

U.S. PIAAC Synthetic Restricted Use File User Guidance

The U.S. PIAAC synthetic restricted use file (S-RUF) is available to researchers outside the U.S. to prepare computer code for the analysis of PIAAC data on the U.S. restricted use file (RUF) for the 2012/2014 household sample. The structure of the S-RUF is similar to the actual restricted use file (RUF) to allow researchers to develop and test computer code. Compared to the synthetic file for the U.S. PIAAC Round 1 (2012), this file contains more records and additional variables to account for both the U.S. PIAAC 2012 and 2014 household study.

Because the data on the S-RUF is synthetic, no conclusions can be drawn from the analysis generated from the computer code. In order to make data-driven analytical decisions, researchers need to submit their SAS, SPSS, or Stata code to piaac@air.org, where the code will be processed using the real U.S. RUF. The output will undergo a confidentiality review and will be returned to the researcher after approval.

Contents of the U.S. PIAAC synthetic restricted use file (S-RUF)

The S-RUF includes variables that were released in the U.S. PIAAC 2012/2014 household sample RUF with synthetic data values. Two types of variables are excluded from the S-RUF but are in the RUF: (1) variables with open-ended/verbatim responses, and (2) variables that were coded to detailed levels (e.g., occupation). A complete list of such variables can be found in Appendix A of this document.

A codebook that contains descriptions for the values of each variable and a SAS format program are also provided at the same location of the S-RUF. In addition to a SAS dataset, an SPSS version of the file is available as well. STATA users can convert the SPSS version to STATA by the following steps: Open the SPSS file in SPSS, click on “File” then select “Save As...”. Once the “Save Data As” window pops up, in the “Save as type:” section, scroll down to the desired STATA version needed, then click on the Save button.

The S-RUF has the same number of records as in the U.S. PIAAC 2012/2014 household sample PUF. Compared to the U.S. PUF available at the U.S. National Center for Education Statistics website (<https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2016667>), the U.S. S-RUF includes additional variables and more detailed variables. The additional variables include, for example:

- Age, continuous
- Income, continuous
- More detailed information on education qualifications and activities
- More detailed information on current work and work history
- More detailed information on family members
- More detailed information on origin and race
- More detailed information on languages spoken
- More detailed information on urbanicity
- Year and month of interview
- Additional sampling and weighting information
- Education-related variables corrected for U.S. routing. Note for the synthetic file, the variables corrected for U.S. routing have the same values as the original variables, but the actual values in RUF are different between the US routing variables and original ones. For more information about the U.S. routing

variables, see appendix E of the U.S. PIAAC 2012/2014 technical report at:
<https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2016036>

The missing value code scheme for the S-RUF is the same as for the RUF. For details on the missing value code scheme, please see table E-1 of the U.S. PIAAC 2012/2014 technical report. Note the SPSS file has a different missing value scheme from the SAS file. It is recommended for STATA users to use the SPSS file instead of SAS file to create a STATA file in order to preserve the missing value codes.

Remote analysis protocol for the U.S. PIAAC synthetic restricted use file (S-RUF)

The U.S. PIAAC S-RUF is released for academic research by researchers outside the U.S. only. You can download the dataset from the OECD website (<http://www.oecd.org/site/piaac/publicdataandanalysis.htm>), then follow the remote analysis protocol as described below:

1. A statement of purpose is provided by the researcher. Include a summary of the requested output, any subsetting of the dataset, and what the output will be used for (presentation, report, etc).
2. The researcher creates code in SAS, SPSS, or Stata to conduct an analysis using the PIAAC S-RUF that is available from the OECD website. The researcher should keep in mind the confidentiality edits that will be conducted by the processing analyst (PA) according to the statistical standards at the National Center for Education Statistics (<http://nces.ed.gov/statprog/2012/pdf/Chapter4.pdf>, accessed 16 November 2016). For example, the following is not allowed:
 - a. release of subsets in the form of datasets or output listings of individual records;
 - b. deduction of small cells or individual information;
 - c. extreme values (e.g., minimum and maximum values from a Proc Univariate in SAS);
 - d. unweighted counts.
3. The researcher provides the code to process on the real PIAAC RUF to the remote analysis center via email (email address: piaac@air.org).
4. The PA in the remote analysis center runs the code provided by the researcher and screens the log for errors. If errors occur, the PA corresponds with the researcher until a clean run is completed.
5. The PA reviews the output according to confidentiality rules. Some output may not be returned due to confidentiality edits. The screened and approved output is then provided to the researcher.

Appendix A. Variables in the RUF but not in the S-RUF

a. Items with open-ended/verbatim responses

Variable	Description
B_S01a1	Education - Highest qualification - Name of foreign qualification
B_S01a2	Education - Highest qualification - Country of foreign qualification (other)
C_S04j	Current status/work history - Last month - Ways of looking for work - Other specified
D_Q01a	Current work - Job title
D_Q01b	Current work - Responsibilities
D_Q02a	Current work - Kind of business, industry or service
D_Q02b	Current work - Main product of firm or organisation
D_S09	Current work - Other type of contract specified
E_Q01a	Last job - Job title
E_Q01b	Last job - Responsibilities
E_Q02a	Last job - Kind of business, industry or service
E_Q02b	Last job - Main product of firm or organisation
E_S08	Last job - Other type of contract specified
J_S04b	Background - Country of birth (other)
J_S05a1	Background - First learned language (other)
J_S05a2	Background - Second learned language (other)
J_S05b	Background - Language spoken at home (other)
B_Q01bUSX	Education - Highest qualification - Area of study verbatim
B_Q02cUSX	Education - Current Qualification - Area of Study Verbatim
B_Q05bUSX	Education - Formal Qualification - Area of Study Verbatim
B_S27eUSX	Activities - Class - Class attendance, other specify
B_Q29bUSX	Activities - Apprentice - Which trade or craft

b. Items that were coded to a very specific level (variables with more general coding are available on the PUF and S-RUF).

Variable	Description
ISCO08_C	Current Job Occupation - Respondent (ISCO 2008) (coded)
ISCO08_L	Last Job Occupation - Respondent (ISCO 2008) (coded)
ISIC4_C	Current Job Industry - Respondent (ISIC rev 4) (coded)
ISIC4_L	Last Job Industry - Respondent (ISIC rev 4) (coded)
LNG_L1	First language learned at home in childhood and still understood - Respondent (ISO 639-2/T) (coded)
LNG_L2	Second language learned at home in childhood and still understood - Respondent (ISO 639-2/T) (coded)
LNG_HOME	Language most often spoken at home - Respondent (ISO 639-2/T) (coded)
REG_TL2	Geographical region - Respondent (OECD TL2) (coded)
ISCO08_US	Trade or Craft - Respondent (ISCO 2008)
USCIP_C	Education - Current qualification - Area of study (coded)
USCIP_H	Education - Highest qualification - Area of study (coded)
USCIP_L	Education - Formal qualification in last 12 months - Area of study (coded)