INFORMATION CONTAINED IN THIS DOCUMENT:

- DESCRIPTION OF FILES
- HOW TO GET STARTED
- DESCRIPTION OF VARIANCE ESTIMATION VARIABLES
- DESCRIPTION OF ID VARIABLES FOR MERGING FILES

DESCRIPTION OF FILES

All of the files on the PISA NCES website contain data for the United States only. Files with data from all countries can be found at the OECD website for PISA: http://www.pisa.oecd.org.

The U.S. files available on this NCES website are:

- Student data

  - The data are contained in STUD06.DAT. This file contains questionnaire items and derived variables and index scores based on the student questionnaire; plausible values for overall science, science competency subscales, science knowledge subscales, science attitude subscales, and the mathematics scale from the assessment; and student sampling weights and replicate weights.

  - There are 5,611 cases in this file. Since the data are hierarchical (students are clustered with schools), each student record contains identification variables that enable the user to merge the school data with the student data, using the variable SCHOOLID.

  - an SPSS syntax file, STUD06.SPS

  - a SAS syntax file, STUD06.SAS

  - a codebook file (STUD06.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and frequencies.

- School data

  - The data are contained in SCHL06.DAT. This file contains items from the school questionnaire, derived variables and index scores based on the school questionnaire, and the school sampling weight. There are 166 cases in this file.

  - an SPSS syntax file, SCHL06.SPS

  - a SAS syntax file, SCHL06.SAS

  - a codebook file (SCHL06.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and frequencies.

- Cognitive (assessment) item data

- The data are contained in ASSESM06.DAT. This file contains student responses to each item in the assessment. Note that the majority of the items have not been released, so there is little descriptive information about them. There are 5,611 cases in this file.

- an SPSS syntax file, ASSESM06.SPS

- a SAS syntax file, ASSESM06.SAS

- a codebook file (ASSESM06.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and frequencies.

- U.S. Questionnaires

  - The U.S. version of the student questionnaire is in the file PISA_MS06_StudentQ_USA_Eng.PDF.

  - The U.S. version of the school questionnaire is in the file PISA_MS06_SchoolQ_USA_Eng.PDF

- Released items used in PISA 2006

  - Science items that were administered in PISA 2006 and subsequently released can be found in PISA_Sample_Items.PDF. Samples of mathematics and reading items are also included.

- PISA 2006 data user's guide

  - PISA_2006_Data_Analysis_Users_Guide.PDF, contains information on the conduct of PISA in the United States.

- PISA 2006 quick guide

  - PISA_QuickGuide.PDF, provides information about using the PISA 2006 ECB and merging data.

- Macros for use with SPSS to produce plausible values and design-corrected standard errors

  - This file, Setup_SPSSreplicates_V4_1.MSI, is designed for use with SPSS 15. It can be used to correct estimates of plausible values and design-corrected standard errors. A complete description of the macros can be found in *PISA 2003 Data Analysis Manual* (OECD 2005).

GETTING STARTED

The PISA Quick Guide found on the PISA 2006 Electronic Codebook (ECB) provides beginning instructions on how to install the ECB software, navigate its features, produce working datasets, merge information from different file sources (student, assessment, and school), and correctly analyze the PISA 2006 data. Users should consult the PISA Data User's Guide, also found on the ECB, for additional information about the PISA 2006 data. Complex analyses of PISA 2006 may require consulting both this PISA 2006 Data User's Guide as well as the PISA 2006 Technical Report (available at

, under "What's New" at the time of this document's writing). Technical notes/appendixes provided as part of PISA 2006 analysis publications from NCES or the OECD may also provide helpful information, as may PISA 2003 publications, particularly the PISA 2003 Data Analysis Manual, available at the OECD PISA website.

The key factor to note is that analysis of PISA 2006 data requires the use of specialized programs because of the use of plausible values and replicate weighting. PISA students do not have just one score for a given scale; rather each student has five "plausible values" for each scale. In addition, besides an overall student weight variable, replicate weights must be used to estimate correct standard errors (see next section). Special programs called macros can estimate scores and correct standard errors. Macros for use with SPSS 15 are available in a set-up file provided on the NCES website. SAS macros created for PISA 2003 can be found through the OECD PISA website.

DESCRIPTION OF WEIGHTS AND VARIANCE ESTIMATION VARIABLES
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. In addition, the macros referenced above under "Getting Started" are designed to apply the weights correctly.

Complete information on the weights and variance estimation variables is available in chapter 8 of the PISA 2006 Technical Report, which can be found on the OECD PISA website (at time of this document's writing, available under "What's New" at ).

DESCRIPTION OF ID VARIABLES USED TO MERGE FILES
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. More information on merging can be found in the PISA 2006 "Quick Guide" included on the PISA 2006 ECB.