
W\_FSTR50= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT50'  
W\_FSTR51= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT51'  
W\_FSTR52= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT52'  
W\_FSTR53= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT53'  
W\_FSTR54= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT54'  
W\_FSTR55= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT55'  
W\_FSTR56= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT56'  
W\_FSTR57= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT57'  
W\_FSTR58= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT58'  
W\_FSTR59= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT59'  
W\_FSTR60= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT60'  
W\_FSTR61= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT61'  
W\_FSTR62= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT62'  
W\_FSTR63= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT63'  
W\_FSTR64= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT64'  
W\_FSTR65= 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT65'

```

W_FSTR66='FINAL STUDENT REPLICATE BRR-FAY WEIGHT66'
W_FSTR67='FINAL STUDENT REPLICATE BRR-FAY WEIGHT67'
W_FSTR68='FINAL STUDENT REPLICATE BRR-FAY WEIGHT68'
W_FSTR69='FINAL STUDENT REPLICATE BRR-FAY WEIGHT69'
W_FSTR70='FINAL STUDENT REPLICATE BRR-FAY WEIGHT70'
W_FSTR71='FINAL STUDENT REPLICATE BRR-FAY WEIGHT71'
W_FSTR72='FINAL STUDENT REPLICATE BRR-FAY WEIGHT72'
W_FSTR73='FINAL STUDENT REPLICATE BRR-FAY WEIGHT73'
W_FSTR74='FINAL STUDENT REPLICATE BRR-FAY WEIGHT74'
W_FSTR75='FINAL STUDENT REPLICATE BRR-FAY WEIGHT75'
W_FSTR76='FINAL STUDENT REPLICATE BRR-FAY WEIGHT76'
W_FSTR77='FINAL STUDENT REPLICATE BRR-FAY WEIGHT77'
W_FSTR78='FINAL STUDENT REPLICATE BRR-FAY WEIGHT78'
W_FSTR79='FINAL STUDENT REPLICATE BRR-FAY WEIGHT79'
W_FSTR80='FINAL STUDENT REPLICATE BRR-FAY WEIGHT80';

PROC FORMAT;
VALUE $COUNTRYF      '840'='United States';
VALUE $SUBNATIF      '84000'='United States';
VALUE ST31Q03F      1='No time'
                    2='Less than 2 hours'
                    3='2 up to 4 hours'
                    4='4 up to 6 hours'
                    5='6 or more hours'
                    7='NA'
                    8='Invalid'
                    9='Missing';

PROC FREQ;
FORMAT
  COUNTRY $COUNTRYF. SUBNATIO $SUBNATIF. ST31Q03 ST31Q03F. ;

TABLES
  COUNTRY SUBNATIO ST31Q03 ;

PROC UNIVARIATE;VAR
  PV1MATH PV2MATH PV3MATH PV4MATH PV5MATH W_FSTUWT
  W_FSTR1 W_FSTR2 W_FSTR3 W_FSTR4 W_FSTR5 W_FSTR6
  W_FSTR7 W_FSTR8 W_FSTR9 W_FSTR10 W_FSTR11 W_FSTR12
  W_FSTR13 W_FSTR14 W_FSTR15 W_FSTR16 W_FSTR17
  W_FSTR18 W_FSTR19 W_FSTR20 W_FSTR21 W_FSTR22
  W_FSTR23 W_FSTR24 W_FSTR25 W_FSTR26 W_FSTR27
  W_FSTR28 W_FSTR29 W_FSTR30 W_FSTR31 W_FSTR32
  W_FSTR33 W_FSTR34 W_FSTR35 W_FSTR36 W_FSTR37
  W_FSTR38 W_FSTR39 W_FSTR40 W_FSTR41 W_FSTR42
  W_FSTR43 W_FSTR44 W_FSTR45 W_FSTR46 W_FSTR47

```



```

W_FSTR48 W_FSTR49 W_FSTR50 W_FSTR51 W_FSTR52
W_FSTR53 W_FSTR54 W_FSTR55 W_FSTR56 W_FSTR57
W_FSTR58 W_FSTR59 W_FSTR60 W_FSTR61 W_FSTR62
W_FSTR63 W_FSTR64 W_FSTR65 W_FSTR66 W_FSTR67
W_FSTR68 W_FSTR69 W_FSTR70 W_FSTR71 W_FSTR72
W_FSTR73 W_FSTR74 W_FSTR75 W_FSTR76 W_FSTR77
W_FSTR78 W_FSTR79 W_FSTR80 ;

```

RUN;

This is the end of the file. Other than the changes to the file locations and output file name, no changes are usually required.

### D.3 Creating an SPSS Working Data File

With the tag file saved, we now want to create program code to create a working data file. Here, we provide an example with SPSS. The SPSS code will contain SPSS statements that will extract only the currently tagged variables from the raw (text) PISA 2006 data file. We want to save these SPSS program statements in a file named SCI\_HMWK.sps. To do so, click on “File | Output | SPSS” on the menu bar. Choose the directory location and type SCI\_HMWK to save. In the next dialogue box, a user typically will choose to label “Categorical” or “None” reserved code. Choosing “All” will label reserved code (i.e., codes for legitimate skips or other types of missing data) for all types of variables; choosing “Categorical” will label the reserved codes only for categorical variables; and choosing “None” will prevent reserved code labeling entirely.

SCI\_HMWK.sps can now be opened and run from within SPSS to produce the SPSS working data file. Quit the ECB and open the SPSS statistical software package. Find the saved .sps file and open it in SPSS.

The code that is produced in SCI\_HMWK.sps is reproduced below. A few lines of the code may require editing before the program can be successfully run. The possible edits and explanations are indicated below the code lines (note that these annotations will not appear in the SPSS program itself).

```

DATA LIST FILE='D:\ECBW\PS6\DATA\STUD06.DAT' FIXED RECORDS=2 /
FULLID 1-15 (A) COUNTRY 20-22 (A) SUBNATIO 23-27 (A)
SCHOOLID 28-32 (A) STIDSTD 33-37 (A) ST31Q03 258-258
PV1MATH 661-669 (4) PV2MATH 670-678 (4) PV3MATH 679-687 (4)
PV4MATH 688-696 (4) PV5MATH 697-705 (4) /2 W_FSTUWT 28-36 (4)
W_FSTR1 37-45 (4) W_FSTR2 46-54 (4) W_FSTR3 55-63 (4)
W_FSTR4 64-72 (4) W_FSTR5 73-81 (4) W_FSTR6 82-90 (4)
W_FSTR7 91-99 (4) W_FSTR8 100-108 (4) W_FSTR9 109-117 (4)
W_FSTR10 118-126 (4) W_FSTR11 127-135 (4) W_FSTR12 136-144 (4)
W_FSTR13 145-153 (4) W_FSTR14 154-162 (4) W_FSTR15 163-171 (4)
W_FSTR16 172-180 (4) W_FSTR17 181-189 (4) W_FSTR18 190-198 (4)
W_FSTR19 199-207 (4) W_FSTR20 208-216 (4) W_FSTR21 217-225 (4)
W_FSTR22 226-234 (4) W_FSTR23 235-243 (4) W_FSTR24 244-252 (4)
W_FSTR25 253-261 (4) W_FSTR26 262-270 (4) W_FSTR27 271-279 (4)

```

W_FSTR28	280-288	(4)	W_FSTR29	289-297	(4)	W_FSTR30	298-306	(4)
W_FSTR31	307-315	(4)	W_FSTR32	316-324	(4)	W_FSTR33	325-333	(4)
W_FSTR34	334-342	(4)	W_FSTR35	343-351	(4)	W_FSTR36	352-360	(4)
W_FSTR37	361-369	(4)	W_FSTR38	370-378	(4)	W_FSTR39	379-387	(4)
W_FSTR40	388-396	(4)	W_FSTR41	397-405	(4)	W_FSTR42	406-414	(4)
W_FSTR43	415-423	(4)	W_FSTR44	424-432	(4)	W_FSTR45	433-441	(4)
W_FSTR46	442-450	(4)	W_FSTR47	451-459	(4)	W_FSTR48	460-468	(4)
W_FSTR49	469-477	(4)	W_FSTR50	478-486	(4)	W_FSTR51	487-495	(4)
W_FSTR52	496-504	(4)	W_FSTR53	505-513	(4)	W_FSTR54	514-522	(4)
W_FSTR55	523-531	(4)	W_FSTR56	532-540	(4)	W_FSTR57	541-549	(4)
W_FSTR58	550-558	(4)	W_FSTR59	559-567	(4)	W_FSTR60	568-576	(4)
W_FSTR61	577-585	(4)	W_FSTR62	586-594	(4)	W_FSTR63	595-603	(4)
W_FSTR64	604-612	(4)	W_FSTR65	613-621	(4)	W_FSTR66	622-630	(4)
W_FSTR67	631-639	(4)	W_FSTR68	640-648	(4)	W_FSTR69	649-657	(4)
W_FSTR70	658-666	(4)	W_FSTR71	667-675	(4)	W_FSTR72	676-684	(4)
W_FSTR73	685-693	(4)	W_FSTR74	694-702	(4)	W_FSTR75	703-711	(4)
W_FSTR76	712-720	(4)	W_FSTR77	721-729	(4)	W_FSTR78	730-738	(4)
W_FSTR79	739-747	(4)	W_FSTR80	748-756	(4)			

The DATA LIST statement indicates the location of the PISA 2006 raw data file from which the chosen variables will be taken. ECB-generated code will automatically choose the correct file and indicate the location of the CD drive. If your CD drive is not "D:\", then you will need to change it.

If your tagged variables come from multiple files (student, school, or assessment), the ECB code will repeat programming steps for each raw data file. For cases involving variables from the student and school raw data files, the SPSS code, at the end of the program, will merge the two created files into a single working file. For cases involving variables from the student and assessment raw data files, you will have to tell the SPSS program to merge the two output files (an example is provided after this program code).

#### VARIABLE LABELS

```

FULLID 'Full student ID 15-digit'
COUNTRY 'Country code 3-digit'
SUBNATIO 'Subnation code'
SCHOOLID 'School ID 5-digit'
STIDSTD 'Student ID 5-digit'
ST31Q03 'Self Study - Science Q31c'
PV1MATH 'Plausible value in math'
PV2MATH 'Plausible value in math'
PV3MATH 'Plausible value in math'
PV4MATH 'Plausible value in math'
PV5MATH 'Plausible value in math'
W_FSTUWT 'FINAL STUDENT WEIGHT'
W_FSTR1 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT1'
W_FSTR2 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT2'

```

W\_FSTR3 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT3'  
W\_FSTR4 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT4'  
W\_FSTR5 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT5'  
W\_FSTR6 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT6'  
W\_FSTR7 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT7'  
W\_FSTR8 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT8'  
W\_FSTR9 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT9'  
W\_FSTR10 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT10'  
W\_FSTR11 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT11'  
W\_FSTR12 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT12'  
W\_FSTR13 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT13'  
W\_FSTR14 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT14'  
W\_FSTR15 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT15'  
W\_FSTR16 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT16'  
W\_FSTR17 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT17'  
W\_FSTR18 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT18'  
W\_FSTR19 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT19'  
W\_FSTR20 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT20'  
W\_FSTR21 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT21'  
W\_FSTR22 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT22'  
W\_FSTR23 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT23'  
W\_FSTR24 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT24'  
W\_FSTR25 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT25'  
W\_FSTR26 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT26'  
W\_FSTR27 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT27'  
W\_FSTR28 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT28'  
W\_FSTR29 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT29'  
W\_FSTR30 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT30'  
W\_FSTR31 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT31'  
W\_FSTR32 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT32'  
W\_FSTR33 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT33'  
W\_FSTR34 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT34'  
W\_FSTR35 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT35'  
W\_FSTR36 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT36'  
W\_FSTR37 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT37'  
W\_FSTR38 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT38'  
W\_FSTR39 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT39'  
W\_FSTR40 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT40'  
W\_FSTR41 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT41'  
W\_FSTR42 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT42'  
W\_FSTR43 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT43'  
W\_FSTR44 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT44'  
W\_FSTR45 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT45'  
W\_FSTR46 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT46'  
W\_FSTR47 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT47'  
W\_FSTR48 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT48'

W\_FSTR49 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT49'  
 W\_FSTR50 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT50'  
 W\_FSTR51 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT51'  
 W\_FSTR52 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT52'  
 W\_FSTR53 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT53'  
 W\_FSTR54 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT54'  
 W\_FSTR55 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT55'  
 W\_FSTR56 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT56'  
 W\_FSTR57 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT57'  
 W\_FSTR58 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT58'  
 W\_FSTR59 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT59'  
 W\_FSTR60 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT60'  
 W\_FSTR61 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT61'  
 W\_FSTR62 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT62'  
 W\_FSTR63 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT63'  
 W\_FSTR64 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT64'  
 W\_FSTR65 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT65'  
 W\_FSTR66 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT66'  
 W\_FSTR67 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT67'  
 W\_FSTR68 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT68'  
 W\_FSTR69 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT69'  
 W\_FSTR70 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT70'  
 W\_FSTR71 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT71'  
 W\_FSTR72 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT72'  
 W\_FSTR73 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT73'  
 W\_FSTR74 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT74'  
 W\_FSTR75 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT75'  
 W\_FSTR76 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT76'  
 W\_FSTR77 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT77'  
 W\_FSTR78 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT78'  
 W\_FSTR79 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT79'  
 W\_FSTR80 'FINAL STUDENT REPLICATE BRR-FAY WEIGHT80'

VALUE LABELS

/ COUNTRY '840' 'United States'  
 / SUBNATIO '84000' 'United States'  
 / ST31Q03 1 'No time'  
 2 'Less than 2 hours'  
 3 '2 up to 4 hours'  
 4 '4 up to 6 hours'  
 5 '6 or more hours'  
 7 'NA'  
 8 'Invalid'  
 9 'Missing'

SAVE OUTFILE='C:\ECBW\PS6\X1' /Compressed.

The SAVE OUTFILE statement indicates where a copy of the SPSS data file will be saved (unlike SAS, this automatically saves a permanent copy), as well as the name of that file (here, X1). If you want to save to a different directory, change the directory information here. You may wish to change the automatically assigned X1 name to something more memorable. You may also wish to change it because X1 is the first name always assigned by the ECB and may cause complications later with other extract files.

```
FREQUENCIES VARIABLES=  
  COUNTRY SUBNATIO ST31Q03 .
```

```
DESCRIPTIVES VARIABLES=  
  PV1MATH PV2MATH PV3MATH PV4MATH PV5MATH W_FSTUWT  
  W_FSTR1 W_FSTR2 W_FSTR3 W_FSTR4 W_FSTR5 W_FSTR6  
  W_FSTR7 W_FSTR8 W_FSTR9 W_FSTR10 W_FSTR11 W_FSTR12  
  W_FSTR13 W_FSTR14 W_FSTR15 W_FSTR16 W_FSTR17  
  W_FSTR18 W_FSTR19 W_FSTR20 W_FSTR21 W_FSTR22  
  W_FSTR23 W_FSTR24 W_FSTR25 W_FSTR26 W_FSTR27  
  W_FSTR28 W_FSTR29 W_FSTR30 W_FSTR31 W_FSTR32  
  W_FSTR33 W_FSTR34 W_FSTR35 W_FSTR36 W_FSTR37  
  W_FSTR38 W_FSTR39 W_FSTR40 W_FSTR41 W_FSTR42  
  W_FSTR43 W_FSTR44 W_FSTR45 W_FSTR46 W_FSTR47  
  W_FSTR48 W_FSTR49 W_FSTR50 W_FSTR51 W_FSTR52  
  W_FSTR53 W_FSTR54 W_FSTR55 W_FSTR56 W_FSTR57  
  W_FSTR58 W_FSTR59 W_FSTR60 W_FSTR61 W_FSTR62  
  W_FSTR63 W_FSTR64 W_FSTR65 W_FSTR66 W_FSTR67  
  W_FSTR68 W_FSTR69 W_FSTR70 W_FSTR71 W_FSTR72  
  W_FSTR73 W_FSTR74 W_FSTR75 W_FSTR76 W_FSTR77  
  W_FSTR78 W_FSTR79 W_FSTR80  
/STATS=DEF.
```

This is the end of the file. Other than the changes to the file locations and output file name, no changes are usually required.

#### **D.4 Merging Student and School Data**

If you tag variables from the student background file (STUD06) and the school file (SCHL06), the SAS and SPSS program code that creates the working data files will create separate student and school data files. To merge these files and save a new output file containing information from both (with school data replicated for each student), add the code described below at the end of the respective SAS or SPSS program codes. This step is also described in the header section (surrounded by asterisks) at the beginning of the program code. Recall that important ID and weighting variables can be included by using the preselected tag files for each file, namely STUD06.TAG and SCHL06.TAG, found in the “Tag” folder of the CD-ROM. Combine these tag files with other tagged variables of interest (as described in section C.2) before merging the separate files.

To merge student and school data in SAS. New code will need to be added after the first “RUN;” line of the automatically generated ECB code. The automatically generated ECB code will have the following structure:

```
LIBNAME PS6 'C:\ECBW\PS6';

DATA X1; INFILE 'D:\ECBW\PS6\DATA\STUD06.DAT' LRECL=1024 PAD; INPUT
  FULLID $ 1-15 SCHOOLID $ 28-32 STIDSTD $ 33-37
  PV2READ 715-723 / ;
```

Next will be other student file statements (e.g., formats, frequencies). Following these will be

```
DATA X2; INFILE 'D:\ECBW\PS6\DATA\SCHL06.DAT' LRECL=1024 PAD; INPUT
  SCHOOLID $ 13-17 SC01Q01 18-22 SC01Q02 23-27 SC02Q01 28-28;
```

Next will be other school file statements (e.g., formats, frequencies). Following these will be

```
RUN;
```

This marks the end of the automatically generated ECB code.

Add the following code to the end of the automatically generated ECB code to merge the school data to the student file. This will create a permanent dataset named “EXTRACT” in the folder referenced by PS6 at the beginning of the automatically generated code. You may change the “EXTRACT” name to something more specific.

```
DATA PS6.EXTRACT; MERGE X1 X2;
  BY SCHOOLID;

RUN;
```

To merge student and school data in SPSS. New code will need to be added after the second “SAVE OUTFILE...” line of the automatically generated ECB code. The automatically generated ECB code will have the following structure:

```
DATA LIST FILE='D:\ECBW\PS6\DATA\STUD06.DAT' FIXED RECORDS=2 /
  FULLID 1-15 (A) SCHOOLID 28-32 (A) STIDSTD 33-37 (A)
  PV2READ 715-723
```

Next will be variable and value label statements for the student data. Following these will be

```
SAVE OUTFILE='C:\ECBW\PS6\X1' /Compressed.
```

Next will be frequencies statements for the student data. Following these will be

```
DATA LIST FILE='D:\ECBW\PS6\DATA\SCHL06.DAT' FIXED RECORDS=1 / SCHOOLID
  13-17 (A) SC01Q01 18-22 SC01Q02 23-27 SC02Q01 28-28
```

Next will be variable and value label statements for the school data. Following these will be

```
SAVE OUTFILE='C:\ECBW\PS6\X2' /Compressed.
```

Next will be frequencies statements for the school data, which mark the end of the automatically generated ECB code.

Add the following code to the end of the automatically generated ECB code to merge the school data to the student file. The command with “/TABLE” should reference the school data (here, “X2”). Note again that you may change the outfile location. You may also change the “XTRACT” name (the output file name) to something else.

```
MATCH FILES /FILE='C:\ECBW\PS6\X1'  
  /TABLE='C:\ECBW\PS6\X2'  
  /BY SCHOOLID.  
EXECUTE.  
  
SAVE OUTFILE='C:\ECBW\PS6\XTRACT' /Compressed.
```

## D.5 Merging Student and Assessment Data

If you tag variables from the student background file (STUD06) and the item-level assessment file (ASSESM06), the SAS and SPSS program code that creates the working data files will create separate student and assessment data files. To merge these files, add the code described below at the end of the respective SAS or SPSS program codes. This step is also described in the header section (surrounded by asterisks) at the beginning of the program code. Recall that important ID and weighting variables can be included by using the preselected tag files for each file, namely STUD06.TAG and ASSESM06.TAG, found in the “Tag” folder of the CD-ROM. Combine these tag files with other tagged variables of interest (as described in section C.2) before merging the separate files.

To merge student and assessment data in SAS. New code will need to be added after the first “RUN;” line of the automatically generated ECB code. The automatically generated ECB code will have the following structure (note that only ID variables appear in this example):

```
LIBNAME PS6 'C:\ECBW\PS6';  
  
DATA X1; INFILE 'D:\ECBW\PS6\DATA\STUD06.DAT' LRECL=1024 PAD; INPUT  
  FULLID $ 1-15 SCHOOLID $ 28-32 STIDSTD $ 33-37  
  / ;
```

Next will be other assessment file statements (e.g., formats, frequencies). Following these will be

```
DATA X3; INFILE 'D:\ECBW\PS6\DATA\ASSESM06.DAT' LRECL=1024 PAD; INPUT  
  SUBNATIO $ 1-5 SCHOOLID $ 6-10 STIDSTD $ 11-15;
```

Next will be other student file statements (e.g., formats, frequencies). Following these will be

```
RUN;
```

This marks the end of the automatically generated ECB code.

Add the following code to the end of the automatically generated ECB code to merge the assessment results to the student file. This will create a permanent dataset named “EXTRACT” in the folder referenced by PS6 at the beginning of the automatically generated code. You may change the “EXTRACT” name to something more specific.

```
PROC SORT DATA=X3;  
BY STIDSTD;  
RUN;  
  
DATA PS6.EXTRACT; MERGE X1 X3;  
BY STIDSTD;  
  
RUN;
```

To merge student and assessment data in SPSS. New code will need to be added after the second “SAVE OUTFILE...” line of the automatically generated ECB code. The automatically generated ECB code will have the following structure (note that only ID variables appear in this example):

```
DATA LIST FILE='D:\ECBW\PS6\DATA\STUD06.DAT' FIXED RECORDS=2 /  
FULLID 1-15 (A) SCHOOLID 28-32 (A) STIDSTD 33-37 (A)
```

Next will be variable and value label statements for the student data. Following these will be

```
SAVE OUTFILE='C:\ECBW\PS6\X1' /Compressed.
```

Next will be frequencies statements for the assessment data. Following these will be

```
DATA LIST FILE='D:\ECBW\PS6\DATA\ASSESM06.DAT' FIXED RECORDS=1 /  
SUBNATIO 1-5 (A) SCHOOLID 6-10 (A) STIDSTD 11-15 (A)
```

Next will be variable and value label statements for the assessment data. Following these will be

```
SAVE OUTFILE='C:\ECBW\PS6\X3' /Compressed.
```

This will be followed by frequencies statements for the student data, which mark the end of the automatically generated ECB code.

Add the following code to the end of the automatically generated ECB code to merge the assessment results to the student file. The command “/TABLE” should refer to the student file (here, “X1”)—note that this is different than when merging the student and school data in SPSS as described in section D.4. Again, you may change the outfile location. You may also change the “XTRACT” name to something more specific.

```
GET FILE='C:\ECBW\PS6\X3'.  
SORT CASES BY STIDSTD.  
SAVE OUTFILE='C:\ECBW\PS6\X3'.
```



MATCH FILES

```
/FILE='C:\ECBW\P6S\X3'  
/TABLE='C:\ECBW\P6S\X1'  
/BY STIDSTD.  
EXECUTE.
```

```
SAVE OUTFILE='C:\ECBW\P6S\XTRACT' /Compressed.
```

## **E. Frequently Asked Questions About PISA 2006**

NCES staff members have received many questions regarding proper techniques for working with PISA data. In this document, these questions (along with NCES responses) have been categorized into topical areas and presented as responses to FAQs. It is hoped that the responses will help users avoid the most commonly made mistakes in working with these data. This document is meant to serve as an introduction or supplement, not a replacement, for other PISA 2006 documentation. To help the data user identify specific topics of interest, questions and responses have been grouped into the following categories:

### **E.1 General and Background Questions**

- Whom can I contact from the National Center for Education Statistics/Department of Education about the PISA 2006 study?
- What are the interrelationships among the different PISA 2006 files?
- How is PISA 2006 related to other international studies?

### **E.2 Sampling**

- In simple terms, explain how the PISA 2006 school and student samples were selected.

### **E.3 Weights**

- What are weights?
- Why do we need to use weights with the PISA 2006 data?
- Why would unweighted estimates not be representative?
- Why do I need to use replicate weights when I compute means, frequencies, etc.?

### **E.4 Electronic Codebooks**

- When I receive my PISA 2006 CD, what are some of the steps that I should follow to check out my CD?

### **E.5 Privileged or Restricted-Use Data**

- When might I need a restricted-use license, and how do I get one?

## **E.1 General and Background Questions**

### **Whom can I contact from the National Center for Education Statistics/Department of Education about the PISA 2006 study?**

For additional information and questions about PISA 2006 and NCES's international activities program, please link to the PISA 2006 web page:  
<http://nces.ed.gov/surveys/pisa>.

Or contact:  
Holly Xie  
(202) 502-7314  
[holly.xie@ed.gov](mailto:holly.xie@ed.gov)

### **What are the interrelationships among the different PISA 2006 files?**

Student questionnaire data and assessment scores are the focal point of the PISA 2006 data files. Student questionnaire responses and estimates of overall student achievement (plausible values) are represented in one file. In this file, each record or case is a single student.

Assessment data are the result of student performance on the PISA assessment. The assessment data file contains responses for each item (i.e., test question) that a student answered. Like the student file, there is one record for each student in the assessment data file. Most users will not need to access this file since plausible values for all scale scores are included in the student file.

School questionnaires were given to school administrators in the schools of the PISA 2006 students. Because PISA 2006 sampled multiple students from each school, the same school questionnaire data apply to all students in the same school. School data can be thought of as providing contextual data about student environments.

Student, school, and assessment data are related by common ID variables. STIDSTD is the student ID variable that permits combining student and assessment data; SCHOOLID is the unique (within the United States) school ID that ties students to schools. Student and school questionnaires, as well as sample assessment items, are available on the NCES PISA website.

### **How is PISA 2006 related to other international studies?**

PISA is a set of international assessments that focuses on 15-year-old students' capabilities in reading literacy, mathematics literacy, and science literacy. PISA is organized by the Organization for Economic Cooperation and Development (OECD), an intergovernmental organization of industrialized countries. Begun in 2000, PISA is administered every 3 years.

The United States participates in two other major international assessments: The Progress in International Reading Literacy Study (PIRLS) and The Trends in International Mathematics and Science Study (TIMSS).

PIRLS is an international comparative study of the reading literacy of young students. Specifically, PIRLS studies the reading achievement and reading behaviors and attitudes of

fourth-grade students in the United States and students in the equivalent of fourth grade in other participating countries. While PISA is age-based and focuses on three subjects in rotation, PIRLS is grade-based and focuses on reading skills.

TIMSS provides data on the mathematics and science achievement of U.S. students compared with that of students in other countries at the fourth- and eighth-grade levels. In contrast to PISA, which includes applied knowledge and practical skills that could have been learned outside of school, TIMSS is curriculum-based and focuses on math and science abilities related to in-school learning. TIMSS data were collected in 1995, 1999, 2003, and 2007.

## **E.2 Sampling**

**In simple terms, explain how the PISA 2006 school and student samples were selected.**

The U.S. sample for PISA 2006 followed the model used in PISA 2003, which was a two-stage sampling process with the first stage a sample of schools and the second stage a sample of students within schools. The PISA school sample was stratified into two explicit groups: large schools and small schools. The frame was implicitly stratified (i.e., sorted for sampling) by five categorical stratification variables: grade span of the school (five levels), control of the school (public or private), region of the country (Northeast, Central, West, Southeast), type of location relative to populous areas (eight levels), and proportion of non-White students (above or below 15 percent). The last variable used for sorting within the implicit stratification was the estimated enrollment of 15-year-old students based on grade enrollments.

The U.S. sample included both public and private schools, randomly selected and weighted to be representative of the nation. The overall weighted school response rate was 69 percent before the use of substitute schools. Following PISA guidelines, substitute schools were identified by assigning as substitutes the two schools neighboring the sampled school in the sorted frame of schools. If an originally sampled school declined to participate, the first substitute was then contacted. If that school also refused to participate, the second school was then contacted.

Participating schools provided lists of 15-year-old students typically in August or September 2006, and a sample of 42 students was selected within each school in an equal probability sample. PISA 2006 assessments and questionnaires were administered between September and November 2006. In total, 166 schools and 5,611 students participated in PISA 2006 in the United States, with an average of about 35 students per school. The final weighted student response rate was 91 percent.

## **E.3 Weights**

**What are weights?**

Weights are variables that are put onto the file to compensate for unequal probabilities of selection and to adjust for the effects of nonresponse. Using weights allows a researcher to

make generalizations to the national population of 15-year-old students represented by PISA 2006. On the PISA 2006 student files, there are three weighting variables:

- W\_FSTUWT: Student final weight for all responding students.
- W\_FSCHWT: Final weight for all responding schools, projecting to the population of schools that 15-year-old students attended in 2006. Since PISA 2006 was designed to yield a representative sample of 15-year-olds enrolled in schools, and the school sample was designed to optimize the selection of these students, it is usually recommended that analyses involving school characteristics be treated as “student characteristics” and conducted at the student level (with W\_FSTUWT).
- W\_REPWT: Replicate weights (80 in total) that enable correct computation of standard errors.

### **Why do we need to use weights with the PISA 2006 data?**

If we do not use weights, the estimates (e.g., counts, proportions, means) that we produce will not be representative of the population about which we are attempting to generalize, that is, U.S. 15-year-old students in 2006.

### **Why would unweighted estimates not be representative?**

The PISA 2006 survey included 5,611 sampled students from across the nation. These participating students represent the 15-year-old students who attended school in the United States in 2006. Because small schools were sampled at a slightly lower rate than large schools, it is necessary to use a weight variable to correct for this. Nonresponse adjustment (correcting for those students who were selected but did not participate in the survey) must also be taken into consideration, because the weights of nonrespondents are distributed among the respondents with similar characteristics. Thus, weights reflect both unequal probabilities of sampling and nonresponse adjustments. It is incumbent upon the researcher to use appropriate weighting variables.

**WARNING:** Researchers should avoid breaking down the sample into such small categories that the analysis is questionable. For example, if a crosstabulation table has a single cell with fewer than 30 cases when the data are not weighted, NCES recommends that the results not be displayed or that the cell be combined with another cell (if appropriate).

### **Why do I need to use replicate weights when I compute means, frequencies, etc.?**

Because the PISA 2006 sample design involved stratification of schools and clustering of students within schools, the resulting statistics are more variable than they would have been had they been based on data collected from a simple random sample (SRS) of the same size. Without accounting for the design, a researcher will underestimate variance estimates, most notably the standard errors (estimated percentages or counts are not affected by the design effects, only by weights), and actually produce estimates that are

often much lower than an SRS design. Underestimating the standard errors will lead to inflated  $t$  values in hypothesis testing. Inflated  $t$  values, in turn, will increase the chance of a type I error, rejecting the null hypotheses. A type I error occurs when a researcher states that two groups are different when in fact they are not. Using replicate weights with the available macro program code (for SAS or SPSS) will produce correct standard errors and enable correct inferences to be drawn about likely true scores of and differences among student populations and subgroups.

## **E.4 Electronic Codebooks**

### **When I receive my PISA 2006 CD, what are some of the steps that I should follow to check out my CD?**

The following steps may help you get a better understanding of the PISA 2006 data.

Step 1—Make sure that you have the right file.

A general rule that should be followed by all researchers when they receive data from the government or any other source is to check the file for accuracy. Does this file include what you think it does? The following questions should be answered for the PISA 2006 CD.

1. Does the PISA 2006 CD contain the files listed in the documentation? Check the directory and subdirectories.
2. Do the weights (final weights) contained on the data files allow you to replicate weighted frequencies found in user manuals or published analyses? You may want to run weighted frequencies on a single variable using each of the weights contained on the file.

Step 2—Understand the data.

Assuming that you performed the above steps and you are confident that the files appear to contain what you hoped they would, it is now time to start learning about the files that you will be working with. Start by asking the following questions:

1. What were the processes involved in getting data from students via questionnaires and cognitive assessments to the medium (CD) that you now possess? Even though you did not collect the data, you need to know the procedures that were involved in collecting and processing the data. You also need to understand the quality control checks that were performed by the contractors in processing the data.
2. How were assessments and questionnaires administered? It is important to realize that students responded to a subset of the assessment items; thus it is necessary to use plausible values for all analyses. In completing the questionnaires, some respondents skipped individual items. For example, (a) some respondents refused to complete selected items, (b) some did not reach the end of the questionnaire, (c) some made illegal skips in the questionnaire, and (d) some responded outside valid ranges. Thus it is important to pay attention to missing data.
3. What can I do to further my understanding of the cases and variables that I plan to use? You can perform your own quality control procedures by answering the

following questions: Are the cases that I selected representative of the population to which I wish to generalize? How do the various breakdowns of the data compare with known population numbers? Is my sample biased—do nonrespondents look different from respondents?

## **E.5 Privileged or Restricted-Use Data**

### **When might I need a restricted-use license, and how do I get one?**

You should consider using the PISA 2006 restricted files when you need data that allow analyses that cannot be accomplished with the public-use data. For example, if you need zip code data to link school information to Census files that contain more detailed information about the community, you should think about obtaining an NCES license agreement. The PISA 2006 restricted-use data file contains information that identifies each school, specifically the NCES identification number (the NCES Common Core of Data ID [for public schools] or the NCES Private School Survey ID) and school name, street address, city, state, and zip code. In addition, the PISA 2006 restricted-use data file contains a race/ethnicity variable with more detail than the public-use version. Contact Brett Wallace at (202) 502-7307 or e-mail him at [brett.wallace@ed.gov](mailto:brett.wallace@ed.gov) for information on how to obtain a license.