

**Slide 1 of 24**

Title Slide: Getting Started with the PIAAC Data

**Slide 2 of 24**

This module provides users with an overview of the micro-level data files and documentation that are available for the Program for the International Assessment of Adult Competencies, or PIAAC. The module also describes which resources are available to learn more about the study, the data, and the data files. Information presented in this module is key to understanding some of the more detailed information presented in subsequent PIAAC modules.

**Slide 3 of 24**

PIAAC micro-level data files are available in both public-use format, or PUF (spelled P-U-F), and restricted-use format, or RUF (spelled R-U-F).

The international public-use files, which contain individual unit record data for all countries, are available for download from the Organization for Economic Cooperation and Development, or OECD. Both public-use and restricted-use files containing individual unit record data and country-specific data for the United States are available from NCES.

All data files contain responses to both the background questionnaire and the cognitive assessment, as well as plausible values and weighting variables for analysis.

The OECD or NCES public-use files and the NCES restricted-use files can be accessed by clicking on the corresponding underlined screen text.

**Slide 4 of 24**

The variables in the PIAAC data files are named using a standard set of conventions. Data users should find these naming conventions intuitive.

For variables associated with the background questionnaire, the first character of the variable name is the letter indicating the section it comes from. For example, the variable name for the first question in section B is “B\_Q01A”.

Background questionnaire items common across PIAAC and the International Adult Literacy Survey, or IALS (spelled I-A-L-S), and/or the Adult Literacy and Life Skills Survey, or ALL, have an additional “\_T” after the question number. This “T” indicates that items are trend items that were used in IALS and ALL.

For cognitive assessment items, the first character of the variable name depends on the item’s mode of delivery or other unique attribute. Computer-based items begin with “C”, “D”, or “E”. Paper-based items begin with “P”. Problem solving in technology-rich environment items begin with “U.”

## Getting Started with the PIAAC Data

Naming conventions for derived variables vary.

### Slide 5 of 24

For the purposes of maintaining confidentiality, certain variables in the OECD and NCES public-use data files were suppressed for U.S. respondents. These variables include:

- Detailed earnings data;
- Occupational data, including detailed occupation codes, job title, and work responsibilities; and
- Sampling selection variables.

### Slide 6 of 24

OECD public-use data files in SPSS and SAS formats are available for download for every country participating in PIAAC, except Australia and Cyprus.

PIAAC data are stored in separate files by country. The separate country files can be merged using the International Association for the Evaluation of Educational Achievement (or IEA) International Database Analyzer, or the IDB Analyzer, which will be discussed in more detail later in this module and in the module titled “Considerations for Analysis of PIAAC Data”.

The data files for each country are stored by the name “prgXXXp1” where XXX is each country’s three-letter country code. For example, the U.S. data file is named “prgUSAp1”.

The OECD data files can be accessed by clicking on the corresponding underlined screen text.

### Slide 7 of 24

NCES public- and restricted-use files containing individual unit record data for U.S. participants are available for download in ASCII, SPSS, and SAS formats. These files include variables for items that were asked of only U.S. respondents, in addition to all of the variables contained in the international public-use file.

### Slide 8 of 24

The U.S.-only variables focus on education, country of origin, language, race/ethnicity, training courses, occupation, health, work, and earnings. They include data for the U.S. routing of questions for those variables that differ from the international routing.

International variables from the background questionnaire that were adapted for the U.S. context end with the letters “US.” U.S.-only background questionnaire variables end with the characters “USX.”

**Slide 9 of 24**

In addition to the variables in the U.S. public-use file, the U.S. restricted-use file also contains variables suppressed in the public-use files. The restricted-use file contains responses for variables at the same level of detail as they were asked in the survey.

The restricted-use file can be accessed through a restricted-use license agreement with NCES. More information about restricted-use license agreements can be obtained by clicking on the corresponding underlined screen text.

**Slide 10 of 24**

Documentation for the OECD micro-data files includes the international technical report, the international data codebook, SAS and Stata data analysis manuals, SAS and Stata macros, and the data compendia for background and cognitive variables.

The OECD data documentation can be accessed by clicking on the corresponding underlined screen text here, or on the slides that follow describing each resource in more detail.

**Slide 11 of 24**

The OECD Technical Report of the Survey of Adult Skills describes the technical details of the survey, including the assessment and instrument design; survey platform development; field operations and quality control; sampling and weighting; and, data analysis and data products.

The U.S. PIAAC Technical Report, published by NCES, differs from the International Technical Report in that it gives U.S.-specific details on sampling, field operations, and platform development.

The OECD Technical Report can be accessed by clicking on the corresponding underlined screen text.

**Slide 12 of 24**

The International Codebook is a guide that describes the layout of the variables in the data file, and documents the values associated with the response options for each survey question. It can be accessed by clicking on the corresponding underlined screen text.

Here is an example item from the International Codebook, which displays the coding values for question B\_Q01a from the PIAAC background questionnaire. Each row contains the variable name, variable label, and coding values, including the reserved coding for missing data.

Users can toggle between the 'Variables' and 'Values' worksheets within the International Codebook Excel file.

**Slide 13 of 24**

The SAS and Stata data analysis manuals provide instructions for using the SAS and Stata macros, including detailed explanations of the options and parameters in the macros.

Macros can be used to calculate statistically valid estimates of percentages, means, proficiency levels, and percentiles, as well as to conduct regression analysis.

The SAS and Stata data analysis manuals can be accessed by clicking on the corresponding underlined screen text in the title of this slide, while the SAS and Stata macros can be accessed by clicking on either the underlined screen text 'SAS macros' or 'Stata macros'.

**Slide 14 of 24**

The PIAAC data compendia are sets of tables that provide previously computed, weighted summary statistics for both cognitive and background items. The purpose of the compendia is to provide a reference for the PIAAC data so that researchers can be sure that they are performing analyses correctly. Researchers may use the compendia to elaborate on their hypotheses, check the results of their own computations, and to search for variables.

It is important to note that, due to the design of the cognitive assessment, comparisons across countries **are not appropriate** for the cognitive item statistics provided for reporting purposes in the compendia.

**Slide 15 of 24**

For each categorical variable in the background questionnaire, the PIAAC data compendia provides information on the sample size, or weighted N; the weighted percent of responses by category; and, the percent missing, skipped, or refused responses. The data compendia for background variables can be accessed by clicking on the underlined screen text, 'background variables'.

Information provided for each cognitive item in the assessment includes the sample size, or weighted N; the weighted percent of correct and incorrect responses; and, the percent of respondents who were missing or did not reach the item. The data compendia for cognitive variables can be accessed by clicking on the underlined screen text, 'cognitive items'.

**Slide 16 of 24**

This is a sample item from the PIAAC Data Compendia. It shows results by country for the item that asks, "In the last week, did you do any PAID work for at least one hour, either as an employee or as self-employed?"

**Slide 17 of 24**

The documentation for the PIAAC survey in the United States provided by NCES includes the U.S. technical report, the U.S. public-use file and restricted-use file codebooks, and the User's Guide for the International Data Explorer, or IDE. The IDE User's Guide provides instructions for using the PIAAC IDE, as well as detailed information on variables contained in the data. Each of these resources can be accessed by clicking on the corresponding underlined screen text.

**Slide 18 of 24**

There are a variety of specialized tools available to conduct analyses of the PIAAC data. Web-based tools include the OECD International Data Explorer, or IDE, which contains national data files for all participating countries except Cyprus and allows users to easily create statistical tables and charts that can be exported for personal use. The NCES IDE is similar to the OECD IDE, but includes U.S.-only variables in addition to the international variables and data for all countries, except Australia and the Russian Federation. The NCES PIAAC Web Portal provides ready-made tables and charts for the most common PIAAC data queries.

**Slide 19 of 24**

More advanced statistical tools for use with the micro-data files include the International Database Analyzer, or IDB Analyzer, which is used with SPSS; SAS and Stata macros developed by the OECD; the AM statistical software developed by the American Institutes for Research, or AIR; and an experimental R-package developed by Statistics Austria.

These tools can be accessed by clicking on the corresponding underlined screen text. Let's look more closely at two of these tools.

**Slide 20 of 24**

The NCES International Data Explorer, or IDE, allows users to create customized statistical tables and charts using data from the PIAAC background questionnaire and cognitive assessments. Performance on the assessments can be analyzed by country, as well as in relation to background characteristics of the respondents. The IDE is a point-and-click method for producing custom results that does not require any advanced statistical knowledge or software, but produces statistically valid results. It can be used to calculate estimates of percentages, means, proficiency levels, and percentiles.

The NCES PIAAC IDE can be accessed by clicking on the corresponding underlined screen text.

**Slide 21 of 24**

The International Database Analyzer, or IDB Analyzer, is a stand-alone software program that facilitates the analysis of PIAAC data using SPSS as the computational

## Getting Started with the PIAAC Data

engine. It allows users to combine data from different countries for cross-country analysis, and to create an analysis dataset by selecting a specific subset of variables. The IDB Analyzer also accounts for PIAAC's complex sample design, including the correct calculation of standard errors. It provides several different procedures for analysis, including the computation of percentages, means, proficiency levels, percentiles, and regression models.

The IDB Analyzer can be freely downloaded by clicking on the corresponding underlined screen text. It will also be discussed in more depth in the module titled "Considerations for Analysis of PIAAC Data".

### Slide 22 of 24

Results from PIAAC can be linked with results from the IALS and the ALL surveys. Micro-data files containing data from IALS and ALL can be downloaded from Statistics Canada by clicking on the corresponding underlined screen text.

These files contain results for literacy from both IALS and ALL, as well as results for numeracy from ALL that use the same scales as the results from the PIAAC Survey of Adult Skills. The process for linking results from PIAAC and those from IALS and ALL is described in Chapter 17 of the OECD Technical Report, which can also be accessed by clicking on the corresponding underlined screen text.

Both the NCES and OECD PIAAC IDEs contain trend data from IALS and ALL as well.

### Slide 23 of 24

This module has introduced users to the data collected for PIAAC. It provided an overview of the micro-data files and data tools that are available for analysis of PIAAC data.

### Slide 24 of 24

Additionally, important resources that have been provided throughout the module are summarized here along with the module's objectives for your reference.

You may now proceed to the next module in the series or click the exit button to return to the landing page.