

Quick Guide to Using the PIRLS 2006 Electronic Codebook (ECB)

Quick Guide to Using the PIRLS 2006 Electronic Codebook (ECB)

March 2009

Ben Dalton
RTI International

Stephen Provasnik
Project Officer
National Center for Education Statistics

Contents

A. Introduction to PIRLS 2006	1
A.1 Overview.....	1
A.2 Major Features	2
B. Description of the PIRLS 2006 ECB	2
B.1 Description of PIRLS 2006 Files.....	2
B.2 CD-ROM.....	3
C. Getting Started	3
C.1 Installing the PIRLS 2006 ECB.....	3
C.2 Using the PIRLS 2006 ECB	4
C.3 ID, Weighting, and Complex Sample Adjustment Variables	6
C.4 Plausible Values.....	7
D. A Sample ECB Session	8
D.1 Selecting Variables for Analysis from the ECB	8
D.2 Creating a SAS Working Data File.....	9
D.3 Creating an SPSS Working Data File	11
D.4 Merging Student and School Data	13
D.4.1 To Merge Student and School Data in SAS.....	13
D.4.2 To Merge Student and School Data in SPSS.....	14
D.5 Merging Student and Assessment Data	14
D.5.1 To Merge Student and Assessment Data in SAS.....	15
D.5.2 To Merge Student and Assessment Data in SPSS	15
D.6 Merging Student and Teacher Data	16
D.6.1 To Merge Student and Teacher Data in SAS.....	17
D.6.2 To Merge Student and Teacher Data in SPSS	18
E. Using Output SPSS Datasets with IDB Analyzer	18
F. Frequently Asked Questions About PIRLS 2006 (FAQs)	19
F.1 General and Background Questions.....	20
F.2 Sampling	21
F.3 Weights	21
F.4 Electronic Codebooks	23
F.5 Privileged or Restricted-Use Data	24

Quick Guide to Using the PIRLS 2006 Data

The purpose of the “Quick Guide” is to orient users of the Progress in International Reading Literacy Study (PIRLS 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 PIRLS 2006 data. The first section of this guide provides a brief overview of the PIRLS 2006 survey and available data files. The next three sections describe the PIRLS 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 is based on the public-use version of the PIRLS 2006 ECB, unless otherwise noted.

This Quick Guide is not designed to replace detailed knowledge of PIRLS 2006 data collection, variable construction, or analysis procedures. For further information on PIRLS 2006, see the *PIRLS 2006 User Guide for the International Database* available at http://timssandpirls.bc.edu/pirls2006/user_guide.html. In addition, users may wish to use a free software program, IDB Analyzer, provided by the PIRLS main study sponsor, the International Association for the Evaluation of Educational Achievement (IEA). IDB Analyzer is briefly described in section E and is available for download at: http://www.iea.nl/iea_studies_datasets.html.

Also note that stand-alone SAS and SPSS program code is available on the NCES website (<http://nces.ed.gov/surveys/pirls/>). 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 PIRLS 2006

A.1 Overview

PIRLS is a complex international effort to assess the reading comprehension of fourth-grade students. PIRLS was first conducted in 2001 and then again in 2006. Forty-five countries or jurisdictions participated in PIRLS 2006, including the United States. At the heart of the PIRLS study is an assessment of student performance on a combined reading literacy scale, a literary subscale (reading for literary experience), and an informational subscale (reading for obtaining and using information). In addition to the student assessments, PIRLS 2006 gathered student information via a questionnaire and information about schools and teachers from questionnaires given to administrators and responding students’ teachers, respectively.

PIRLS in the United States was sponsored by the National Center for Education Statistics (NCES) of the Institute of Education Sciences, U.S. Department of Education. Development and administration of PIRLS at the international level occurs under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). In the United States, a total of 183 schools participated in the 2006 study, with 5,190 students in those schools completing the PIRLS assessment. For the assessment, each student was asked to read two passages and answer

questions about them within 80 minutes. Considerable effort was made to ensure that assessment items (as well as questionnaire items) were comparable before and after translation across participating countries and jurisdictions. In the United States, students were in the fourth grade of schooling; in other countries, students were in the grade most comparable to fourth grade in the United States.

A.2 Major Features

PIRLS 2006 features internationally comparable samples of fourth-grade students; an internationally standardized set of assessment items in reading literacy; internationally standardized student, teacher, and school administrator questionnaires; allowances in the questionnaires for jurisdiction-specific items or responses; the inclusion of multiple imputed estimates of individual student performance (known as plausible values); and individual assessment items' performance data reported along with overall performance plausible values.

B. Description of the PIRLS 2006 ECB

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

- search the names and labels of variables in the PIRLS 2006 data;
- examine the unweighted response categories, frequencies, and percentages of responses for all PIRLS 2006 variables;
- create a list of “tagged” variables (called a tag list or tag file) to be extracted from the PIRLS 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 PIRLS 2006 Files

The PIRLS 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 titles identify the content of files you may see:

ASAUSAR2	Item-level student assessment responses
ASGUSAR2	Student questionnaire responses and assessment results
ASTUSAR2	Student-teacher linkage file (this file is not typically used in analyses of stand-alone U.S. data)
ATGUSAR2	Teacher information and questionnaire responses
ACGUSAR2	School information and school administrator questionnaire 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
.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 (i.e., the assessment data on the restricted-use ECB is the same as on the public-use ECB).

B.2 CD-ROM

The public-use datasets, ECB, and supporting documentation for PIRLS 2006 are located on one CD-ROM, NCES 2009-061. This data product contains the following files:

- PIRLS 2006 student, school, and assessment data;
- ECB software (discussed above);
- this Quick Guide to Using PIRLS 2006 Data;
- an ECB help file (HELP.PDF); and
- an electronic copy of the PIRLS 2006 Data Analysis User's Guide.

The restricted-use CD-ROM (NCES 2009-051) contains the ECB software, the ECB help file, and a restricted-use supplement to the PIRLS 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 PIRLS 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 PIRLS 2006 ECB

Files on PIRLS 2006 CD-ROM

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

- The root folder includes this document (QuickGuide.PDF); HELP.PDF; and folders ECBW, Report, and Tag.

- The “ECBW” folder includes five data files (student general, student item-level assessment, student-teacher linkage, teacher, and school), documentation for data files, and an installation program (Setup.exe) for the PIRLS 2006 ECB.
- The “Report” folder includes an electronic copy of PIRLS 2006 documentation.
- The “Tag” folder includes five tag files (one for each file type listed in B.1) 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 five tag files with preselected variables allow users to easily select critical identification and sampling variables before selecting other variables of interest.

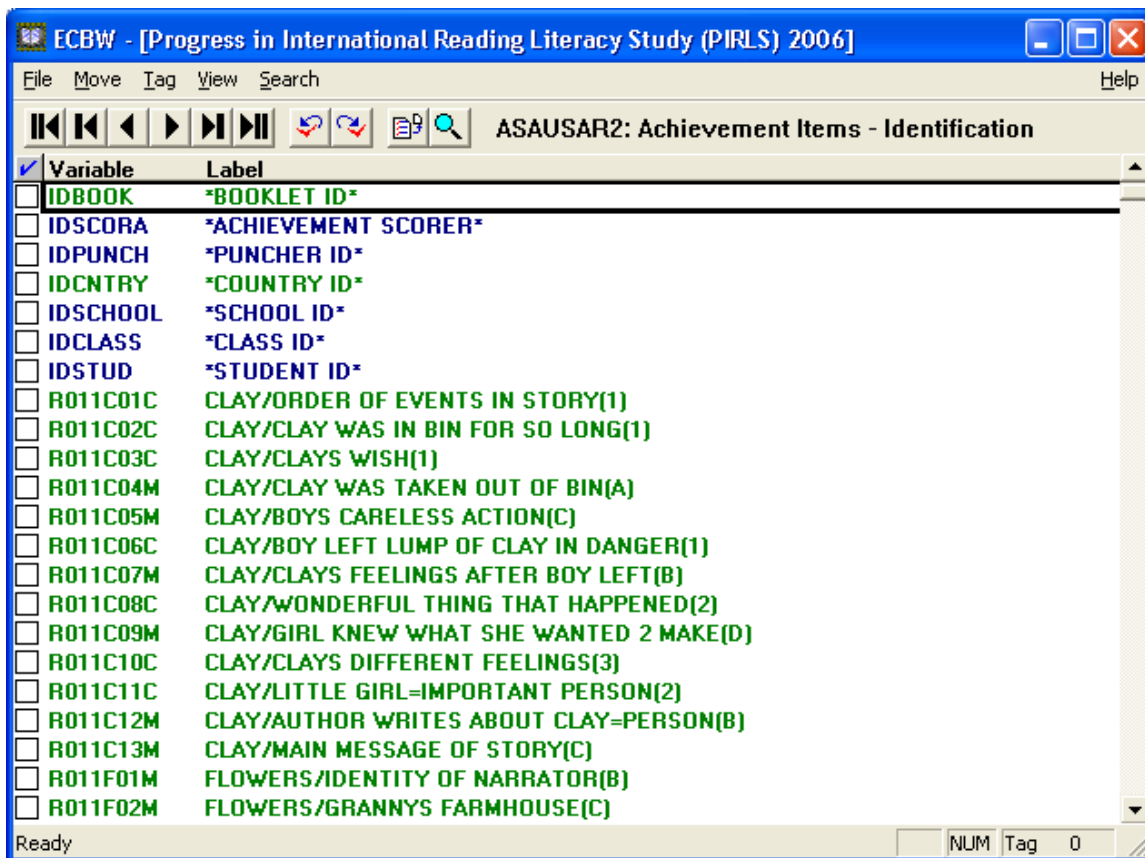
Installing and Loading the ECB

The PIRLS 2006 ECB will be automatically installed and launched on your computer when you run the ECB’s file “Setup.exe.” The following steps describe how to do this.

1. Insert the PIRLS 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 | PIRLS 2006 (named “PR7” for the public-use version, “P7R” for the restricted-use version) program, or double-click on the PIRLS 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 PIRLS 2006 ECB

When the PIRLS 2006 is correctly installed and operating, you should see an ECB window that looks like the screenshot below, with a list of variable names and labels in 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., ASGUSAR2).
- 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 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 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

For any analysis of PIRLS 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 PIRLS 2006 ECB, selection of variables from different files will produce SAS or SPSS program code that creates separate files. These files must be manually merged if the user wants different-file data in the same working dataset. Use of appropriate ID variables is critical in that case. See the *PIRLS 2006 User Guide for the International Database* for further information about ID variables.

PIRLS 2006 includes two main types of weights: total weights and replicate weights. Total weights are designed to produce estimates that pertain to the full student population of each country. Failing to use these total weights will render results that reflect only the specific study

sample and not the fourth-grade population that the sample in PIRLS 2006 was designed to represent. The main total weights include the following:

- Total student weight (TOTWGT) is necessary to produce correct population estimates of student-level variables.
- Teacher weight (TCHWGT) is designed for student-level analyses involving teacher information.
- School weight (SCHWGT) is designed for school-level analyses of school information.

In addition, the data files contain a senate weight (SENWGT) designed for cross-national analyses that involve more than one country. The senate weight adjusts sample sizes to be the same (500) in each country. Also included is a house weight (HOUWGT) which adjusts sample sizes so that the weighted sample is the same as the sample size in each country, which may be useful if analysis software uses the larger TOTWGT-weighted sample size for statistical tests.

Replicate weights are designed to be used with the jackknife replication procedure 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. Replicate weights are stored in variables with names that begin with the letters “JK.”

Other weights are available as well for cross-national analyses. The *PIRLS 2006 User Guide for the International Database* contains more information on these additional weights and information about the weighting procedures used for PIRLS.

SAS and SPSS software, as well as other statistical packages, have options for using total weights and replicate weights. In addition, SAS and SPSS program code files specially written for PIRLS are available at the IEA website (http://timss.bc.edu/pirls2006/user_guide.html). Further, IEA’s free analysis software package, the IDB Analyzer, is specifically designed to analyze international datasets like PIRLS 2006, and has routines for easily handling the weighting variables. See section E below for more information.

Preselected tag lists of ID and weighting variables for the five individual files, as well as for all files combined, are available in the “Tag” folder of the CD-ROM.

C.4 Plausible Values

PIRLS 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 PIRLS sample members did not take the full battery of assessment items (each student was given a subset of items). In the case of PIRLS 2006, five plausible values were computed for each student respondent, indicating possible “true” values of the student’s score on the underlying conceptual dimension (science, mathematics, or reading literacy, 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 IEA website (http://timss.bc.edu/pirls2006/user_guide.html). Also see the *PIRLS 2006 User Guide for the International Database* for more information.

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 PIRLS 2006 ECB software. In this example, a user wants to create a file of variables associated with students' reading achievement and time spent doing reading homework. The actions the user would perform are described step by step.

Open the PIRLS 2006 ECB by double-clicking on the PIRLS 2006 ECB icon on the computer desktop or by clicking on the START button, going to Programs, selecting "Electronic Codebook for Windows," and clicking on "PR7." Read and respond to the warning; if you "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 six variables will appear, identifying variables that record student responses (student variable names start with the letters "AS") and teacher answers to questions about homework (teacher variable names start with the letters "AT"). Click OK to close this box. Then click the "Next Tag" button on the toolbar (third from right). This brings you to the first set of tagged variables. The third variable here, ASBGRHT, identifies time spent on reading homework. Since this is the variable of interest, we want to select this and not the other. To remove all tagged variables (including the ones not immediately adjacent to ASBGRHT), click "Tag | Clear All Tag" on the menu bar. Then tag ASBGRHT again.

We know that the student file also contains the plausible values variables for the overall reading literacy 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," then "Close." The main window will have highlighted the first of the five plausible values in overall reading literacy. Tag all five by selecting the check-box for each (ASRREA01 to ASRREA05), or hold down the Shift key, select ASRREA01, then tag ASRREA05, 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 the user will create from the raw PIRLS 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 READ_HWK. To do so, click "File | Output | Tag File" on the menu bar, and choose the location to which to save. Type READ_HWK 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 lists" from the menu bar (a warning will come up about clearing existing tags; since READ_HWK has already been saved, you do not need to worry about this). Browse for or type in the directory

path for READ_HWK in the File #1 box, then browse for or type in the directory path for the ASGUSAR2.tag file (student background 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 READ_HWK.tag to the same location as previous, 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) PIRLS 2006 data file. We want to save these SAS program statements in a file named READ_HWK.sas. To do so, click on “File | Output | SAS-PC” on the menu bar. Choose the directory location and type READ_HWK to save. In the next dialogue box, a user will be asked to label reserve codes and should typically choose “Categorical” or “None”. 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.

READ_HWK.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 READ_HWK.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 PR7 'C:\ECBW\PR7';
```

The libname is assigned automatically by SAS with the directory path in which the program is saved. The libname PR7 may be changed to a different name if desired.

```
DATA X2;
```

The DATA statement indicates that a temporary SAS working file (called X2) 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 “PR7.” (with the period) prior to X2. 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 also wish to change the automatically assigned X2 name to something more memorable and because X2 is the name always assigned by the ECB to files created from the student questionnaire and assessment results file, a fact which may cause complications later with other extract files.

```
INFILE 'D:\ECBW\PR7\DATA\ASGUSAR2.DAT' LRECL=1024 PAD;
```

The INFILE statement indicates the location of the PIRLS 2006 raw data file from which the chosen variables will be taken. ECB-generated code will automatically chose 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 (e.g., student general, student item-level assessment, teacher), the ECB code will repeat these steps for each raw data file. Separate data sets will be generated for each file in which a variable has been tagged.

INPUT

```
IDCNTRY 4-8 IDSCHOOL 9-12 IDSTUD 20-27 ILTEACH1 46-51
ASBGRHT 106-106 ASRREA01 175-187 .6 ASRREA02 188-200 .6
ASRREA03 201-213 .6 ASRREA04 214-226 .6 ASRREA05 227-239 .6
HOUWGT 505-517 .6 TOTWGT 518-530 .6 SENWGT 531-543 .6
JKZONE 622-623 JKREP 624-624;
```

LABEL

```
IDCNTRY='*COUNTRY ID*'
IDSCHOOL='*SCHOOL ID*'
IDSTUD='*STUDENT ID*'
ILTEACH1='*FIRST TEACHER ID*'
ASBGRHT='GEN/READING/TIME SPENT ON READING'
ASRREA01='PLAUSIBLE VALUE: OVERALL READING PV1'
ASRREA02='PLAUSIBLE VALUE: OVERALL READING PV2'
ASRREA03='PLAUSIBLE VALUE: OVERALL READING PV3'
ASRREA04='PLAUSIBLE VALUE: OVERALL READING PV4'
ASRREA05='PLAUSIBLE VALUE: OVERALL READING PV5'
HOUWGT='HOUSE WEIGHT'
TOTWGT='TOTAL STUDENT WEIGHT'
SENWGT='SENATE WEIGHT'
JKZONE='JACKKNIFE ZONE'
JKREP='JACKKNIFE REPLICATE CODE';
```

PROC FORMAT;

```
VALUE IDCNTRYF      840='United States';
VALUE ASBGRHTF     1='I NEVER HAVE READING TO DO FOR HOMEWORK'
                   2='HALF HOUR OR LESS'
                   3='MORE THAN A HALF HOUR TO 1 HOUR'
                   4='MORE THAN 1 HOUR'
                   8='not administered'
                   9='omitted';
VALUE JKREPF       0='0'
                   1='1';
```

PROC FREQ;

FORMAT

```
IDCNTRY IDCNTRYF. ASBGRHT ASBGRHTF. JKREP JKREPF. ;
```


TABLES

```
IDCNTRY ASBGRHT JKREP ;
```

```
PROC UNIVARIATE;VAR
```

```
IDSCHOOL IDSTUD ILTEACH1 ASRREA01 ASRREA02 ASRREA03
```

```
ASRREA04 ASRREA05 HOUWGT TOTWGT SENWGT JKZONE ;
```

```
RUN;
```

This is the end of the file. Other than the changes to the file locations and output file name, no additional 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) PIRLS 2006 data file. We want to save these SPSS program statements in a file named READ_HWK.sps. To do so, click on “File | Output | SPSS” on the menu bar. Choose the directory location and type READ_HWK 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.

READ_HWK.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 READ_HWK.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\PR7\DATA\ASGUSAR2.DAT' FIXED RECORDS=1 /
  IDCNTRY 4-8 IDSCHOOL 9-12 IDSTUD 20-27 ILTEACH1 46-51
  ASBGRHT 106-106 ASRREA01 175-187 (6) ASRREA02 188-200 (6)
  ASRREA03 201-213 (6) ASRREA04 214-226 (6) ASRREA05 227-239 (6)
  HOUWGT 505-517 (6) TOTWGT 518-530 (6) SENWGT 531-543 (6)
  JKZONE 622-623 JKREP 624-624
```

The DATA LIST statement indicates the location of the PIRLS 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 (e.g., student general, student item-level assessment, teacher), the ECB code will repeat programming steps for each raw data file. Separate data sets will be generated for each file in which a variable has been tagged.

VARIABLE LABELS

```
IDCNTRY '*COUNTRY ID*'
IDSCHOOL '*SCHOOL ID*'
IDSTUD '*STUDENT ID*'
ILTEACH1 '*FIRST TEACHER ID*'
ASBGRHT 'GEN/READING/TIME SPENT ON READING'
ASRREA01 'PLAUSIBLE VALUE: OVERALL READING PV1'
ASRREA02 'PLAUSIBLE VALUE: OVERALL READING PV2'
ASRREA03 'PLAUSIBLE VALUE: OVERALL READING PV3'
ASRREA04 'PLAUSIBLE VALUE: OVERALL READING PV4'
ASRREA05 'PLAUSIBLE VALUE: OVERALL READING PV5'
HOUWGT 'HOUSE WEIGHT'
TOTWGT 'TOTAL STUDENT WEIGHT'
SENWGT 'SENATE WEIGHT'
JKZONE 'JACKKNIFE ZONE'
JKREP 'JACKKNIFE REPLICATE CODE'
```

VALUE LABELS

```
/ IDCNTRY      840 'United States'
/ ASBGRHT      1  'I NEVER HAVE READING TO DO FOR HOMEWORK'
                2  'HALF HOUR OR LESS'
                3  'MORE THAN A HALF HOUR TO 1 HOUR'
                4  'MORE THAN 1 HOUR'
                8  'not administered'
                9  'omitted'
/ JKREP        0  '0'
                1  '1'
```

```
SAVE OUTFILE='C:\ECBW\PR7\X2' /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, X2). If you want to save to a different directory, change the directory information here. You may also wish to change the automatically assigned X2 name to something more memorable, and because X2 is the name always assigned by the ECB to files created from the student questionnaire and assessment results file, a fact which may cause complications later with other extract files.

```
FREQUENCIES VARIABLES=
```

```
  IDCNTRY ASBGRHT JKREP .
```

```
DESCRIPTIVES VARIABLES=
```

```
  IDSCHOOL IDSTUD ILTEACH1 ASRREA01 ASRREA02 ASRREA03
  ASRREA04 ASRREA05 HOUWGT TOTWGT SENWGT JKZONE
  /STATS=DEF.
```

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

D.4 Merging Student and School Data

If you tag variables from the student background file (ASGUSAR2) and the school file (ACGUSAR2), 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 ASGUSAR2.TAG and ACGUSAR2.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.

D.4.1 To Merge Student and School Data in SAS

The automatically generated ECB code will have the following structure; after the end of the first “RUN;” line, add the indicated code:

```
LIBNAME PR7 'C:\ECBW\PR7';

DATA X2;INFILE 'D:\ECBW\PR7\DATA\ASGUSAR2.DAT' LRECL=1024 PAD;INPUT
  IDCNTRY 4-8 IDSCHOOL 9-12 IDSTUD 20-27 ILTEACH1 46-51
  HOUWGT 505-517 .6 TOTWGT 518-530 .6 SENWGT 531-543 .6
  JKZONE 622-623 JKREP 624-624;
```

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

```
DATA X3;INFILE 'D:\ECBW\PR7\DATA\ACGUSAR2.DAT' LRECL=1024 PAD;INPUT
  IDSCHOOL 9-12;
```

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 merge the school data to the student file. This creates a permanent dataset named “EXTRACT” in the folder referenced by PR7 at the beginning of the automatically generated code. You may change the “EXTRACT” name to something more specific.

```
DATA PR7.EXTRACT; MERGE X2 X3;
  BY IDSCHOOL;

RUN;
```

D.4.2 To Merge Student and School Data in SPSS

The automatically generated ECB code will have the following structure; after the end of the first “SAVE OUTFILE...” line, add the indicated code:

```
DATA LIST FILE='D:\ECBW\PR7\DATA\ASGUSAR2.DAT' FIXED RECORDS=1 /
  IDCNTRY 4-8 IDSCHOOL 9-12 IDSTUD 20-27 ILTEACH1 46-51
  HOUWGT 505-517 (6) TOTWGT 518-530 (6) SENWGT 531-543 (6)
  JKZONE 622-623 JKREP 624-624
```

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

```
SAVE OUTFILE='C:\ECBW\PR7\X2' /Compressed.
```

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

```
DATA LIST FILE='D:\ECBW\PR7\DATA\ACGUSAR2.DAT' FIXED RECORDS=1 /
  IDSCHOOL 9-12
```

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

```
SAVE OUTFILE='C:\ECBW\PR7\X3' /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 merge the school data to the student file. The command with “/TABLE” should reference the school data (here, “X3”). 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\PR7\X2'
  /TABLE='C:\ECBW\PR7\X3'
  /BY IDSCHOOL.
EXECUTE.

SAVE OUTFILE='C:\ECBW\PR7\XTRACT' /Compressed.
```

D.5 Merging Student and Assessment Data

If you tag variables from the student background file (ASGUSAR2) and the item-level assessment file (ASAUSAR2), 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 ASGUSAR2.TAG and ASAUSAR2.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.

D.5.1 To Merge Student and Assessment Data in SAS

The automatically generated ECB code will have the following structure; after the end of the first “RUN;” line, add the indicated code (note that only the merging variable appears in this example):

```
LIBNAME PR7 'C:\ECBW\PR7';  
  
DATA X1; INFILE 'D:\ECBW\PR7\DATA\AS AUSAR2.DAT' LRECL=1024 PAD; INPUT  
    IDSTUD 24-31;
```

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

```
DATA X2; INFILE 'D:\ECBW\PR7\DATA\AS GUSAR2.DAT' LRECL=1024 PAD; INPUT  
    IDCNTRY 4-8 IDSCHOOL 9-12 IDSTUD 20-27 ILTEACH1 46-51  
    HOUWGT 505-517 .6 TOTWGT 518-530 .6 SENWGT 531-543 .6  
    JKZONE 622-623 JKREP 624-624;
```

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 merge the assessment results to the student file. This creates a permanent dataset named “EXTRACT” in the folder referenced by PR7 at the beginning of the automatically generated code. You may change the “EXTRACT” name to something more specific.

```
PROC SORT DATA=X2;  
    BY IDSTUD;  
RUN;  
  
DATA PR7.EXTRACT; MERGE X1 X2;  
    BY IDSTUD;  
  
RUN;
```

D.5.2 To Merge Student and Assessment Data in SPSS

The automatically generated ECB code will have the following structure; after the end of the first “SAVE OUTFILE...” line, add the indicated code (note that only the merging variable appears in this example):

```
DATA LIST FILE='D:\ECBW\PR7\DATA\AS AUSAR2.DAT' FIXED RECORDS=1 /
```

IDSTUD 24-31

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

```
SAVE OUTFILE='C:\ECBW\PR7\X1' /Compressed.
```

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

```
DATA LIST FILE='D:\ECBW\PR7\DATA\ASGUSAR2.DAT' FIXED RECORDS=1 /  
  IDCNTRY 4-8 IDSCHOOL 9-12 IDSTUD 20-27 ILTEACH1 46-51  
  HOUWGT 505-517 (6) TOTWGT 518-530 (6) SENWGT 531-543 (6)  
  JKZONE 622-623 JKREP 624-624
```

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

```
SAVE OUTFILE='C:\ECBW\PR7\X2' /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 merge the assessment results to the student file. The command “/TABLE” should refer to the student file (here, “X2”)—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\X2'.  
SORT CASES BY IDSTUD.  
SAVE OUTFILE='C:\ECBW\PS6\X2'.  
  
MATCH FILES  
  /FILE='C:\ECBW\PR7\X1'  
  /TABLE='C:\ECBW\PR7\X2'  
  /BY IDSTUD.  
EXECUTE.  
  
SAVE OUTFILE='C:\ECBW\PR7\XTRACT' /Compressed.
```

D.6 Merging Student and Teacher Data

If you tag variables from the student background file (ASGUSAR2) and the teacher file (ATGUSAR2), the SAS and SPSS program code that creates the working data files will create separate student and teacher 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 ASGUSAR2.TAG and ATGUSAR2.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.

The code below is based on the fact that ILTEACH1 (in the student file) is identical to IDTEACH (in the teacher file). However, this is only true if data from one country such as the United States are being used. If you plan to use U.S. data in conjunction with data from other countries, see the PIRLS User Guide for the International Database for information on combining student and teacher data.

D.6.1 To Merge Student and Teacher Data in SAS

The automatically generated ECB code will have the following structure; after the end of the first “RUN;” line, add the indicated code (note that only the merging variable appears in the teacher file in this example):

```
LIBNAME PR7 'C:\ECBW\PR7';

DATA X2; INFILE 'D:\ECBW\PR7\DATA\ASGUSAR2.DAT' LRECL=1024 PAD; INPUT
  IDCNTRY 4-8 IDSCHOOL 9-12 IDSTUD 20-27 IDTEACH 46-51
  HOUWGT 505-517 .6 TOTWGT 518-530 .6 SENWGT 531-543 .6
  JKZONE 622-623 JKREP 624-624;
```

NOTE: In this example, ILTEACH1 normally appears before its designated location “46-51”, but has been renamed IDTEACH to permit it to merge with teacher data. This step MUST be accomplished for the merge to work properly.

Next will be other student file statements (formats, frequencies). Any references to ILTEACH1 should be changed to IDTEACH in these statements, otherwise an error message may occur. Following these statements will be:

```
DATA X4; INFILE 'D:\ECBW\PR7\DATA\ATGUSAR2.DAT' LRECL=1024 PAD; INPUT
  IDTEACH 13-18;
```

Next will be other teacher file statements (formats, frequencies). Following these will be

```
RUN;
```

This marks the end of the automatically generated ECB code.

Add the following code to merge the teacher data to the student file. This creates a permanent dataset named “EXTRACT” in the folder referenced by PR7 at the beginning of the automatically generated code. You may change the “EXTRACT” name to something more specific.*/

```
DATA PR7.EXTRACT; MERGE X2 X4;
  BY IDTEACH;

RUN;
```

D.6.2 To Merge Student and Teacher Data in SPSS

The automatically generated ECB code will have the following structure; after the end of the first “SAVE OUTFILE...” line, add the indicated code (note that only the merging variable appears in the teacher file in this example):

```
DATA LIST FILE='D:\ECBW\PR7\DATA\ASGUSAR2.DAT' FIXED RECORDS=1 /
  IDCNTRY 4-8 IDSCHOOL 9-12 IDSTUD 20-27 IDTEACH 46-51
  HOUWGT 505-517 (6) TOTWGT 518-530 (6) SENWGT 531-543 (6)
  JKZONE 622-623 JKREP 624-624
```

NOTE: In this example, ILTEACH1 normally appears before its designated location “46-51”, but has been renamed IDTEACH to permit it to merge with teacher data. This step **MUST** be accomplished for the merge to work properly.

Variable and Value label statements for the student data follow. NOTE that any references to ILTEACH1 in these statements should be changed to IDTEACH, otherwise an error message may occur. Following these statements will be:

```
SAVE OUTFILE='C:\ECBW\PR7\X2' /Compressed.
```

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

```
DATA LIST FILE='D:\ECBW\PR7\DATA\ATGUSAR2.DAT' FIXED RECORDS=1 /
  IDTEACH 13-18
```

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

```
SAVE OUTFILE='C:\ECBW\PR7\X4' /Compressed.
```

Next will be frequencies statements for the teacher data. These statements mark the end of the automatically generated ECB code.

Add the following code to merge the teacher data to the student file. The command “/TABLE” should refer to the teacher file (here, “X4”). Again, you may change the outfile location. You may also change the “XTRACT” name to something more specific.

```
MATCH FILES
  /FILE='C:\ECBW\PR7\X2'
  /TABLE='C:\ECBW\PR7\X4'
  /BY IDTEACH.
EXECUTE.

SAVE OUTFILE='C:\ECBW\PR7\XTRACT' /Compressed.
```

E. Using Output SPSS Datasets with IDB Analyzer

SPSS datasets, once created, can be imported into IDB Analyzer, a free software program available from the IEA that enables analysis of PIRLS 2006 data. The IDB Analyzer program,

along with instructions on its use, can be found on the IEA's website at http://www.iea.nl/iea_studies_datasets.html. IDB Analyzer is specifically designed to analyze international datasets such as PIRLS 2006, and has explicit procedures for handling plausible values. Note that IDB Analyzer requires SPSS datasets that contain weighting (e.g., TOTWGT for student-level analysis) and complex survey adjustment variables (those beginning "JK"—these variables should be tagged in the ECB (by the user or by importing preexisting tag files from the "Tag" folder of the CD) before outputting SPSS program code to create an SPSS dataset.

F. Frequently Asked Questions About PIRLS 2006 (FAQs)

NCES staff members have received many questions regarding "proper techniques" for working with PIRLS 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 PIRLS 2006 documentation. To help the data user identify specific topics of interest, questions and responses have been grouped into the following categories:

F.1 General and Background Questions

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

F.2 Sampling

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

F.3 Weights

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

F.4 Electronic Codebooks

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

F.5 Privileged or Restricted-Use Data

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

F.1 General and Background Questions

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

For additional information and questions about PIRLS 2006 and NCES' international activities program, please link to the PIRLS 2006 web page:

<http://nces.ed.gov/surveys/pirls>.

Or contact:

Dr. Stephen Provasnik

(202) 502-7480

stephen.provasnik@ed.gov

What are the interrelationships among the different PIRLS 2006 files?

PIRLS 2006 surveyed students, teachers, and school administrators. All PIRLS 2006 files come from these sources. Students provided background information as well as participated in the reading literacy assessment. Overall reading assessment results as well as student background questionnaire responses are provided in the student background file (ASGUSAR2). Item-level responses to the student assessment are provided in a separate file (ASAUSAR2).

The teachers and school administrators of participating students were also surveyed. School information and school administrator questionnaire responses are found in the school file (ACGUSAR2) and teacher responses are found in the teacher file (ATGUSAR2). A fifth file is the student-teacher linkage file (ASTUSAR2). The student-teacher linkage file is for cases in which multiple teachers answered questionnaires for the same student. In the case of the United States, only one teacher per student answered a questionnaire, so this file is not typically used in analyses of the stand-alone U.S. data.

How is PIRLS 2006 related to other international studies?

The Progress in International Reading Literacy Study (PIRLS) is an international comparative study of the reading literacy of young students. 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. PIRLS is grade-based and focuses on reading skills. PIRLS was first administered in 2001 and included 35 countries, and was administered again in 2006 to students in 45 education systems.

The United States participates in two other major international assessments: The Program for International Student Assessment (PISA) and The Trends in International Mathematics and Science Study (TIMSS).

PISA is a system of international assessments that focus on 15-year-olds' 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 three years.

TIMSS provides reliable and timely 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. TIMSS has typically surveyed fourth- and eighth-grade students. In contrast to PISA, TIMSS focuses more broadly on math and science abilities and less on practical skills. TIMSS data has been collected in 1995, 1999, 2003, and 2007.

F.2 Sampling

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

The U.S. PIRLS sample was designed to be representative of all fourth-grade students in the 50 states and Washington, D.C. In addition to the base sample (designed to yield 150 participating schools), the United States sampled additional private schools and high-poverty schools, defined as those schools in which 50 percent or more of students were eligible to receive free or reduced-price lunch, in order to increase the precision of the estimates for these subgroups. Thus the U.S. sample was designed to yield 180 participating schools altogether.

The PIRLS school sample was based on a sampling frame from NCES' Common Core of Data (CCD) and Private School Universe Survey (PSS). To be consistent with the sampling design for PIRLS 2001, the frame was divided into one 10-largest-cities group (each called a metropolitan statistical area or MSA) and 451 other units (primary sampling units or PSU) outside of the 10 largest cities. Schools were sorted by state, percentage of racial/ethnic minority students, control of school (public/private), percentage of students eligible for free or reduced-priced lunch, and locale before the selection process. Within each selected PSU or MSA stratum, schools were selected on the basis of the number of fourth-grade students in the school so that larger schools had a higher selection probability than smaller schools.

Data collection occurred between January and June 2006. A total of 183 schools participated (including 63 schools selected as replacements for schools from the original sample that declined to participate), for a final weighted school response rate of 86 percent. One or two classrooms from each school were selected, depending on school size, and 5,190 students completed the PIRLS assessment for a weighted student response rate of 95 percent.

F.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 fourth-grade students represented by PIRLS 2006. On the PIRLS 2006 student files, there are three main weight variables:

- **TOTWGT:** Total student weight for analysis of individual student characteristics or performance.

- TCHWGT: Teacher weight designed for student-level analyses involving teacher information (note this is not designed for analyzing teachers as the unit of analysis; PIRLS 2006 was not designed to support such analyses).
- SCHWGT: School weights designed for school-level analysis.

In addition, the data files contain a senate weight (SENWGT) designed for cross-national analyses that involve more than one country. The senate weight adjusts sample sizes to be the same (500) in each country. Also included is a house weight (HOUWGT), which adjusts sample sizes so that the weighted sample is the same as the sample size in each country, which may be useful if analysis software uses the larger TOTWGT-weighted sample size for statistical tests.

Replicate weights are designed to be used with the jackknife replication procedure 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. Replicate weights are stored in variables with names that begin with the letters “JK.”

Other weights are available as well for cross-national analyses and which provide more information about the weighting procedures used for PIRLS as a whole; see the *PIRLS 2006 User Guide for the International Database* for more information on these additional weights.

Why do we need to use weights with the PIRLS 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, the population of U.S. fourth-grade students in 2006.

Why would unweighted estimates not be representative?

In PIRLS 2006, 5,190 sampled students participated in the survey from across the nation. These participating students represent the fourth-grade students who attended school in the United States in 2006. Because schools were sampled in proportion to the size of the fourth-grade enrollment, and in some schools two classes were sampled rather than one, it is necessary to use weights to accurately represent U.S. fourth-grade students. 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 PIRLS 2006 sample design involved the clustering of schools within PSUs and 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 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 a simple random sample design. Underestimating your standard errors will lead to inflated t values in hypothesis testing. This, in turn, increases the chance of a type I error, rejecting the null hypotheses. This is 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.

F.4 Electronic Codebooks

When I receive my PIRLS 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 PIRLS 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 PIRLS 2006 CD.

1. Does the PIRLS 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—Understanding 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 do 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?

F.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 PIRLS 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 PIRLS 2007 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. Contact Brett Wallace at (202) 502-7307 or e-mail brett.wallace@ed.gov for information on how to obtain a license.