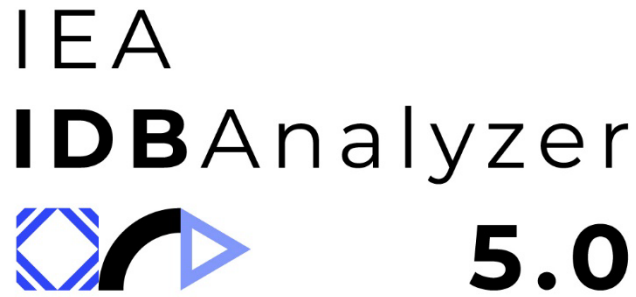


Help Manual for the IEA IDB Analyzer (Version 5.0)¹



©IEA 2023
www.iea.nl

¹ Please cite as: IEA (2025). Help Manual for the IEA IDB Analyzer (Version 5.0). Hamburg, Germany. (Available from www.iea.nl)

Table of Contents

About the IEA IDB Analyzer.....	4
About this Help Manual.....	6
The Convert Module	6
The Merge Module	6
The Analysis Module.....	6
Selecting Cases for Analysis.....	8
Significance Testing.....	8
What's New in the IEA IDB Analyzer (Version 5.0)?.....	8
Installing the IEA IDB Analyzer.....	9
Installation Notes.....	9
System Requirements for PC	9
Preparing to Run the IEA IDB Analyzer in a Mac Environment	10
Working with SPSS.....	10
Working with R.....	10
Package dependencies when using R	10
Merging Files with R prior to using the IDB Analyzer	11
Variable and Value Labels when using R.....	11
Starting the Application.....	11
Using the Convert Module.....	13
Using the Merge Module.....	14
Using the Analysis Module.....	18
Output Files.....	19
Handling of Special Characters in File and Path Names.....	20
Computing Percentages Only.....	21
Computing Percentages and Means of non-Plausible Values.....	24
Computing Percentages and Means of Plausible Values.....	28
Computing Percentages of the Population Meeting User-Specified Benchmarks.....	32
Computing Percentages within Benchmarks.....	37
Computing Percentiles.....	38
Computing Percentiles of Plausible Values.....	41
Computing Pearson Correlation Coefficients.....	45
Computing Pearson Correlation Coefficients with Plausible Values.....	48
Computing Spearman Rank-Order Correlation Coefficients.....	52
Computing Linear Regression Coefficients.....	55
Computing Linear Regression Coefficients with Plausible Values.....	60
Computing Logistic Regression Coefficients.....	65
Computing Logistic Regression Coefficients with Plausible Values.....	69
Troubleshooting.....	74
Appendix A: Working with IEA Data from Multiple Years.....	75
Appendix B: Cut points for International Proficiency Levels.....	76
ICCS – Civics and Citizenship Scale	76
ICILS – Computer and Information Literacy Scale.....	76
LaNA – Mathematics and Reading Scales	76
PIAAC – Literacy Scale.....	76
PIAAC – Numeracy Scale	78
PIAAC – Problem Solving in Technology Rich Environments Scale	78
PIAAC – Adaptive Problem Solving	78

PIRLS – Reading Literacy Scale.....	78
PISA – Collaborative Problem-Solving Scale	79
PISA – Financial Literacy Scale	79
PISA – Mathematics Literacy Scale.....	79
PISA – Problem Solving Scale	79
PISA – Reading Literacy Scale.....	80
PISA – Science Literacy Scale	80
TIMSS – Mathematics and Science Scales (Grades 4 and 8)	81
TIMSS Advanced – Advanced Mathematics and Physics Scales	81
Appendix C: Interpreting Contrast Coded Variables in Linear Regression	82
Dummy Coding	82
Effect Coding.....	82
Other Types of Coding	83
Appendix D: Working with TALIS Data from across ISCED Levels	84
Appendix E: Calculating Standard Errors in TIMSS and PIRLS.....	85
Appendix F: Analysis Notes for Specific Studies.....	86
OECD - PISA.....	86
U.S. NAEP.....	86
Appendix G: Calculating Design Effects.....	87
Appendix H: Using the IDB Analyzer with SAS OnDemand for Academics.....	89
About SAS OnDemand for Academics	89
Using the IDB Analyzer with SAS ODA.....	89
Merging Files	91
Conducting Analysis.....	91
Appendix I: Analysis Settings.....	93
Appendix J: IEA’s IDB Analyzer License Agreement.....	96

About the IEA IDB Analyzer

The IEA's International Database Analyzer (IEA IDB Analyzer) is an application developed by the International Association for the Evaluation of Educational Achievement (IEA) that can be used to combine and analyze data from all IEA's large-scale assessments, as well as analyze data from most major large-scale assessment surveys, including those from the U.S. National Assessment of Educational Progress (NAEP).

The IEA IDB Analyzer consists of two main modules: the Merge Module and the Analysis Module. These two modules are integrated in one single application. A third module introduced with Version 5.0 is one to convert SPSS system files to R data files. This module is activated when the user selects R as the statistical software of choice.

The Merge Module creates syntax that can be used to combine files from across different countries and levels (student, teacher, school, etc.). The Analysis Module creates syntax that can be used to perform analysis with these international databases. The syntax generated uses information from the sampling design in the computation of sampling variance and handles multiple plausible values. The code generated by the IEA IDB Analyzer enables the user to compute descriptive statistics and conduct statistical hypothesis testing among groups in the population without having to write any programming code.

At its core, the IEA IDB Analyzer is a graphical user interface that reads the contents of data files, and generates syntax in SPSS, SAS, or R to combine data files and conduct analysis with these data. In order to use the IEA IDB Analyzer, you will need to have SPSS, SAS, or R installed in your machine. Figure 1 below lists, in alphabetical order, the studies that have been configured in the IEA IDB Analyzer.²

Figure 1: Studies that can be analyzed with the IEA IDB Analyzer (in alphabetical order)

Study/Organization	Capability	SPSS	SAS	R
3S/OECD: Starting Strong Teaching and Learning International Survey	Merge and Analysis	Y	Y	Y
ALLS/StatsCan & OECD: Adult Literacy and Life Skills Study ³	Analysis	Y	Y	N
CivEd/IEA: Civics Education Study	Merge and Analysis	Y	Y	Y
IALS/OECD: International Adult Literacy Study	Analysis	Y	Y	N
ICCS/IEA: International Civics and Citizenship Education Study	Merge and Analysis	Y	Y	Y
ICILS/IEA: International Computer and Information Literacy Study	Merge and Analysis	Y	Y	Y
IELS/OECD: International Early Learning and Child Well-being Study	Merge and Analysis	Y	Y	Y
NAEP/U.S. National Assessment of Educational Progress	Analysis ⁴	Y	Y	N

² Upon request, the IEA IDB Analyzer can be configured to work with other large-scale assessment databases. For further information please contact the IEA at idb-analyzer@iea-hamburg.de.

³ When working with ALLS and IALS data you will need to rename the replicate weight variables to eliminate leading zeroes from the replicate number.

⁴ Please see the Appendix of this Help Manual for the available options.

Study/Organization	Capability	SPSS	SAS	R
PIAAC/OECD: Programme of International Assessment of Adult Competencies	Merge and Analysis	Y	Y	Y
PIRLS ⁵ /IEA: Progress in International Literacy Study	Merge and Analysis ⁶	Y	Y	Y
PISA/OECD: Programme for International Student Assessment	Analysis ⁷	Y	Y	Y
PRIDI/IADB: Project on Child Development Indicators	Merge and Analysis	Y	Y	Y
REDS/UNESCO: Responses to Educational Disruptions	Merge and Analysis	Y	Y	Y
SITES/IEA: Second Information Technology in Education Study	Merge and Analysis	Y	Y	Y
TALIS/OECD: Teachers and Learning International Study	Merge and Analysis	Y	Y	Y
TEDS-M/IEA: Teacher Education Study	Merge and Analysis	Y	Y	Y
TIMSS ⁸ /IEA: Trends in International Mathematics and Science Study	Merge and Analysis ⁹	Y	Y	Y

The IEA IDB Analyzer is licensed free of cost, not sold, and is for use only in accordance with the terms of the licensing agreement. While you can use the software for free, you do not have ownership of the software itself or its components, including the accompanying macros, and you are only authorized to use these macros in combination with the IEA IDB Analyzer, unless explicitly authorized by the IEA.

The software and license expire at the end of each calendar year when you will again have to download and reinstall the most current version of the software and agree to the new license. The complete copy of the licensing agreement is included in an Appendix of this Help Manual.

Support for the IEA IDB Analyzer can be obtained by contacting IEA's IDB Analyzer support team at idb-analyzer@iea-hamburg.de.

⁵ This includes PIRLS, ePIRLS and PIRLS Literacy, as well as the TIMSS/PIRLS Combined Study.

⁶ As of calendar year 2016, official PIRLS publications calculate standard errors using the "FULL" method. This change will be applied retroactively to data from previous cycles. For more information about this change, please refer to the Appendix of this Help Manual.

⁷ Please see the Appendix of this Help Manual for the available options.

⁸ This includes TIMSS, TIMSS Numeracy, as well as the TIMSS/PIRLS Combined Study.

⁹ As of calendar year 2016, official TIMSS publications calculate standard errors using the "FULL" method. This change will be applied retroactively to data from previous cycles. For more information about this change, please refer to the Appendix of this Help Manual.

About this Help Manual

The IEA IDB Analyzer can be used to merge and analyze datasets from a diverse array of studies. This Help Manual uses data from the PIRLS 2011 assessment for the examples. Examples on using data for specific IEA studies can be found in the “User Guide for the International Database” for the specific study. Documentation on the procedures used for calculating statistics and standard errors can be found in the corresponding technical documents for the study.

Throughout this Help Manual we present several example analyses. These are copied to the installation folder during installation of the program.

The Convert Module

The Convert Module is used to convert SPSS system files (.sav) into R data files (.RData). When running the Convert Module, the IEA IDB Analyzer creates R syntax that converts all SPSS system files in a directory into R data files.

The Merge Module

The Merge Module is used to combine data files from different countries¹⁰ and when necessary, merge data files from different sources (e.g., student background questionnaires and achievement files, student background files with teacher or school level files, etc.). It also allows the user to select individual or sets of variables to create a smaller and more manageable dataset.

When running the Merge Module, the IEA IDB Analyzer creates syntax that merges and combines files specified by the user, keeping only the selected variables. The syntax is automatically saved as a text file and opened in a window of the statistical software selected at the beginning of the process. The data files created using the Merge Module can be processed either with the Analysis Module of the IEA IDB Analyzer or by any other analysis software that accepts these files as input.

The Merge Module is only available to use with IEA databases and others in which the data are published separate by country (i.e., OECD-TALIS, OECD-PIAAC, IADB-PRIDI, etc.). Please refer to Figure 1 for a complete list of compatibility with studies.

The Merge Module is only designed to combine data from the same assessment year. To combine data across 2 or more years, please refer to the Appendix of this Help Manual.

The Analysis Module

The Analysis Module of the IEA IDB Analyzer provides procedures for the computation of means, percentages, standard deviations, correlations, and regression coefficients for any variable of interest overall for a country, and for specific subgroups within a country. It also provides procedures to compute percentages of people in the population that are within, at, or above benchmarks of performance or within user-defined cut points in the proficiency distribution, percentiles based on the achievement scale, or any other continuous variable.

The Analysis Module can be used to create syntax to analyze data files from the studies mentioned above, regardless of whether they have been preprocessed with the IEA IDB Analyzer Merge Module. The Analysis Module can create syntax for several analysis procedures. Like the Merge Module, the Analysis Module creates syntax that computes the statistics specified by the user. The following analyses can be performed with the Analysis Module:

Analysis Type	SPSS	SAS	R
1. Percentages only: Computes percentages by subgroups defined by the user.	Yes	Yes	Yes

¹⁰ Throughout this manual, the term “country” is used to refer to reporting jurisdictions in the corresponding study. These include countries per se, “states”, “provinces”, “benchmarking participants”, “partner countries”, etc. Each study has adopted its own convention for naming and identifying these entities, but in general these are identified by a unique value in the variables usually called IDCNTRY, CNT, CNTRYID, etc.

Analysis Type	SPSS	SAS	R
2. Percentages and Means: Computes percentages, means, design effects and standard deviations for selected variables by subgroups defined by the user. The percentage of missing responses is included in the output. It also performs the computation of t-test statistics of group mean and percent differences taking into account sample dependency.	Yes	Yes	Yes
3. Benchmarks: Computes the percent of the population in two modes: a) as the percent within a group meeting a set of user-specified performance or achievement benchmarks, or b) as the percent within a benchmark that belong to a specific group of the population. It computes the percentages within a group meeting the benchmarks in two modes: cumulative (percent of the population at or above given points in the distribution) or discrete (percent of the population within given points of the distribution). It can also compute the mean of an analysis variable for those within a particular group when the discrete option is selected. In doing this, it performs the computation of group mean differences and percent differences between groups taking into account sample dependency.	Yes	Yes	Yes
4. Percentiles: Computes the score points that separate a given proportion of the distribution of scores by subgroups defined by the grouping variable(s).	Yes	Yes	Yes
5. Correlations: Computes Pearson or Spearman Rank-Order correlation for selected variables by subgroups defined by the grouping variable(s). The IEA IDB Analyzer can also compute the correlation between sets of plausible values.	Yes	Yes	Yes
6. Linear Regression: Computes linear regression coefficients for selected variables predicting a dependent variable by subgroups defined by the user. The IEA IDB Analyzer has the capability of including plausible values as dependent or independent variables in the linear regression equation. It also has the capability of contrast coding categorical variables (dummy or effect) and including them in the linear regression equation.	Yes	Yes	Yes ¹¹
7. Logistic Regression: Computes logistic regression coefficients for selected variables predicting a dependent dichotomous variable, by subgroups defined by the user. The IEA IDB Analyzer has the capability of including plausible values as independent variables in the logistic regression equation. It also has the capability of contrast coding categorical variables and including them in the logistic regression equation. Under the SAS version you can run multinomial logistic regressions.	Yes	Yes	No

When calculating these statistics, the IEA IDB Analyzer has the capability of using any continuous or categorical variable in the database or make use of proficiency estimates in the form of plausible values. When using plausible values, the IEA IDB Analyzer generates code that takes into account the multiple imputation methodology in the calculation of the variance for statistics, as it applies to the corresponding study.

All procedures offered within the Analysis Module of the IEA IDB Analyzer also make use of appropriate sampling weights and standard errors of the statistics that are computed according to the variance estimation procedure required by the design as it applies to the corresponding study.

¹¹ Currently only listwise deletion is available for R.

Before conducting data analysis with the IEA IDB Analyzer, we recommend you become familiar with the specifics of the study design of interest. Each study has its own technical documentation available online. The Appendix of this Help Manual shows the analysis settings as implemented in the IEA IDB Analyzer. The selection of the values for these parameters is based on the recommendations made in the corresponding technical documentation.

Videos are also available illustrating examples of how to use the IEA IDB Analyzer. These videos are available from <http://www.iea.nl>.

Selecting Cases for Analysis

All the analysis types have the option to manually subset the data used for the analysis. By using the “SELVAR” and “SELCRIT” parameters in the macros produced by the IEA IDB Analyzer, the user can conduct analysis using only a subset of the data. For example, by setting “SELVAR = YEAR” and “SELCRIT = (year = 1990)”, a user can conduct the analysis only with those cases for which the variable YEAR equals 1990¹². Please note that when using a selection criterion, the program still needs to read through the entire file to search for the records that meet the selection criteria, so the gain in processing speed will depend on the size of the file that contains all the cases. When selecting cases using multiple variables you will need to list them individually in the “SELVAR” statement (for example “SELVAR = var1 var2 /”). When using the SELCRIT option you need to write the syntax using the same format you’d use to write an IF statement in SAS or SPSS, or a filter function in the R package dplyr. This can include AND, &, |, or OR operators, or any other available operator for selecting a subset of the cases. Please refer to the corresponding documentation of the statistical software of choice for more information on acceptable syntax.

Significance Testing

Most of the analysis types, except linear and logistic regression, calculate significance testing statistics. These are explained throughout this manual. Depending on the specific analysis you are conducting, these calculations could take a relatively long time. If you want to suppress the calculation of significance testing statistics, you can use the parameter DOSIG = N. This will suppress the calculation of any significance testing statistics.

What’s New in the IEA IDB Analyzer (Version 5.0)?

Version 5.0 of the IEA IDB Analyzer replaces Version 4.0, which is not available or supported any longer. ALL functionality of version 4.0 has been preserved in version 5.0.

The main differences between Version 4.0 and Version 5.0 are the following:

- Introduction of functionality with R.
- Introduction of the Convert Module to convert files from SPSS sav to RData format.
- The Merge Module and the Analysis Module have been updated to recognize files from newer studies since Version 4.0 was last released.

¹² When selecting using string/text variables, you will need to enclose values in single quotes ('). Double quotes will cause an error in processing.

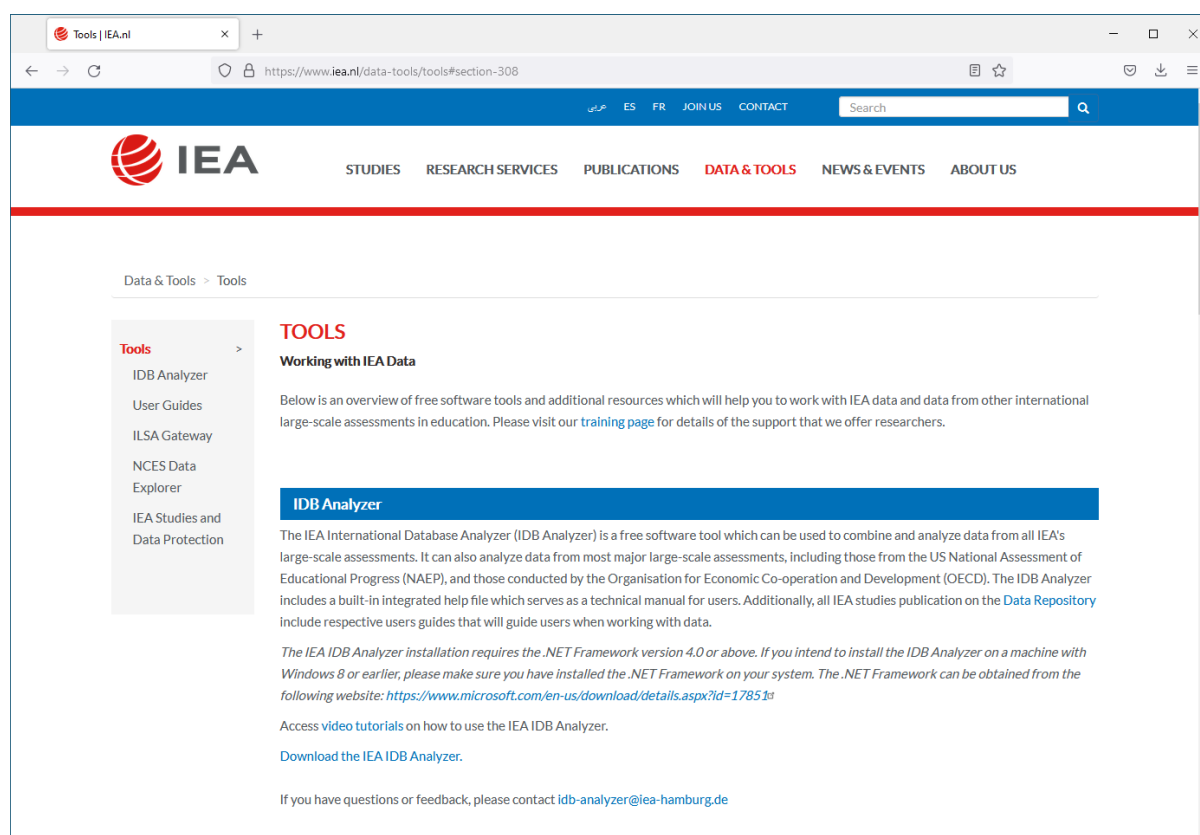
Installing the IEA IDB Analyzer

The most current version of the IEA IDB Analyzer is always available, free of charge, from the IEA website (<http://www.iea.nl>, see Figure 2).

To install the IEA IDB Analyzer on your computer we recommend you take the following steps:

1. Download or copy the installation file to a directory of your choice.
2. Double click on the IEA IDB Analyzer installation file to start the installation process.
3. When installing the IEA IDB Analyzer, you will be asked if the previously user-defined country list shall be deleted. If you choose “Yes” the default version of the country list will be added. We recommend choosing the default setting “Yes” as this ensures that your IDB Analyzer version also recognizes data files from newly added studies.
4. Choose the default directory for the installation suggested by the IEA IDB Analyzer setup: C:\Users\- 5. Once the installation process is completed you are ready to use the IEA IDB Analyzer.

Figure 2: Screenshot of IEA IDB Analyzer Download Page¹³



Installation Notes

System Requirements for PC

The IEA IDB Analyzer will work on most computers using the most recent versions of the Microsoft Windows operating system. You will also need SPSS, SAS, or R and RStudio installed in your machine to fully execute the code created by the IEA IDB Analyzer.

¹³ Actual look of this web page might have changed after the publication of this Help Manual.

Preparing to Run the IEA IDB Analyzer in a Mac Environment

Currently there is no standalone Mac version of the IEA IDB Analyzer. However, the software can be used on Mac through a virtual machine and Windows installed on it. The current version was tested using Windows installed on Parallels Desktop for Mac (<http://www.parallels.com/products/desktop/>), although other virtual machine software, e.g. VirtualBox (<https://www.virtualbox.org/>) could also be used. In order to install a working copy of the IEA IDB Analyzer on a MAC, please follow these steps:

1. Install the virtual machine software on your Mac computer.
2. Install Windows on the virtual machine.
3. Install SPSS, SAS, or R and RStudio on the Windows virtual machine.
4. Install the IEA IDB Analyzer on the Windows virtual machine following the steps described in the previous section.

Please note that running Windows on a virtual machine can increase the hardware requirements of your Mac computer. Running the IEA IDB Analyzer, and the SPSS macros it produces, will be possible only using the Windows operating system installed on the virtual machine.

Working with SPSS

On some computers SPSS and the IEA IDB Analyzer cannot read labels in languages other than English. This depends on the SPSS character encoding settings. To overcome this problem, do the following:

1. Start SPSS (**Start -> IBM SPSS Statistics ## -> IBM SPSS Statistics ##**).
2. Open the SPSS options (Edit > Options) dialog box.
3. Under the General or Language tab (depending on your SPSS version), click on the radio button Unicode (universal character set) in the Character Encoding for Data and Syntax group.
4. Click the **OK** button to apply the settings and close the dialog box.
5. Close SPSS.
6. Start SPSS again.

Working with R

To enable the R functionality within the IEA IDB Analyzer, follow these steps:

1. Install the latest free version of R Studio Desktop by visiting the following webpage:
<https://posit.co/download/rstudio-desktop/#download>
2. Install the latest version of R by visiting the following webpage:
<https://cran.r-project.org/bin/windows/base/>
3. Verify the installation directory for R. After installing R, ensure it is located in the following directory:
C:\Program Files\R\R-[Version Number]

This directory is crucial for the IEA IDB Analyzer to locate the installed version of R. If the directory structure differs, the IEA IDB Analyzer may encounter difficulties in identifying the R version on your machine. Consult your IT administrator to ensure you have the necessary administrator rights for installation.

Ensure that only one version of R is available on your computer to prevent any conflicts.

Package dependencies when using R

When using R, the IEA IDB Analyzer checks automatically if the user has installed the R packages that are needed to run the analysis of choice. If the user has not installed one of the required R packages, the user is asked in a pop-up window if the missing R package shall be downloaded and installed.

Since R is a rapidly evolving free-of-charge software with many people contributing to its development, compatibility issues between packages can sometimes occur.

Therefore, we strongly recommend that users update the required R packages on a regular basis. The IEA IDB Analyzer checks automatically if the user has installed the same or a newer version of packages that were used to verify the current version of the IEA IDB Analyzer. The user will be informed in a pop-up window if a package needs to be updated.

Merging Files with R prior to using the IDB Analyzer

Please note that the Merge Module uses the **left_join()** function, from the tidyverse/dplyr package. If you want to use in the Analysis Module files merged outside the Merge Module, you will need to use this function in your merge.

The code for merging files should look like this:

```
data_new <- left_join(data1, data2, by = c("A"="B"))
```

Variable and Value Labels when using R

The IEA IDB Analyzer R module requires all variables to have labels for the variable and the values of categorical variables. Using variables without labels will have unpredictable results.

Starting the Application

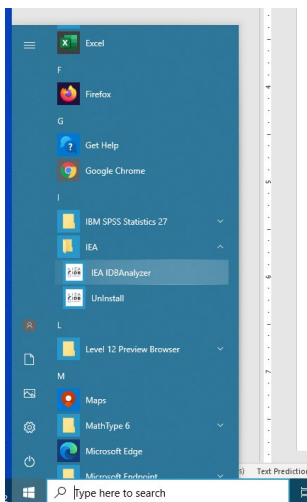
The IEA IDB Analyzer consists of three separate modules integrated in a single application: the Convert Module, the Merge Module and the Analysis Module.

You can start the IEA IDB Analyzer and access these modules by doing the following:

Start -> IEA -> IEA IDB Analyzer

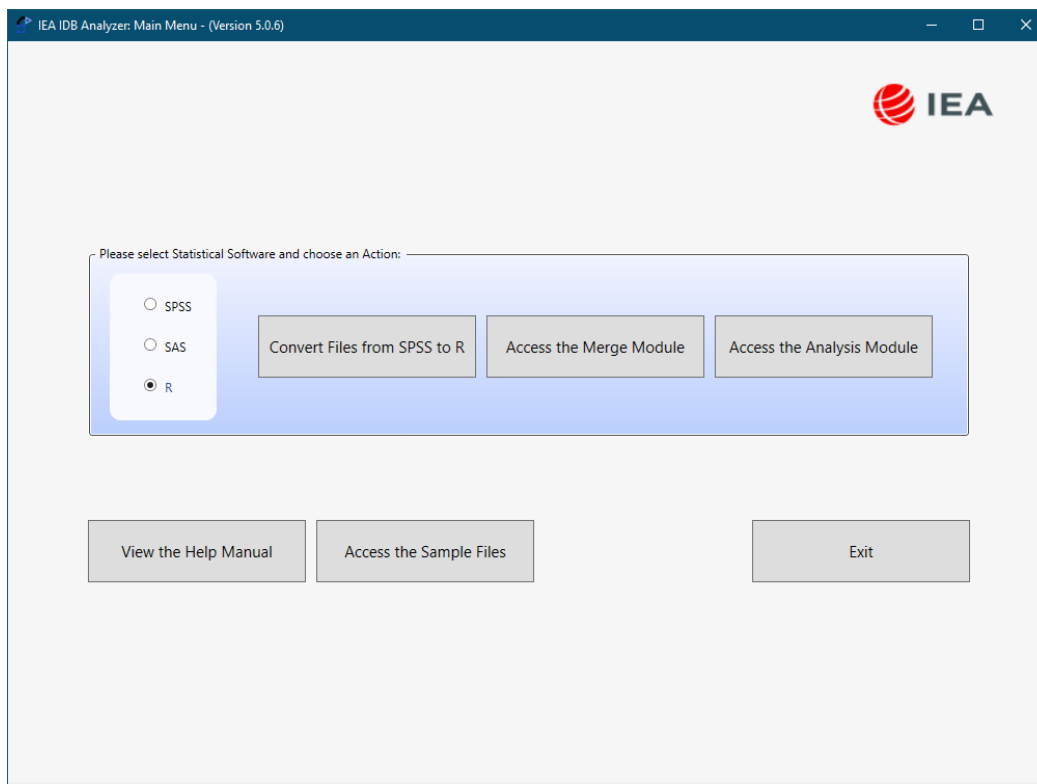
Figure 3 below shows the selection sequence for starting the application from the Start menu.

Figure 3: Starting the Application from the Start Menu



After starting the program and loading the main window of the application, you will have the option to select the statistical software you will be working with (SPSS, SAS, or R), and then choose between the Convert Module, the Merge Module and the Analysis Module, as shown in Figure 4. The Convert Module is only available when choosing R as the statistical software.

Figure 4: The Main Menu of the IEA IDB Analyzer



After selecting the statistical software of your choice (SPSS, SAS, or R), click on the corresponding button for the corresponding module. Usually, when using the IDB Analyzer, the first step is to prepare the data for analysis using the Merge Module, but when using R, you might first need to convert existing SPSS system files (.sav) to R data files.

Using the Convert Module¹⁴

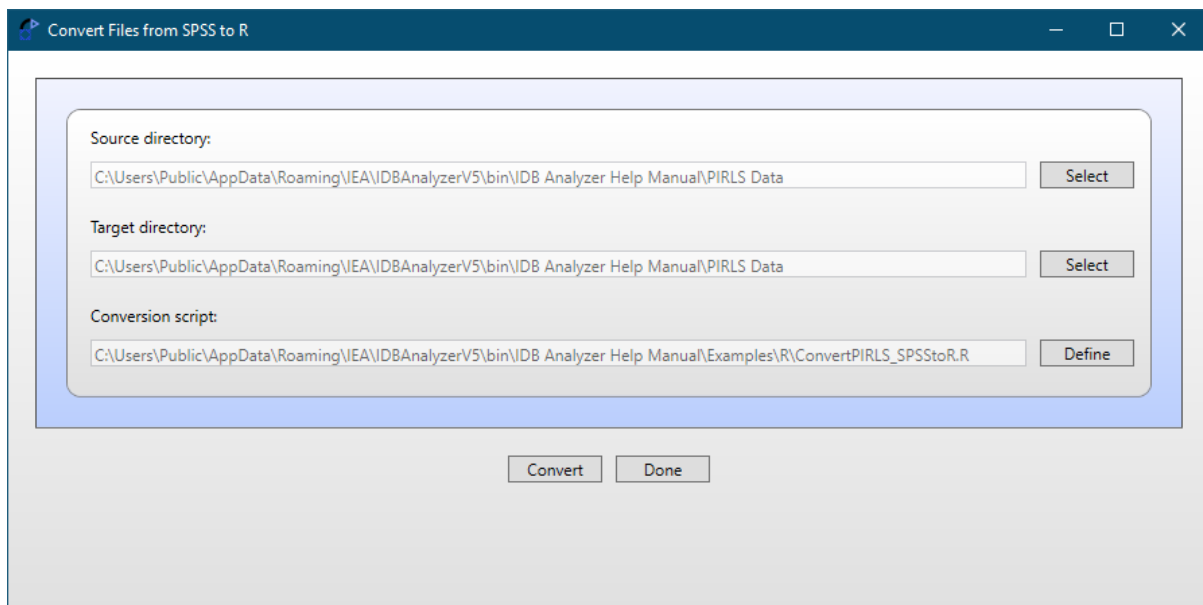
The Convert Module will convert all SPSS system files (those with the extension .sav) in a specified Source directory, to R Data frames with the extension .RData, and save them in the Target directory.

When using the Convert Module, you will need to follow these steps:

1. Start the IEA IDB Analyzer from the Start menu by selecting **Start -> IEA -> IEA IDB Analyzer**.
2. Select R as the statistical software you want to work with.
3. Click the **Convert Files from SPSS to R** button. This button is only active if you choose R as the **Statistical Software**.
4. Select the Source directory. This is where the source SPSS system files are located.
5. Select the Target directory. This is the directory where the SPSS system files converted to RData format will be saved.
6. Provide the name and location to use to save the Conversion script.
7. Click on the **Convert** button. In doing so, the IDB Analyzer will save the conversion script to the specified location and will also open R, or whichever program you have assigned as the default program to open files with the extension .R.
8. Select the conversion script created by the IDB Analyzer and run it.
9. Check that the files were converted and saved to the target directory.

Figure 5 shows the IEA IDB Analyzer Setup Screen for a sample conversion. The corresponding sample syntax and output files can be viewed by selecting the box **Access the Sample Files** from the main menu.

Figure 5: Sample setup for Convert module



Using the setup above, the IDB Analyzer will convert all *.sav files located in the **C:\Users\...\PIRLS Data** directory to *.RData files, will save these also in **C:\Users\...\PIRLS Data** and will save the conversion script to **C:\Users\...\R\ConvertPIRLS_SPSStoR.R**.

¹⁴ This module is only available when R is selected as the statistical software of choice.

Using the Merge Module

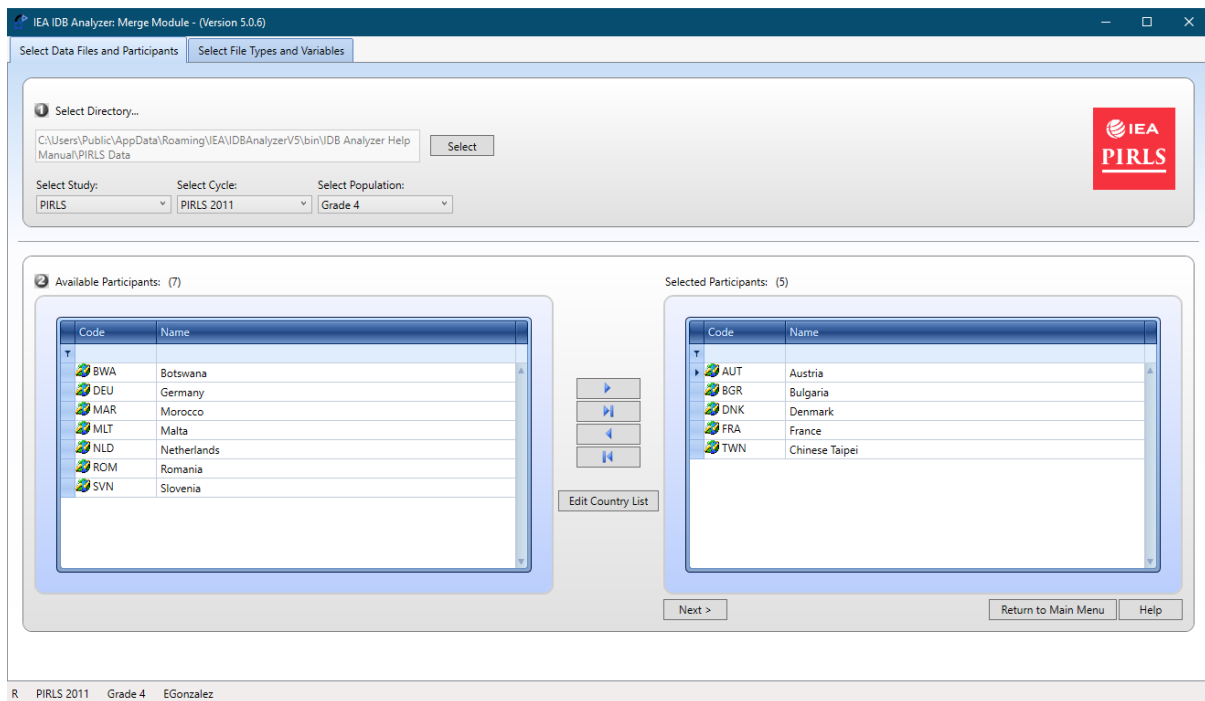
The Merge Module allows the user to combine data files from different countries, and data files collected from different sources within a country. The Merge Module is customized to recognize data files from all IEA studies and other OECD studies where the data are published separate by country (currently PIAAC and TALIS). Figure 1 shows in alphabetical order the databases that the IEA IDB Analyzer Merge Module is able to recognize at the time this Help Manual was prepared. The Merge Module will be configured to recognize other databases as they become available so if you do not find a more recent study listed here, please contact us for an update.

The Merge Module recognizes the data files for the specific study by reading the file names in the selected directory and matching them to the file naming convention pre-specified in the IEA IDB Analyzer configuration files. Because of this, we recommend that you do not save the merged file in the same directory where the source files are located as this might cause unexpected result. We also recommend that you keep files from different studies and years in separate directories as keeping them in the same location might have unexpected consequences.

When using the Merge Module, you will need to follow these steps:

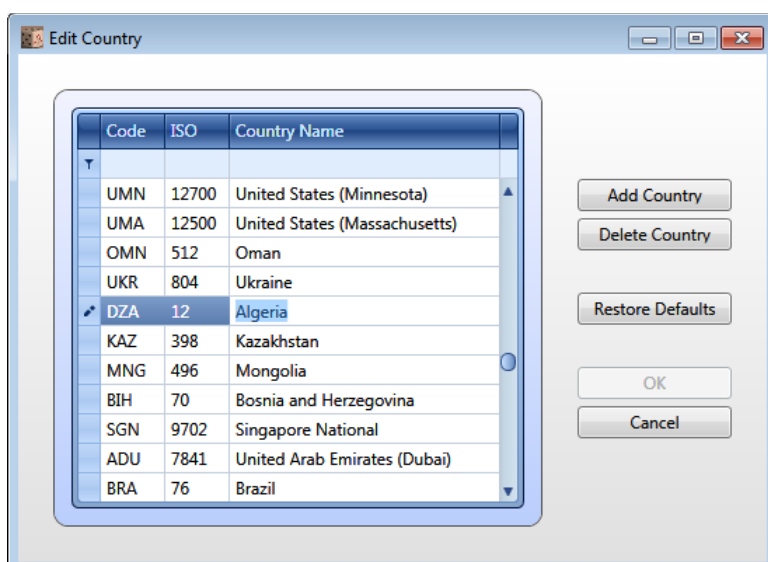
1. Start the IEA IDB Analyzer from the Start menu by selecting **Start -> IEA -> IEA IDB Analyzer**.
2. Select the Statistical Software you want to work with (Choose between R, SAS, or SPSS).
3. Click the **Access the Merge Module** button.
4. The Merge Module interface is divided into two distinct tabs. In the first one you can select the countries and edit country labels. In the second tab you can select the file types you want to merge, the variables you want to include in your analysis, and specify the name of the file where you want to save the resulting merge.
5. Under the **Select Data Files and Participants** section, select the directory where the individual country data files are located. In our example, we are using a subset of the PIRLS 2011 data located in "C:\users\public\...\PIRLS Data".
6. After making this selection, the program automatically recognizes and completes the study, survey, and grade information and lists all countries that are available in that directory as available participants for the merge. If the directory contains data for more than one study and more than one cycle and grade, the IEA IDB Analyzer will prompt you to select the study, survey, and grade of the files with which you want to work.
7. Select from the available participants those that you want to use for your analysis and move them to the panel on the right of the screen (Selected Participants). For multiple selections of countries, you need to hold the CTRL key of your keyboard when making selections. In our example, we have selected Austria, Bulgaria, Chinese Taipei, Denmark, and France. Notice that the countries appear sorted in alphabetical order by their 3-letter code. If you click the column header you can sort the list by country code or by country name. If you want to select all available countries, click on the arrow button with a vertical bar and all countries will be moved over to the right selection panel. The selections should look as seen in Figure 6. Typing text in the blue space between the column header and the country list will filter the entries from the corresponding column.

Figure 6: Merge Module Setup for Selecting Countries



- To edit the country labels, click on the **Edit Country List** button and you will be prompted with a list of country labels for editing. Figure 7 shows the **Edit Country List** window, which allows you to edit the names of the countries. Additionally, you can delete countries from the list by clicking on the **Delete Country** button or add countries by clicking on the **Add Country** button. When you add a new country to the list, you first need to add the three-letter country code, then the numerical ISO code, and the complete country name field. Clicking the **Save** button will save the changes and the **OK** button will close the **Country List Editor**. Please note that if you duplicate an existing country code, the IEA IDB Analyzer will display a warning message that there is a duplicate in the country list and will not implement the changes. Clicking the button **Restore Defaults** will revert the Country List to its default content.

Figure 7: Country List Editor



9. After selecting the country or countries of interest, click the **Next** button at the bottom of the screen or the tab at the top of the screen to move on to the **Select File Types and Variables** section. Here you can choose the file types for merging and the variables you want included in the merged data file. The selections on this screen will vary depending on the study files with which you will be working.
10. Selecting variables for a merge is a **two-step process**. First you need to select the file type by ticking the corresponding **checkbox** next to it. You then proceed to select the corresponding variables by moving them from the **Available Variables** to the **Selected Variables** section of the interface. For our example we will select each of the file types available: student background, student achievement and school background, and select all the available variables. Note that selecting a file type will only select a default set of ID and sampling variables, and a default set of scores, should these exist in the file. You will need to select some or all remaining variables, should you be interested in them. De-selecting a file type by unchecking the corresponding checkbox will automatically de-select all variables from that file type and exclude them from the merge.

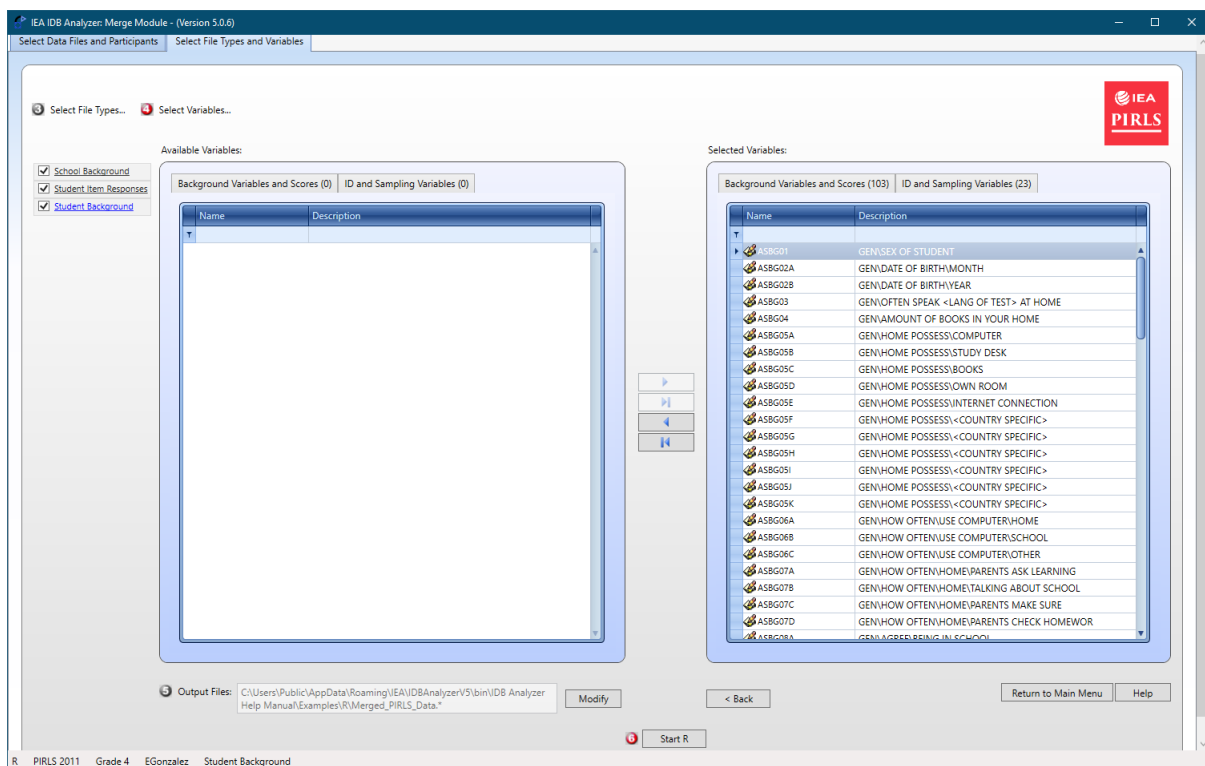
Not all combinations of file types are possible. Please refer to the technical documentation of the specific study for information about file types available, and allowable combinations.

NOTE: The list of variables for each file type is generated from the first available file of that type sorted alphabetically. This assumes all files of the same type have the same internal structure (variables, formats, labels, etc.). Any deviation from this can cause unexpected results. Should you want to modify the contents of a file for a country, or set of them, we recommend you do this on the resulting merged file, after the merge is completed, and not on the individual file.

11. You can select variables to merge using the arrow buttons in the middle of the screen, moving the variables from the **Available Variables** list on the left side to the **Selected Variables** list on the right. If you want to select all available variables, click on the arrow-bar button and all the variables will be copied over to the panel on the right of your screen.
12. When selecting the variables, you can search variables by variable name, or by description label using the filter boxes (light blue space between column header and list of variables) in the **Available Variables** list and **Selected Variables** list. You can also sort the variable list by name, description label, or location in the file by reselecting the file. By default, the variables appear sorted by location within the file.
13. After you have completed selecting files and variables, specify the name and the path of your syntax and merged data file under **Output Files**, clicking the **Define** button (this will change to **Modify** after you have selected a file destination for the first time). The file name can contain only alphanumeric characters (a-z, A-Z, 0-9) and underscores. It cannot contain any special characters and spaces. To avoid overwriting any of the original files, we recommend you save the merged file in a different directory from where the original files are located. We also recommend using a different convention for naming your merged file, although this is not mandatory.
14. Click on the **Start SPSS/SAS/R** button to create the SPSS/SAS/R syntax file and open it in an SPSS/SAS/R syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 8 shows the selected file types, variables, and the file name specified for the merged data from PIRLS 2011.

Figure 8: Merge Module Setup for Selecting File Types and Variables



For the examples that follow in this chapter, we have selected all available variables in the student background, student achievement, and the school background files. All the plausible values and sampling information available in the file has been automatically selected. The resulting data file will be located in the directory specified and will be named **Merged_PIRLS_Data**, and the R, SPSS, or SAS syntax that is created for generating this file is also named **Merged_PIRLS_Data** with the corresponding file extension. You can take a look at the resulting files by selecting the box **Access the Sample Files** from the main menu.

Once you have finished creating this file, you are ready to proceed with your analysis.

Please note the following:

- a. After running the syntax, you can verify your file was created by checking that there were no errors or warnings in the R, SPSS, or SAS output window, and looking in the target directory for the existence of the file.
- b. When creating the merged file, **SPSS** and **R** will create one or more temporary files. These are placed in the same directory as the merged file and are named "tmp*. *". You can safely delete these files once the merged file is created.
- c. When the SPSS code is executed, it creates temporary files in the target directory or where you are saving the files. For this reason, we recommend you always direct your work to a directory located in your local machine (rather than, for example, to a network drive or cloud storage). This avoids potentially overwriting of temporary files created by others and will considerably speed up processing of the files. By default, SAS and R use the local machine to create such temporary files and therefore does not present this problem.

Using the Analysis Module¹⁵

To access the Analysis Module, you need to click the corresponding button in the main menu of the IEA IDB Analyzer. If you are currently using the Merge Module, you need to click the **Return to the Main Menu** button located at the bottom right corner of the screen.

To open the Analysis Module, you first need to select in the **Main Menu**, if you want to work with **R**, **SPSS**, or **SAS**. Afterwards, click the Analysis Module button. The Analysis Module will automatically load the last merged or analyzed data file. You can choose a different file if needed.

The Analysis Module generates R, SPSS, or SAS syntax for the computation of means, percentages, standard deviations, correlations, and linear and logistic regression coefficients for any variable of interest for a country overall, and for specific subgroups within a country. It also computes percentages of people in the population that are within benchmarks of performance, or within user-defined cut points in the proficiency distribution, and the percent of people who have exceeded such benchmarks or cut points in the distribution, as well as user-defined percentiles for continuous variables.

Please note that when the SPSS code is executed, it creates temporary files in the target directory or where you are saving the files. For this reason, we recommend you to always direct your work to a directory located in your local machine. This avoids potentially overwriting of temporary files and will considerably speed up processing of the files. By default, SAS uses the local machine to create such temporary files and therefore does not present this problem.

Regardless of the analysis type you choose, there are some selections that need to be made for all analyses. Specifically, you will need to select the data files that contain the data you will be working with, the analysis type you want to conduct, the statistics you want to compute, whether you want to use plausible values for the analysis, and whether to include cases with missing values in any of the classification or grouping variables. By default, the IEA IDB Analyzer excludes those cases that have missing information for any of the classification or grouping variables. You can override this by deselecting this option.

For all statistic types you will need to use at least one grouping variable. By default, the program always performs analysis on a country-by-country basis, using the variable IDCNTY or equivalent as the default grouping variable¹⁶. You can add other grouping variables as your analysis requires.

The first step after entering the Analysis Module is to select the **Analysis Type** based on the data you have merged and selected when using the Merge Module. When selecting the **Analysis Type**, the IEA IDB Analyzer will check that the file has the necessary variables for the analysis. For each analysis you will need the sampling weight, and either the replicate weights or variables with replication information. If these are not found, the IEA IDB Analyzer will issue a warning message and not let you continue. For further information on these variables, and the analysis types possible for each study, please refer to the technical documentation corresponding to each study.

Please note that the **Analysis Types** are preconfigured in the IEA IDB Analyzer based on the analysis specifications of each of the studies. While new analysis types can be added to the configuration file, this is only possible by contacting the IEA software unit for further instructions.

Each analysis procedure using SPSS and R in the IEA IDB Analyzer calculates a so called "Table Average"¹⁷. This corresponds to the average of the statistics calculated across the countries and/or jurisdictions included in the file used for the analysis and included in the table. These countries and jurisdictions are uniquely identified by the country identification variable specific to the study (variable IDCNTY in most IEA studies). Depending on the set of countries and jurisdictions included in this file, the "Table Average" calculated by the IEA IDB Analyzer might or might not correspond to the "International Average" presented in the International Report for the corresponding study. To suppress

¹⁵ Please refer to the Appendix for notes on using te IDB Analyzer with R.

¹⁶ Throughout this manual, the term "country" is used to refer to reporting jurisdictions in the corresponding study. These include countries per se, "states", "provinces", "benchmarking participants", "partner countries", etc. Each study has adopted its own convention for naming and identifying these entities, but in general these are identified by a unique value in the variables usually called IDCNTY, CNT, CNTRYID, etc.

¹⁷ Currently only available in SPSS and R.

calculation of the table average you can set the parameter “INTAVG = N /” in the SPSS syntax created by the IEA IDB Analyzer.

Output Files

Depending on the software you are working with, you will get the following output files:

With SPSS:

Type	Generated by	Extension	Format	Open with
Syntax	IEA IDB Analyzer	*.sps	Text	SPSS or text editor
Results	SPSS	*.spv	SPSS Pivot Tables	SPSS
Results	SPSS	*.sav	SPSS Dataset	SPSS
Results	SPSS	*.xlsx	MS Excel	MS Excel
Results	SPSS	*.csv	Comma separated variables	MS Excel or other CSV file reader

With SAS:

Type	Generated by	Extension	Format	Open with
Syntax	IEA IDB Analyzer	*.sas ¹⁸	Text	SAS or other text editor
Results	SAS	*.lst	SAS Listing file	SAS or other text editor
Results	SAS	*.html	HTML	SAS or other HTML reader
Results	SAS	*.csv	Comma separated variables	MS Excel or other CSV file reader
Results	SAS	*.sas7bdat	SAS Dataset	SAS
Analysis Steps	SAS	*.log	SAS log file	SAS or other text editor

With R:

Type	Generated by	Extension	Format	Open with
Syntax	IEA IDB Analyzer	*.R	Text	R or R Studio
Results	R	*.html	HTML	Any HTML reader
Results	R	*.RData	R Data Frame	R
Results	R	*.xlsx	MS Excel	MS Excel

¹⁸ SAS syntax can be run in interactive or batch mode.

All output files will use the name specified for the syntax file created by the IEA IDB Analyzer and will be written to the same directory where the syntax file is written. The output files will be differentiated by the extension used for the file, and some additional text within the file name clarifying the contents of the file.

Handling of Special Characters in File and Path Names

While the IDB Analyzer can work in a wide variety of environments, it might not always handle some special characters correctly. Filenames provided by the user are used to name the output files, and therefore have a maximum length and can only use characters allowed by the R, SAS, or SPSS engines. If you encounter an error related to a filename, please review the corresponding documentation of your software. There are also some limitations with path names. Whenever possible avoid using special characters (such as #, \$, %, &, !, ?, quotes, or punctuation marks) in the path names you use for your input and output files.

Computing Percentages Only

To compute percentages of variables with the corresponding standard errors, you will need to select “Percentages only” from the **Statistic Type** dropdown menu.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables to be used to define the subgroups for analysis. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTRY or its equivalent as the first grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
Separate Tables by	This is the list of classification variables used to generate each table with results. When omitted, it will generate tables using only the grouping variables.
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

As an example, we will compute the percentages for boys and girls (ASBG01) and their standard errors within each country (IDCNTRY), using the weighting variable TOTWGT. The data will be read from the data file **Merged_PIRLS_Data** and the standard errors will be computed based on replicate weights.

The steps in the IEA IDB Analyzer are as follows:

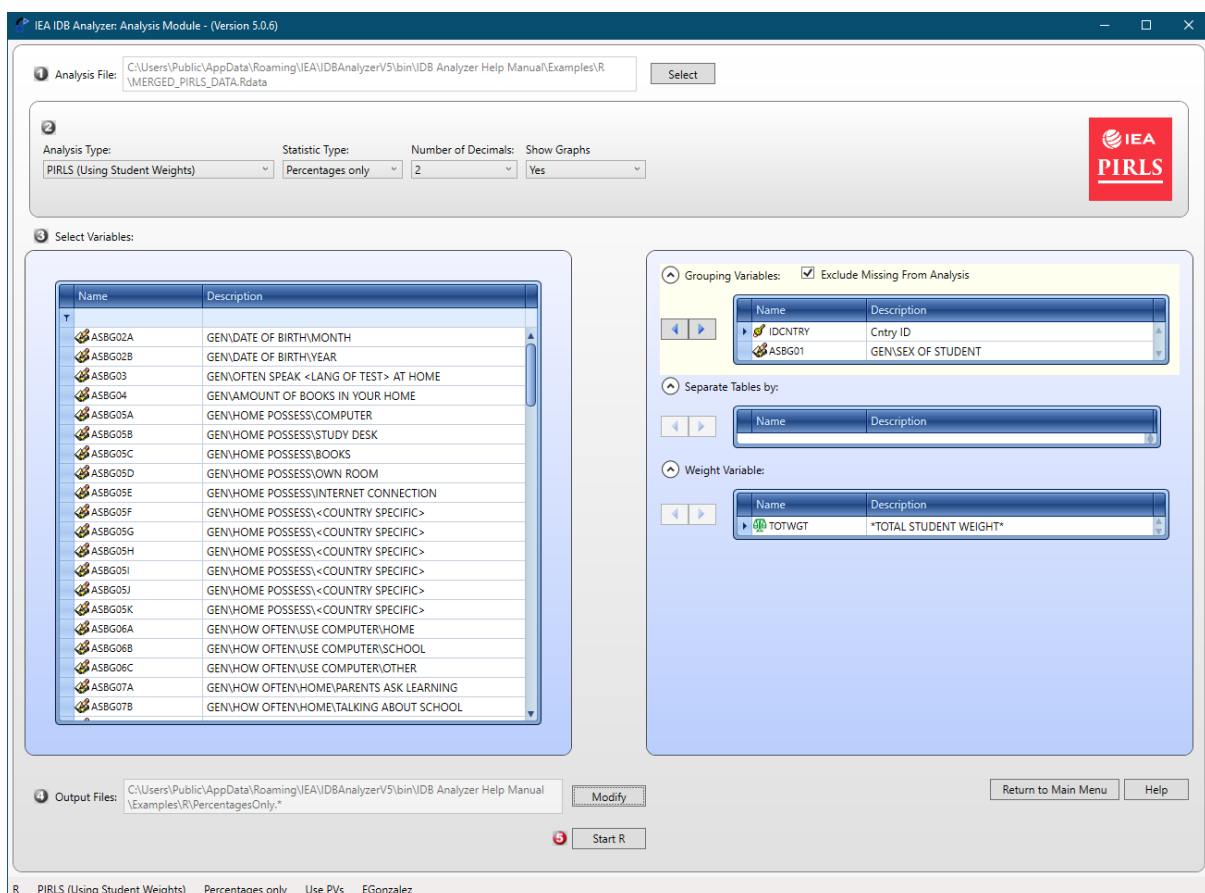
1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As type of the analysis select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and variance estimation variables, please refer to the technical documentation specific to the study.
4. From the **Statistic Type** dropdown menu, select **Percentages Only**. Leave the other dropdown menus unchanged.
5. If you want the IEA IDB Analyzer to create graphs leave the default option “Yes” under the option **Show Graphs**, otherwise change to “No”.¹⁹
6. In the next step you will need to define the grouping variables. As **Grouping Variable**, the software always selects variable IDCNTRY by default. You will need to add ASBG01 for this example. To do this, mark the variable from the variable list on the left side of the window and press the right arrow button belonging to the section of the grouping variable or double click on the variable in the list. This will move the variable ASBG01 from the variable list on the left side into the field for the grouping variables on the right. In this case, adding the variable ASBG01 as a grouping variable, or in the “Separate Tables by” has the same effect. If you want to create separate tables for each of several variables, add the variables to the “Separate Tables by” box. For example, using IDCNTRY as a grouping variable, and ASBG01 and another variable as “Separate Tables by” will create one set of results for IDCNTRY by ASBG01 and a separate set of results for IDCNTRY by the other variable.

¹⁹ Graphics are only available with SPSS.

7. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “PercentagesOnly”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis.
8. Click on the **Start SPSS/SAS/R Start SPSS** or **Start SAS** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 9 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 9: Analysis Module Setup for Computing Percentages only



There will be several output files created. The first one will have percentages for each of the subgroups created using the grouping variables. The other(s) will have results from the differences between the groups formed using the grouping variable(s). In our example, there will be two files. One contains the differences between boys and girls within each country, and the second one the differences between the countries for the boys and for the girls. These output files will have “_Sig” and “_Sig2” attached to their name, respectively. Note that the “_Sig2” file is only created when there are two or more grouping variables.

If a list of variables is provided to “Separate Tables by”, the name of these variables is added to the name of the file.

The columns in the output files are the following:

- Grouping Variables: Grouping variables used in defining the groups for the analysis
- N: Number of cases in group
- TOTWGT: Sum of the weights for cases in the group
- SUMW_SE: The standard error for the sum of the weights in the group.
- PCT: Percentage of cases in the group
- PCT_SE: Standard error of the percentage of cases in the group
- NVARSTRAT: when using the parameter STRATVAR, the number of variance strata contributing cases to the group. Currently only configured for IEA studies.
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- METHOD: The method of replication used for the analysis
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

The columns in the output file(s) with the comparisons (ending in “_Sig” and “_Sig2”) are the following:

- Grouping Variables: The grouping variables that define the groups in the analysis.
- GROUPVAR: The name of the variable that defines the groups that are being compared.
- REFGROUP: The label of the reference group.
- COMPGROUP: The label of the comparison group.
- PCT, PCT_SE, CPCT and CPCT_SE: The percent of people in the reference and the comparison group, respectively, and their corresponding standard errors.
- PCTDIFF: The difference between the percent of people in the comparison group and the reference group.
- PCTDIFF_SE: The standard error of the percent difference (PCTDIFF). This standard error is computed assuming independent samples when comparing countries, and assuming dependent samples otherwise.
- PCTDIFF_T: The t-statistics for the percent difference between the reference and comparison group. This is simply the difference divided by the corresponding standard error (PCTDIFF/PCTDIFF_SE)
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- METHOD: The method of replication used for the analysis
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: The name of the analysis file used in the analysis
- SELCRIT: the selection criteria, if any, used for the analysis.

Computing Percentages and Means of non-Plausible Values

To compute percentages and means of continuous variables not involving plausible values, you will need to select “**Percentages and Means**” from the **Statistic Type** dropdown menu.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables to be used to define the subgroups for analysis. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTY or its equivalent as the first grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
Separate Tables by	This is the list of classification variables used to generate each table with results. When omitted, it will generate tables using only the grouping variables.
Analysis Variables	This is the list of variables for which means are to be computed. The variable(s) to be selected should be numeric variables. You can select more than one analysis variable. If you want to compute means for plausible values, you will need to select “Use PVs” under the “Plausible Value Option” dropdown menu (please refer to the next section for more information).
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis based on the analysis type selected.

As an example, we will compute the mean age (ASDAGE) for boys and girls (ITSEX) and their standard errors within each country (IDCNTY), using the weighting variable TOTWGT. We will also compute the percentages of boys and girls and their standard errors within each country. The data will be read from the data file **Merged_PIRLS_Data** and the standard errors will be computed based on replicate weights.

The steps in the IEA IDB Analyzer are as follows:

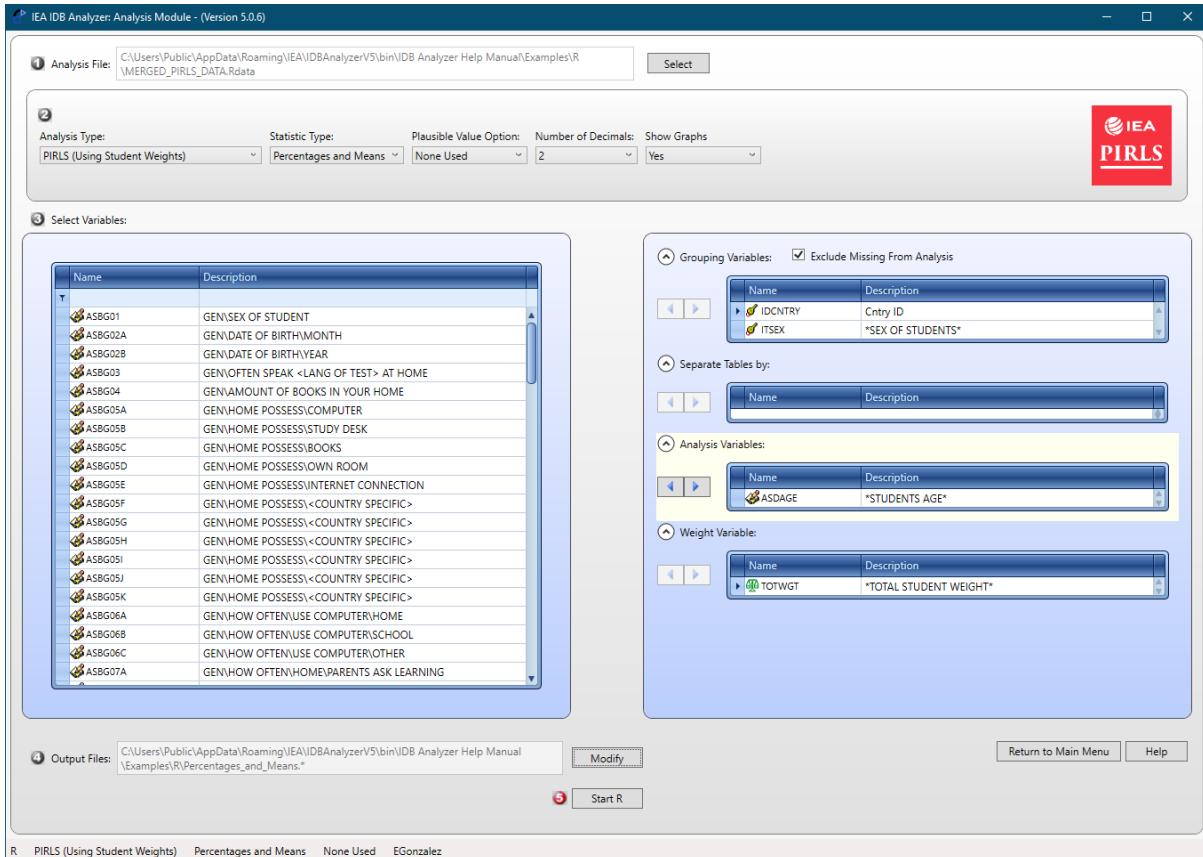
1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As **Analysis Type**, select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and jackknifing variables, please refer to the technical documentation specific to the study.
4. From the **Statistic Type** dropdown menu, select **Percentages and Means**. Leave the other dropdown menus unchanged as these are not relevant or available for this analysis.
5. Under the **Plausible Value Option** select **None Used**.
6. If you want the IEA IDB Analyzer to create graphs, leave the default option “Yes” under the option **Show Graphs**.
7. In the next steps the variables for the analysis need to be selected:
 - As **Grouping Variables**, the software always selects variable IDCNTY by default. You will need to add ITSEX for this example. To do this, select the variable from the variable list on the left-hand side of the window and press the right arrow button belonging to the section of the grouping variable, or just double click on the variable name. This will move the variable ITSEX from the variable list on the left side into the field for the grouping variables on the right. Make sure the corresponding section is activated. In this case,

adding the variable ITSEX as a grouping variable, or in the “Separate Tables by” has the same effect. If you want to create separate tables for each of several variables, add the variables to the “Separate Tables by” box. For example, using IDNCTRY as a grouping variable, and ITSEX and another variable as “Separate Tables by” will create one set of results for IDCNTRY by ITSEX and a separate set of results for IDCNTRY by the other variable.

- Next the analysis variables need to be selected. To activate this section, you will need to click into the area around the **Analysis Variables** field. This time you will need to select the variable ASDAGE from the list of variables and move it to the analysis variables field by pressing the right arrow button in this section. Note that you can select more than one analysis variable for your analysis. The output will contain separate tables with statistics for each one of them.
8. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “Percentages_and_Means”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis.
 9. Click on the **Start SPSS/SAS/R** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 10 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 10: Analysis Module Setup for Computing Percentages and Means of non-Plausible Values



There will be several output files created. The first one will have percentages and means for each of the subgroups created using the grouping variables. The other(s) will have results from the differences between the groups formed using the grouping variable(s). In our example, there will be two files. One contains the differences between boys and girls within each country, and the second one the differences between the countries for the boys and for the girls. These output files will have “_Sig” and “_Sig2” attached to their name, respectively. Note that the “_Sig2” file is only created when there are two or more grouping variables specified in the analysis. If a list of variables is provided to “Separate Tables by”, the name of these variables is added to the name of the file.

The columns in the output files with the percentages and means are the following:

- Grouping Variables: Grouping variables used in defining the groups for the analysis
- XVAR: Analysis Variable used in the analysis
- N: Number of cases in group
- TOTWGT: Sum of the weights for cases in the groups defined by the Grouping Variables (excludes cases with missing values for the analysis variable).
- SUMW_SE: Standard error of the sum of the weights
- TSUMW: Sum of the weights for cases in the groups defined by the Grouping Variables (includes all cases in the analysis file).
- TSUMW_SE: Standard error for TSUMW.
- PCT: Percentage of cases in the group (excludes cases with missing values for the analysis variable)
- PCT_SE: Standard error of the percentage of cases in the group
- TPCT: Percentage of cases in the group (including all cases in the analysis file)
- TPCT_SE: Standard error of TPCT.
- MNX: Average of the analysis variable
- MNX_SE: Standard error of the mean analysis variable
- SDX: Standard deviation of the analysis variable
- SDX_SE: Standard error of the mean analysis variable
- VRX: variance of the analysis variable
- VRX_SE: standard error of the variance of the analysis variable
- DEff: the design effect²⁰
- PCTMISS: Percent of cases within the group with missing analysis variable
- NVARSTRAT: when using the parameter STRATVAR, the number of variance strata contributing cases to the group. Currently only configured for IEA studies.
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- METHOD: The method of replication used for the analysis
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: the name of the analysis file used in the analysis
- SELCRIT: the selection criteria, if any, used for the analysis.

²⁰ Please refer to the Appendix of this Help Manual for information about calculating the design effects.

The columns in the file(s) with the percent and mean comparisons (ending in “_Sig” and “_Sig2”) are the following:

- Grouping Variables: The grouping variables that define the groups in the analysis.
- DVAR: The name of the analysis variable that is used for the comparison.
- GROUPVAR: The name of the variable that defines the groups that are being compared.
- REFGROUP: The label of the reference group.
- COMPGROUP: The label of the comparison group.
- PCT, PCT_SE, CPCT and CPCT_SE: The percent of people in the reference and the comparison group, respectively, and their corresponding standard errors.
- PCTDIFF: The difference between the percent of people in the comparison group and the reference group.
- PCTDIFF_SE: The standard error of the percent difference (PCTDIFF). This standard error is computed assuming independent samples when comparing countries, and assuming dependent samples otherwise.
- PCTDIFF_T: The t-statistics for the percent difference between the reference and comparison group. This is simply the difference divided by the corresponding standard error (PCTDIFF/PCTDIFF_SE)
- MNX, MNX_SE, CMNX and CMNX_SE: The mean ASDAGE of the reference and the comparison group, respectively, and their corresponding standard errors.
- MNXDIFF: The difference between the comparison group and the reference group on the analysis variable.
- MNXDIFF_SE: The standard error of the mean difference (MNXDIFF). This standard error is computed assuming independent samples when comparing countries, and assuming dependent samples otherwise.
- MNXDIFF_T: The t-statistics for the mean difference between the reference and comparison group. This is simply the difference divided by the corresponding standard error (MNXDIFF/MNXDIFF_SE)
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- METHOD: The method of replication used for the analysis
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: the name of the analysis file used in the analysis
- SELCRIT: the selection criteria, if any, used for the analysis.

Computing Percentages and Means of Plausible Values

To compute percentages and means of plausible values you will need to select “**Percentages and Means**” from the **Statistic Type** dropdown menu, and under **Plausible Value Options** select “**Use PVs**”.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables to be used to define the subgroups for analysis. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTY or its equivalent as the first grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
Separate Variables by	This is the list of classification variables used to generate each table with results. When omitted, it will generate tables using only the grouping variables.
Plausible Values	This section is used to identify the set of plausible values for the analysis.
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

As an example, we will compute the mean reading achievement (using plausible values ASRREA01-5) for boys and girls (ITSEX) within each country (IDCNTY) and their standard errors, using the weighting variable TOTWGT. The program uses all plausible values to compute these statistics. It will also compute the percentages of boys and girls within each country, and their standard errors. The data will be read from the data file **Merged_PIRLS_Data** and the standard errors will be computed based on replicate weights.

The steps in the IEA IDB Analyzer are as follows:

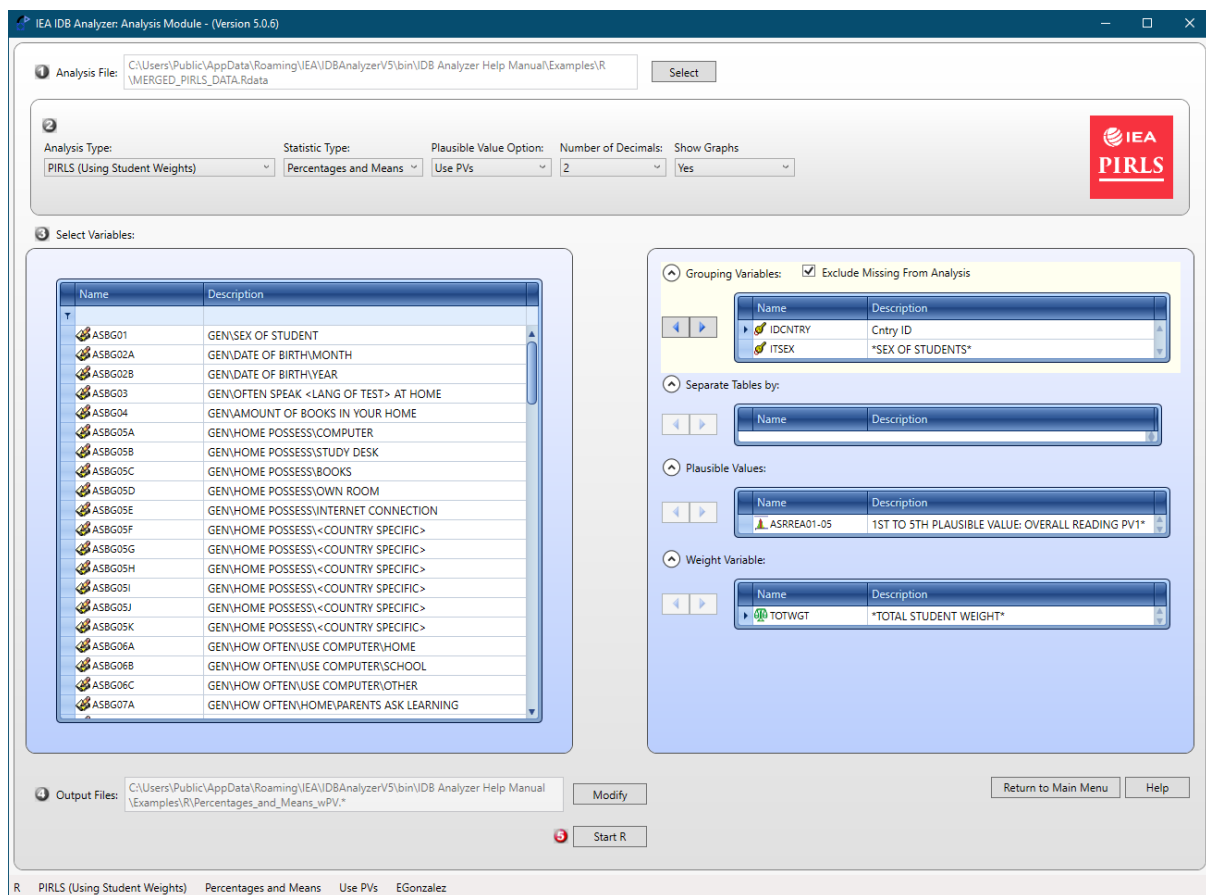
1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As Analysis Type, choose **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and jackknifing variables, please refer to the technical documentation specific to the study.
4. From the **Statistic Type** dropdown menu, select **Percentages and Means**.
5. From the **Plausible Value Option** dropdown menu, select **Use PVs**. Leave the other dropdown menus unchanged.
6. If you want the IEA IDB Analyzer to create graphs leave the default option “Yes” under the option **Show Graphs**, otherwise select “No”.
7. In the next steps the variables for the analysis need to be specified:
 - For Grouping Variables, the software always selects variable IDCNTY by default. You will need to add ITSEX for this example. To do this, select the variable from the list on the left side of the window and press the right arrow button or double click on the variable. This will move the variable ITSEX from the variable list on the left side into the field for the grouping variables on the right. You can also drag the variable from one panel to the other. In this case, adding the variable ITSEX as a grouping variable, or in the “Separate Tables by” has the same effect. If you want to create separate tables for each of several variables, add the variables to the “Separate Tables by” box. For example, using IDCNTY as a grouping variable, and ITSEX and another variable as “Separate Tables by” will create one

set of results for IDCNTRY by ITSEX and a separate set of results for IDCNTRY by the other variable.

- Next select the plausible values. To activate this section, you will need to click into the area of the **Plausible Values** field. Now you will need to select variable ASRREA01-05 from the list of variables and move it to the analysis variables field by pressing the right arrow button in this section, or just double click on the variable name. You are able to select more than one set of plausible values for the analysis. Each set will be analyzed sequentially.
8. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “Percentages_and_Means_wPV”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis.
 9. Click on the **Start SPSS/SAS/R** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 11 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 11: Analysis Module Setup for Computing Percentages and Means of Plausible Values



There will be several output files created. The first one will have percentages and means for each of the subgroups created using the grouping variables. The other(s) will have results from the differences between the groups formed using the grouping variable(s). In our example, there will be two files. One

contains the differences between boys and girls within each country, and the second one the differences between the countries for the boys and for the girls. These output files will have “_Sig” and “_Sig2” attached to their name, respectively. Note that the “_Sig2” file is only created when there are two or more grouping variables specified in the analysis. If a list of variables is provided to “Separate Tables by”, the name of these variables is added to the name of the file.

The columns in the output files with the percentages and means are the following:

- Grouping Variables: Grouping variables used in defining the groups for the analysis.
- DVAR: The name of the plausible value used in the analysis
- N: Number of cases in group
- TOTWGT: Sum of the weights for cases in the groups in the groups defined by the Grouping Variables
- SUMW_SE: Standard error of the sum of the weights
- PCT: Percentage of cases in the group
- PCT_SE: Standard error of the percentage of cases in the group
- MNPV: Average of the plausible values
- MNPV_SE: Standard error of the mean of the plausible values
- MNPV_CI: The 95% confidence interval of the average of the plausible values
- SDPV: Standard deviation of the plausible values
- SDPV_SE: Standard error of the standard deviation of the plausible values.
- VRPV: Variance of the plausible values
- VRPV_SE: Standard error of the variance of the plausible values
- DEff1, DEff2, DEff3, DEff4, DEff5²¹: The design effects
- PCTMISS: Percent missing the plausible values within the group
- NVARSTRAT: when using the parameter STRATVAR, the number of variance strata contributing cases to the group. Currently only configured for IEA studies.
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- METHOD: The method of replication used for the analysis
- NPV: The number of plausible values used for the analysis
- SHORTCUT: Whether only one plausible value was used to calculate the sampling error (Y) or all of them were used (N)
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: File used for the analysis
- SELCRIT: Selection criteria used in the analysis, if any.

The columns in the output file(s) with the mean comparisons (ending in “_Sig” and “_Sig2”) are the following:

- Grouping Variables: Grouping variables used in defining the groups for the analysis.
- PVVAR: The name of the plausible value that is used for the comparison.

²¹ Please refer to the Appendix of this Help Manual for information about calculating design effects.

- GROUPVAR: The name of the variable that defines the groups that are being compared.
- REFGROUP: The label of the reference group.
- COMPGROUP: The label of the comparison group.
- PCT, PCT_SE, CPCT and CPCT_SE: The percent of people in the reference and the comparison group, respectively, and their corresponding standard errors.
- PCTDIFF: The difference between the percent of people in the comparison group and the reference group.
- PCTDIFF_SE: The standard error of the percent difference (PCTDIFF). This standard error is computed assuming independent samples when comparing countries, and assuming dependent samples otherwise.
- PCTDIFF_T: The t-statistics for the percent difference between the reference and comparison group. This is simply the difference divided by the corresponding standard error (PCTDIFF/PCTDIFF_SE)
- MNPV, MNPV_SE, CMPV and CMNPV_SE: The mean plausible value of the reference and the comparison group, respectively, and their corresponding standard errors.
- MNPVDIFF: The difference between the comparison group and the reference group on the analysis variable.
- MNPVDIFF_SE: The standard error of the mean difference (MNPVDIFF). This standard error is computed assuming independent samples when comparing countries, and assuming dependent samples otherwise.
- MNPVDIFF_T: The t-statistics for the mean difference between the reference and comparison group. This is simply the difference divided by the corresponding standard error (MNPVDIFF/MNPVDIFF_SE)
- MNPVDIFF_CI: The confidence interval around the difference between the comparison group and the reference group on the analysis variable.
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- METHOD: The method of replication used for the analysis
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: The name of the analysis file used in the analysis
- SELCRIT: the selection criteria, if any, used for the analysis.

Computing Percentages of the Population Meeting User-Specified Benchmarks²²

To compute percentages of the population meeting or exceeding benchmarks or cut points in the distribution of achievement, you need to select “**Benchmark**” from the **Statistic Type** dropdown menu. In addition to computing percentages, you also have the option to compute the average of one of more analysis variables by achievement group.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables that will be used to define the subgroups. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTRY or its equivalent as the first grouping variable and there should always be at least one grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
Plausible Values	The achievement score or set of plausible values that will be used for the analysis.
Benchmarks	The cut points of the distribution. See the Appendix of this Help Manual for a listing of the cut points used in international large-scale assessment studies supported by the IEA IDB Analyzer. The cut points entered will be used as the lower bound values for the achievement grouping.
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

As examples, we will conduct 3 separate analysis: a) compute the percent of boys and girls within the population meeting each of the PIRLS benchmarks, b) compute the percent of boys and girls within the population within each of the achievement levels as defined by the benchmarks, and c) compute the percent of boys and girls within the population within each of the achievement levels as defined by the benchmarks and their average on the “Motivated to read” (ASBGSMR) scale. The data will be read from the data file **Merged_PIRLS_Data**.

The steps in the IEA IDB Analyzer are as follows:

1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As type of the analysis, select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally, the corresponding variance estimation variables are preselected by the software. For the correct weight and variance estimation variables, please refer to the technical documentation specific to the study.
4. From the **Statistic Type** dropdown menu, select **Benchmarks**. Leave the other dropdown menus unchanged.

Please note that there are additional parameters for the **Benchmarks** statistics type which you can choose from the **Benchmarks Options** dropdown menu:

²² “Benchmarks” are points along the distribution of proficiency that have a defined significance when describing what respondents know and can do. Depending on the study or survey, different terms might be used, such as “achievement levels”, “proficiency levels”, etc. Appendix B of this Help Manual lists the corresponding cut points used in the major international studies.

- a) **Cumulative:** Compute the percent of people “at or above” the achievement benchmarks by selecting this option.
 - b) **Discrete:** Compute the percentage of people within each of the groups in the distribution as defined by the cut points provided. The cut points are treated as the lower bound value for the categories.
 - c) **Discrete with Analysis Variable(s):** In addition to the percent within achievement groups, this option also computes the mean value of an analysis variable (or variables) for the groups as defined by the achievement benchmarks. For example, this option allows you to compute the average motivation to read per achievement group within each of the achievement groups.
5. If you want the IEA IDB Analyzer to create graphs leave the default option “Yes” under the option **Show Graphs**²³ otherwise change to “No”.
 6. In the next step you will need to define the grouping variables. As **Grouping Variable**, the software always selects variable IDCNTY by default, or an equivalent variable depending on the study. In addition, select ITSEX as a grouping variable.
 7. If you selected the option **Discrete with Analysis Variable(s)** you then select the analysis variable(s) for your analysis.
 8. Select the **Plausible Value** that will be used in this analysis. In our example we will use the overall reading plausible values (ASRREA01 – ASRREA05). When choosing the plausible values, you will have the option to include in your analysis cases in your file that have no plausible values. These will be presented in a category labeled “Not Classified” and these cases will be used in the denominator when calculating percentages. If all cases in your file have plausible values, selecting this option will have no effect in your analysis and results.²⁴ Leaving this box unchecked will conduct the analysis using only those cases that have plausible values.
 9. Specify the cut points in the distribution. For our example, we will use the PIRLS international benchmarks of achievement: 400, 475, 550, and 625. These numbers need to be typed in increasing order separated by spaces. Click on the **Achievement Benchmarks** field to add them. You will have the choice to select the ones provided by the interface or edit them.
 10. The weight and variance estimation variables are automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and variance estimation variables, please refer to the technical documentation specific to the study.
 11. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our examples we will use the names “CumulativeBenchmarks” (a), “DiscreteBenchmarks” (b) and “DiscretewAnalysisBenchmarks” (c). These filenames will be used to create the syntax files to perform the analysis, and the corresponding output files with the statistics from the analysis.
 12. Click on the **Start SPSS/SAS/R Start SPSS** or **Start SAS** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

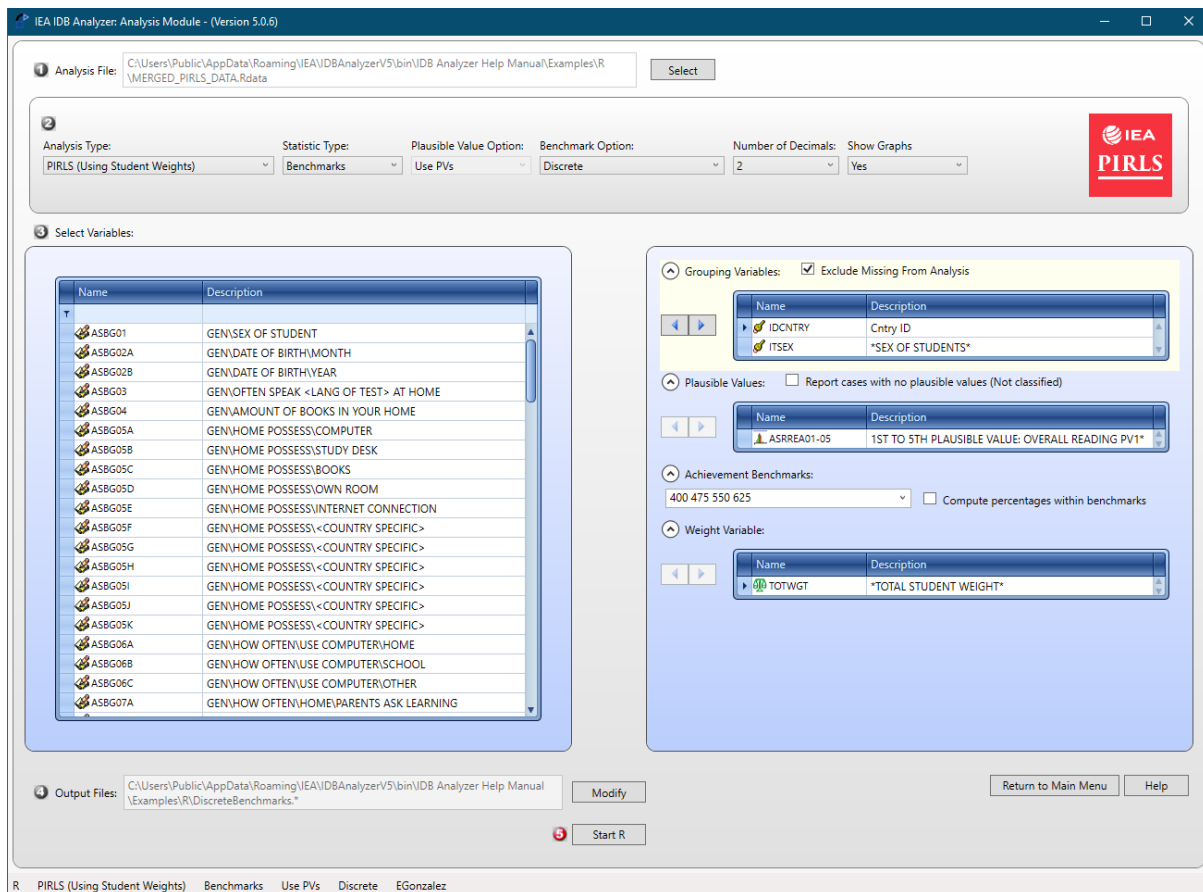
As with any analysis involving plausible values, when computing benchmarks of achievement, the analyses are conducted separately with each plausible value. The results presented are the average of the results obtained with each plausible value, as described in the technical documentation for the corresponding study.

²³ Graphs are only available when the “Discrete” or “Discrete with Analysis Variable(s)” benchmark options are chosen.

²⁴ This option was originally developed for use with the Problem-Solving scale in PIAAC, but is available for any database that would not have assigned plausible values to all participants.

Figure 12 shows the IEA IDB Analyzer Setup Screen for computing benchmarks, discrete option (example b). Corresponding syntax and output files with the results from the three sample analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 12: Analysis Module Setup for Computing Percentages with Benchmarks (Discrete)



The columns in the output files when computing discrete or cumulative benchmarks are the following:

- Grouping Variables: Grouping variables used in defining the groups for the analysis (IDCNTY and ITSEX in this case)
- PVVAR: The plausible value used for the analysis
- CUTVAR: The achievement group as defined by the benchmarks used in the analysis
- BNCHMRKS: The values used as cut points of the distribution
- N: Number of cases in group (note that since you are using plausible values as achievement scores, this is the average number of cases obtained using each of the plausible values used in the analysis, and therefore will not necessarily be an integer)
- TOTWGT: Sum of the weights for cases in the group (as with the number of cases, this is also averaged across the 5 computations using each of the plausible values)
- SUMW_SE: The standard error for the sum of the weights in the group.
- PCT: Percentage of cases in the group within the categories of the last grouping variable
- PCT_SE: Standard error of the percentage of cases in the group
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis

- SHORTCUT: Whether the sampling variance was calculated using all plausible values (N), or just the first plausible value (Y).
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- NPV: Number of plausible values used in the analysis.
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

When using an analysis variable, you get the following columns:

- Grouping Variables: Grouping variables used in defining the groups for the analysis (IDCNTRY and ITSEX in this case)
- PVVAR: The plausible value used for the analysis
- DVAR: The name of the analysis variable used in the analysis
- CUTVAR: The achievement group as defined by the benchmarks used in the analysis
- BNCHMRKS: The values used as cut points of the distribution
- N: Number of cases in group that are NOT missing the analysis variable
- TOTWGT: Sum of the weights for cases in the group that are NOT missing the analysis variable
- SUMW_SE: the standard error for TOTWGT.
- TSUMW: Sum of the weights for cases in the group regardless of missing the analysis variable.
- TSUMW_SE: The standard error for TSUMW.
- PCT: Percentage of cases in the group within the categories of the last grouping variable, using only cases that are NOT missing the analysis variable.
- PCT_SE: Standard error of PCT.
- TPCT: Percentage of cases in the group within the categories of the last grouping variable, using all cases in the file.
- TPCT_SE: Standard error of TPCT.
- MNX: Average of the analysis variable within the group
- MNX_SE: Standard error of MNX.
- SDX: Standard deviation of the analysis variable within the group
- SDX_SE: Standard error of SDX.
- PCTMISS: Weighted percent of cases missing the analysis variable within the group.
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- SHORTCUT: Whether the sampling variance was calculated using all plausible values (N), or just the first plausible value (Y).
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- NPV: Number of plausible values used in the analysis.

- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

When calculating discrete percentages of people within achievement benchmarks, with or without an analysis variable, you get an additional output file with comparisons between the percentages of people within the benchmark groups across all the subgroups. If you select to work with an analysis variable, you also get the difference between the mean for the analysis variable between the groups, within achievement level.

The columns in the output file(s) with the percent and mean comparisons (ending in “_Sig”) are the following:

- Grouping Variables: All but the last grouping variable will be listed. In our example, you will see IDCNTY as the comparisons will be made between groups formed by the variable ITSEX.
- CUTVAR: The achievement group as defined by the benchmarks used in the analysis.
- PVVAR: The plausible value used for the analysis
- GROUPVAR: The name of the variable that defines the groups that are being compared.
- REFGROUP: The label of the reference group.
- COMPGROUP: The label of the comparison group.
- PCT, PCT_SE, CPCT and CPCT_SE: The percent of people in the reference and the comparison group, respectively, and their corresponding standard errors.
- PCTDIFF: The difference in percent of people between the comparison group and the reference group.
- PCTDIFF_SE: The standard error of the difference between the percent in the comparison group and the reference group. This standard error is computed assuming independent samples when comparing countries, and assuming dependent samples otherwise.
- PCTDIFF_T: The t-statistics for the mean difference between the percent of people in the reference and comparison group. This is simply the difference divided by the corresponding standard error (PCTDIFF/PCTDIFF_SE)
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- METHOD: The method of replication used for the analysis
- NPV: The number of plausible values used for the analysis
- SHORTCUT: Whether only one plausible value was used to calculate the sampling error (Y) or all of them were used (N)
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: File used for the analysis
- SELCRIT: Selection criteria used in the analysis, if any.

When your analysis involves one or more analysis variables, you will see one “_Sig” file for each analysis variable, with the following additional columns:

- DVAR: The name of the analysis variable that is used for the comparison.
- MNX, MNX_SE, CMNX and CMNX_SE: The mean on the analysis variable for the reference and the comparison group, respectively, and their corresponding standard errors.
- MNXDIFF: The difference in average analysis variable between the comparison group and the reference group.

- MNXDIFF_SE: The standard error of the difference in average analysis variable between the comparison group and the reference group. This standard error is computed assuming independent samples when comparing countries, and assuming dependent samples otherwise.
- MNXDIFF_T: The t-statistics for the mean difference in average analysis variable between the reference and comparison group. This is simply the difference divided by the corresponding standard error (MNXDIFF/MNXDIFF_SE).

Computing Percentages within Benchmarks

By default, the statistic type “Benchmarks” computes the percent of people in a group that are at each of the achievement levels or benchmarks. In our examples, we computed the percent of boys and girls that are in each of the 5 groups defined by the achievement benchmarks. These percentages add up to 100% for each group.

However, a research question could be posed as “what percent of those students at each of the achievement levels are boys, and what percent of them are girls?” This can be answered by checking the box “Compute percentages within benchmarks”. When doing so, the program will compute the percent of students within each of the groups defined by the benchmarks that are boys and that are girls. It uses the last variable in the list of grouping variables as the variable for which the percentages are computed.

Computing Percentiles

To compute percentiles for a variable, you will need to select “**Percentiles**” from the **Statistic Type** dropdown menu.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables that are to be used to define the subgroups. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTRY or its equivalent as the first grouping variable and there should always be at least one grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
Analysis Variables	The variable(s) for which the percentiles will be calculated. You can select one or more analysis variables for this analysis.
Percentiles	These are the percentiles that will be calculated from the distribution of values for the variables. These need to be sorted in increasing order separated by spaces and written with no decimals.
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

In this example, we will compute the 25th, 50th, and 75th percentiles for the scales “Students Like Reading” (ASBGSLR) and “Students Motivated to Read” (ASBGSMR).

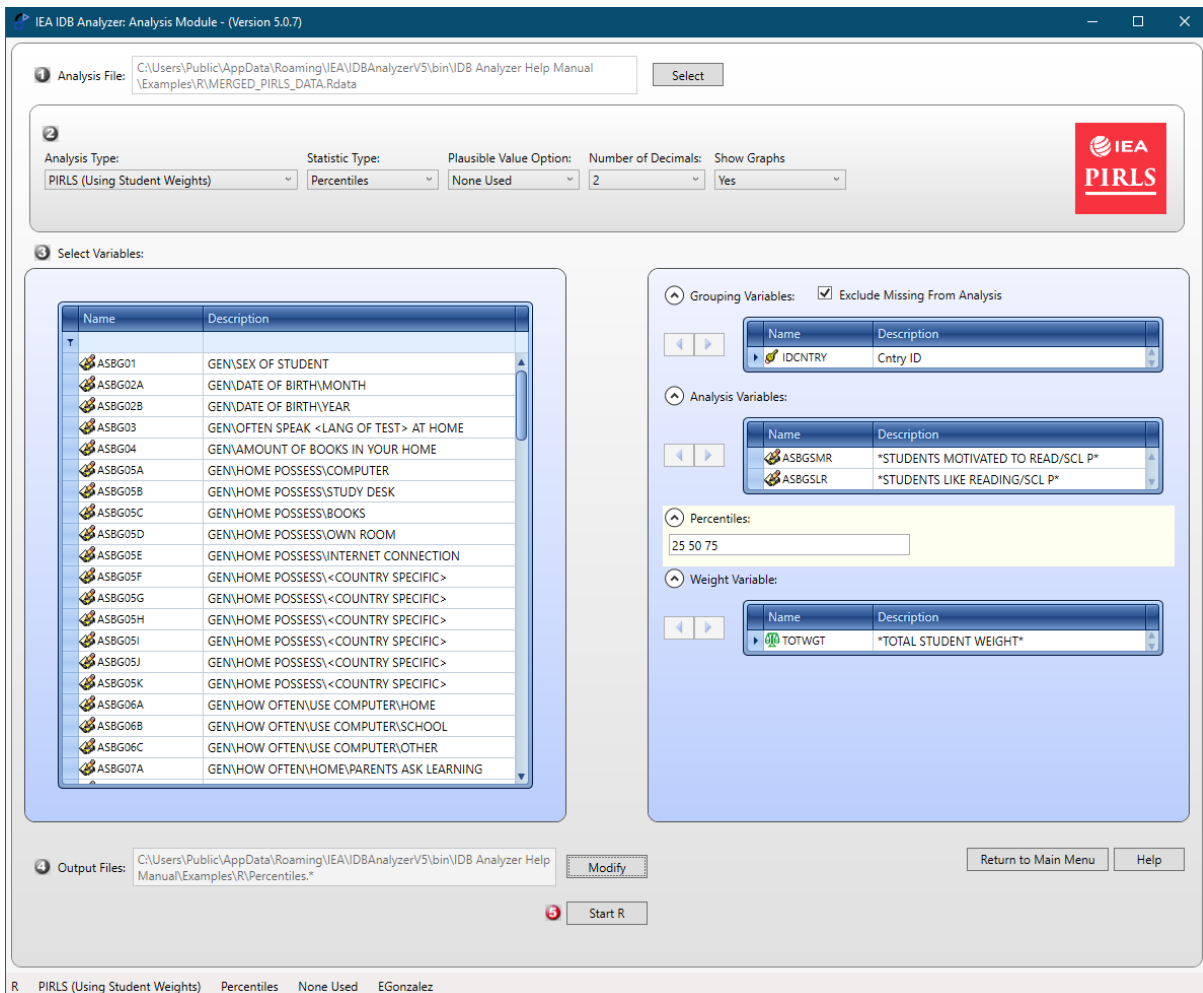
The steps in the IEA IDB Analyzer are as follows:

1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As Analysis Type, select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and variance estimation variables, please refer to the technical documentation specific to the study.
4. From the **Statistic Type** dropdown menu, select **Percentiles**. Leave the other dropdown menus unchanged.
5. In the next steps all parameters for the analysis need to be defined:
 - As Grouping Variable, the software always selects variable IDCNTRY by default. For this analysis, no other grouping variables will be used.
 - Click on the **Analysis Variables** field to activate it, choose ASBGSLR and ASBGSMR from the list of available variables on the left side, and select it as analysis variable using the right arrow button.
 - Specify the percentile points in the distribution. For our example, we will compute the 25th, 50th, and 75th percentiles. The numbers 25, 50 and 75 need to be typed in increasing order separated by spaces (i.e., 25 50 75). Click on the **Percentiles** field to add them.
6. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “Percentiles”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis.

- Click on the **Start SPSS/SAS/R Start SPSS** or **Start SAS** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 13 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 13: Analysis Module Setup for Computing Percentiles



The columns in the output files are the following:

- **Grouping Variables:** Grouping variables used in defining the groups for the analysis (only IDCNTRY in this case)
- **DVAR:** Analysis Variable used in the analysis
- **N:** Number of cases in group
- **TOTWGT:** Sum of the weights for cases in the group
- **P25, P50 and P75:** The percentiles computed
- **P25_SE, P50_SE and P75_SE:** The standard error of the computed percentiles
- **WEIGHT:** The weighting variable used for the analysis
- **REPS:** The number of replicates used for the analysis

- METHOD: The method of replication used for the analysis
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

When calculating percentiles, you get additional output files with comparisons of the percentiles between all the subgroups. In the example above, you would get the comparison of each of the percentiles requested between each country pairing. The columns in the output file(s) with the comparisons of the percentiles (ending in “_Sig”) are the following:

- Grouping Variables: Grouping variables used in defining the groups for the analysis.
- Percentile: The percentile being compared.
- DVAR: the analysis variable used to calculate the percentiles
- GROUPVAR: The name of the variable that defines the groups that are being compared.
- REFGROUP: The label of the reference group.
- COMPGROUP: The label of the comparison group.
- VAL, VAL_SE, CVAL, CVAL_SE: the percentiles for the reference and comparison groups, respectively, with their corresponding standard errors.
- VALDIF: the difference in the percentile values between the groups.
- VALDIFF_SE: The standard error of the difference between the percentiles. This standard error is computed assuming independent samples when comparing countries, and assuming dependent samples otherwise.
- VALDIF_T: The t-statistics for the mean difference between the percentile in the reference and comparison group. This is simply the difference divided by the corresponding standard error (VALDIFF/VALDIFF_SE)
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- METHOD: The method of replication used for the analysis
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: File used for the analysis
- SELCRIT: Selection criteria used in the analysis, if any.

In addition, the user can request a GAP analysis using the option DOGAP = Y²⁵. When this is requested, you get two additional outputs. In the first one, with the file name ending in _GAP, the program compares the percentiles within each group (e.g., the gap between the 25th and the 50th percentile, the gap between the 25th vs. the 75th percentile, etc.). This shows whether differences between percentiles are significantly different from zero. The second output, with the filename ending in _GAP2, shows the comparisons of the differences between the percentiles between the groups (e.g., compares the difference/gap between the 25th and the 75th percentile in group A vs. the difference/gap in group B). In both cases it tests the size of the gap, and the columns in the output are labelled accordingly.

²⁵ GAP Analysis is currently only available for SAS and SPSS.

Computing Percentiles of Plausible Values

To compute percentiles of Plausible Values you will need to select “**Percentiles**” from the **Statistic Type** dropdown menu, and under **Plausible Value Options** select “**Use PVs**”.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables that are to be used to define the subgroups. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTRY or its equivalent as the first grouping variable and there should always be at least one grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
Plausible Values	The set of plausible values that will be used for the analysis. You can select several plausible values for the analysis. The percentiles will be computed for each of them separately.
Percentiles	These are the percentiles that will be calculated from the distribution. These need to be sorted in increasing order and written with no decimals.
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

As an example, we will compute the 25th, 50th and 75th percentiles of student achievement scores and their standard errors for each country by gender. The data will be read from the data file **Merged_PIRLS_Data** and the standard errors will be computed based on replicate weights.

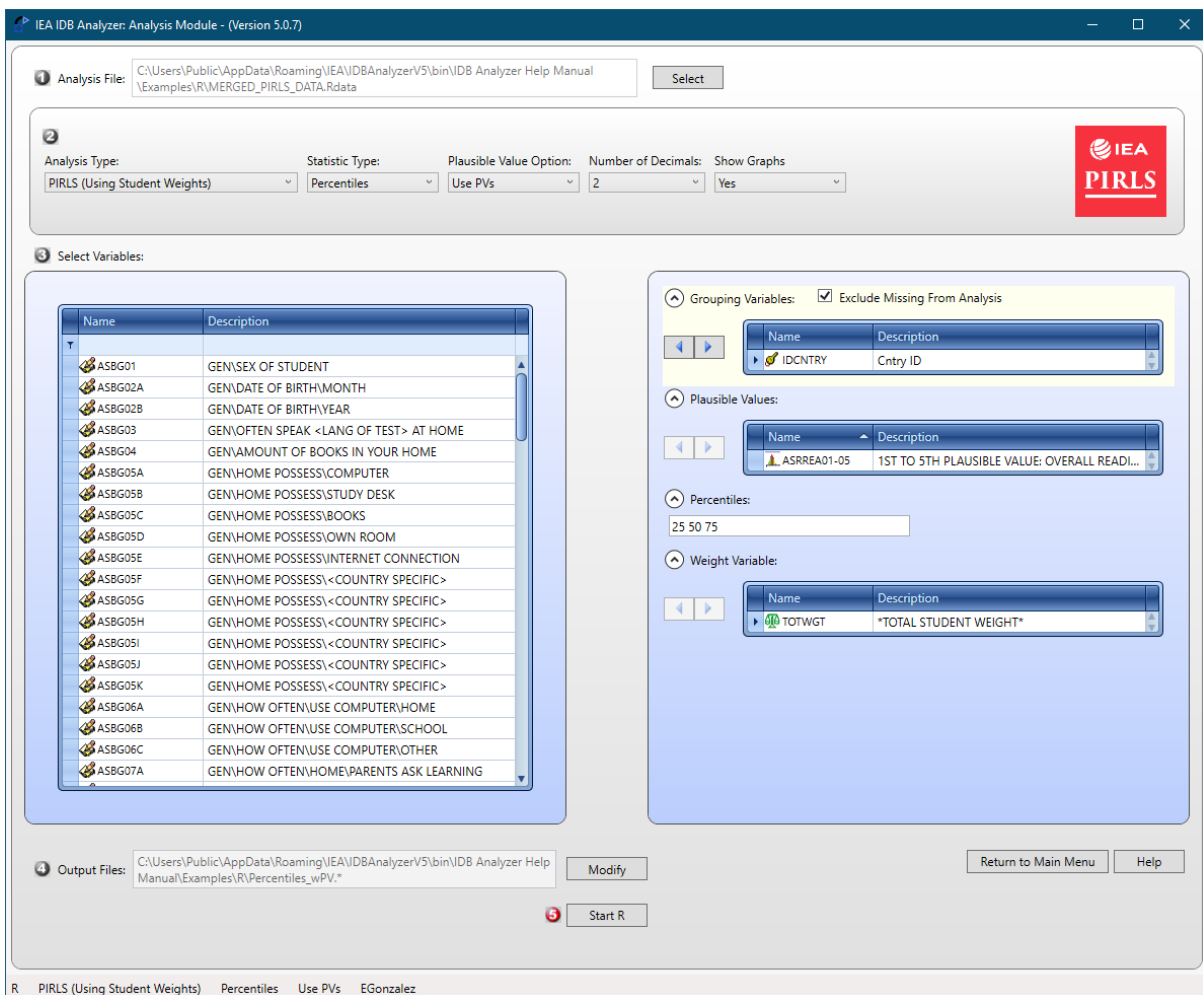
The steps in the IEA IDB Analyzer are as follows:

1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As Analysis Type, select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and variance estimation variables, please refer to the technical documentation specific to the study.
4. From the **Statistic Type** dropdown menu, select **Percentiles**. From the **Plausible Values Option** dropdown menu, choose **Use PVs**. Leave the other dropdown menus unchanged.
5. In the next steps, all parameters for the analysis need to be defined:
 - As **Grouping Variable**, the software always selects variable IDCNTRY or its equivalent by default. Add the variable ITSEX as second grouping variable to obtain the percentiles of the plausible values by gender.
 - Click on the **Plausible Values** field, select the variable ASRREA01-05 (1st to 5th Overall Reading Plausible Values) and use the arrow button to place it as the plausible value variable or by double clicking on the variable name. While you can select more than one set of plausible values for the analysis, only one will be used for this example.
 - Specify the percentile points in the distribution clicking on the **Percentiles** field. For our example, we will compute the 25th, 50th and 75th percentiles. The numbers 25, 50 and 75 need to be typed in increasing order separated by spaces (25 50 75).

6. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “Percentiles_wPV”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis.
7. Click on the **Start SPSS/SAS/R Start SPSS** or **Start SAS** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 14 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 14: Analysis Module Setup for Computing Percentiles using Plausible Values



The columns in the output files Excel file and SPSS datasets are the following:

- Grouping Variables: Grouping variables used in defining the groups for the analysis (only IDCNTRY in this case)
- DVAR: Analysis Variable used in the analysis
- N: Number of cases in group
- TOTWGT: Sum of the weights for cases in the group
- P25, P50 and P75: The percentiles computed

- P25_SE, P50_SE and P75_SE: The standard error of the computed percentiles
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- METHOD: The method of replication used for the analysis
- NPV: The number of variables used in the analysis.
- SHORTCUT: Whether the sampling variance was calculated using all plausible values (N), or just the first plausible value (Y).
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

When calculating percentiles, you get additional output files with comparisons of the percentiles between all the subgroups. In the example above, you would get the comparison of each of the percentiles requested between each country pairing. The columns in the output file(s) with the comparisons of the percentiles (ending in “_Sig”) are the following:

- Grouping Variables: Grouping variables used in defining the groups for the analysis.
- Percentile: The percentile being compared.
- DVAR: the analysis variable used to calculate the percentiles
- GROUPVAR: The name of the variable that defines the groups that are being compared.
- REFGROUP: The label of the reference group.
- COMPGROUP: The label of the comparison group.
- VAL, VAL_SE, CVAL, CVAL_SE: the percentiles for the reference and comparison groups, respectively, with their corresponding standard errors.
- VALDIF: the difference in the percentile values between the groups.
- VALDIFF_SE: The standard error of the difference between the percentiles. This standard error is computed assuming independent samples when comparing countries, and assuming dependent samples otherwise.
- VALDIF_T: The t-statistics for the mean difference between the percentile in the reference and comparison group. This is simply the difference divided by the corresponding standard error (VALDIFF/VALDIFF_SE)
- WEIGHT: The weighting variable used for the analysis
- REPS: The number of replicates used for the analysis
- METHOD: The method of replication used for the analysis
- NPV: The number of plausible values used for the analysis
- SHORTCUT: Whether only one plausible value was used to calculate the sampling error (Y) or all of them were used (N)
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- INFILE: File used for the analysis
- SELCRIT: Selection criteria used in the analysis, if any.

In addition, the user can request a GAP analysis using the option `DOGAP = Y`²⁶. When this is requested, you get two additional outputs. In the first one, with the file name ending in `_GAP`, the program compares the percentiles within each group (e.g., the gap between the 25th and the 50th percentile, the gap between the 25th vs. the 75th percentile, etc.). This shows whether differences between percentiles are significantly different from zero. The second output, with the filename ending in `_GAP2`, shows the comparisons of the differences between the percentiles between the groups (e.g., compares the difference/gap between the 25th and the 75th percentile in group A vs. the difference/gap in group B). In both cases it tests the size of the gap, and the columns in the output are labelled accordingly.

²⁶ GAP Analysis is currently only available for SAS and SPSS.

Computing Pearson Correlation Coefficients

To compute Pearson correlation coefficients, not using plausible values, you need to select “**Correlations (Pearson)**” from the **Statistic Type** dropdown menu.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables that are to be used to define the subgroups. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTY or its equivalent as the first grouping variable and there should always be at least one grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
Analysis Variables	The list of analysis variables that will be used to calculate the Pearson correlation coefficients.
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

In this example, we will compute the Pearson correlation between the scales “Students like reading” (ASBGLR), “Students motivated to read” (ASBGSMR) and “Students confident in their reading” (ASBGSCR).

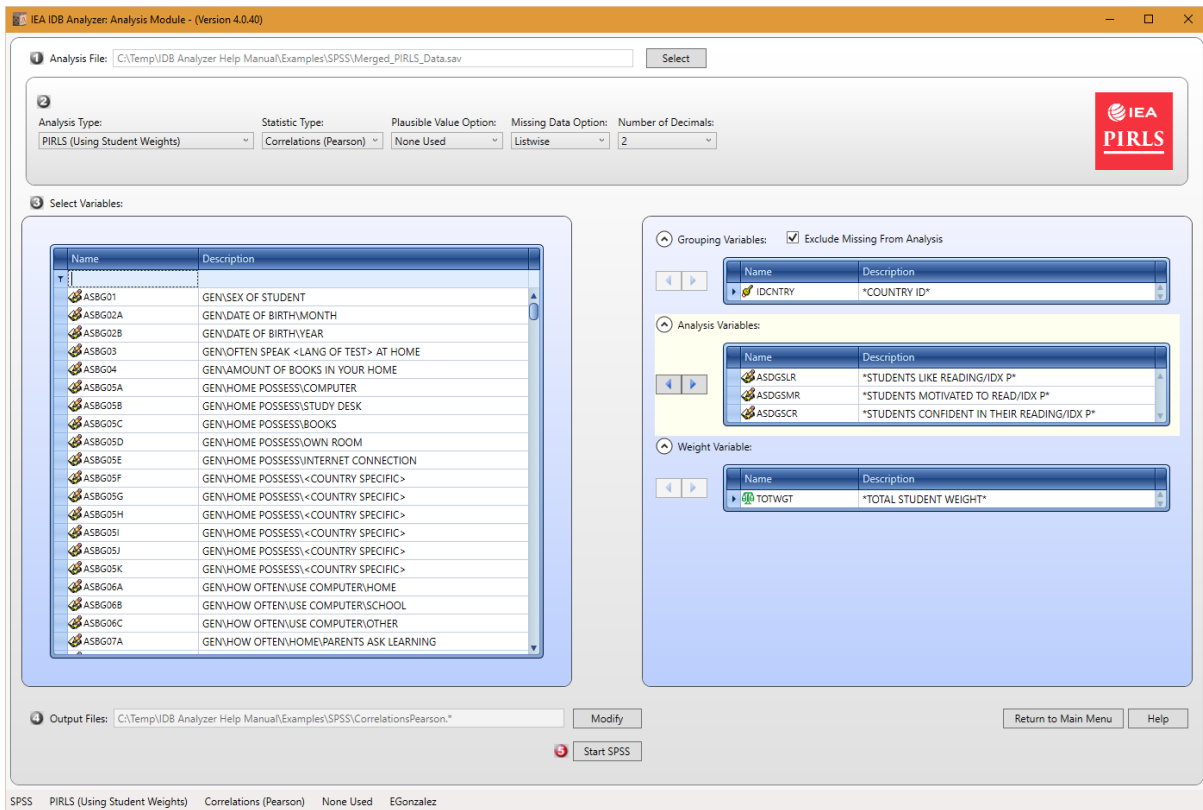
The steps in the IEA IDB Analyzer are as follows:

1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As Analysis Type, select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and variance estimation variables, please refer to the technical documentation specific to the study.
4. From the **Statistic Type** dropdown menu, select **Correlation (Pearson)**. Leave the other dropdown menus unchanged. Note that there are two options under the **Missing Data Option** dropdown menu – **Pairwise** and **Listwise**. Depending on how you want to treat the missing data, you might want to change it. For the time being, we will leave the default setting (**Listwise**). When choosing **Pairwise**, all available data are used in the analysis, when choosing **Listwise** only cases with complete data are used in the analysis.
5. In the next steps, all variables need to be specified:
 - As Grouping Variable, the software always selects variable IDCNTY by default, or its equivalent. No other variable needs to be added for this example.
 - The next step is to select the analysis variables. To activate this section, you will need to click into the area of the **Analysis Variables** field. Now you will need to select variables ASBGLR, ASBGSMR and ASBGSCR from the list of variables and move them to the analysis variables window by pressing the right arrow button in this section.
6. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “CorrelationsPearson”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis. The suffixes **_Desc**, or **_Corr** are added to the filename to identify the statistics contained in the corresponding file. This name will also be used to create and name a new output window with the results from this analysis.

- Click on the **Start SPSS/SAS/R Start SPSS** or **Start SAS** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 15 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 15: Analysis Module Setup for Pearson Correlations



The columns in the “_Desc” output files are the following:

- **Grouping Variables:** Grouping variables used in defining the groups for the analysis (only IDCNTRY in this case)
- **VARIABLE:** The analysis variable specified in the analysis setup
- **MEAN:** Means for the analysis variables
- **STDEV:** Standard deviations of the analysis variables
- **TOTWGT:** Sum of the weights for cases in the groups defined by the Grouping Variables
- **TOTWGT.SE:** Standard error of the weights
- **NOBS:** The number of cases used for this variable.
- **MEAN.SE:** Standard errors of the means of the analysis variables
- **STDEV.SE:** Standard errors of the standard deviations of the analysis variables
- **XVAR:** The list of variables in the analysis
- **WEIGHT:** The weighting variable used for the analysis
- **METHOD:** The method of replication used for the analysis

- MISSOPTN: How the missing responses were handled (e.g., Pairwise, Listwise)
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

The columns in the “_Corr” output files are the following:

- Grouping Variables: Grouping variables used in defining the groups for the analysis (only IDCNTY in this case)
- VARIABLE and ASBGSLR, ASBGSMR and ASBGSCR: Reading across, these cells contain the Pearson correlation coefficient between the variable in the column VARIABLE and the variable listed in the column name.
- ASBGSLR.se, ASBGSMR.se, and ASBGSCR.se: The standard errors for the Pearson correlation coefficients for the corresponding pair of variables.
- XVAR: The name of the variables in the analysis
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- MISSOPTN: How the missing responses were handled (e.g., Pairwise, Listwise)
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

When using the parameter STRATVAR, an additional file will be created ending in “_NVS” (for Number of Variance Strata). This file contains one record per group defined by the grouping variables, and a column labeled NVARSTRAT indicating the number of variance strata contributing cases to the corresponding group. Currently only configured for IEA studies.

Computing Pearson Correlation Coefficients with Plausible Values

To compute Pearson correlation coefficients with plausible values, you need to select “**Correlations (Pearson)**” from the **Statistic Type** dropdown menu, and under **Plausible Value Options** select “**Use PVs**”.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables that are to be used to define the subgroups. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTRY or its equivalent as the first grouping variable and there should always be at least one grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
Analysis Variables	The list of analysis variables that will be used to calculate the Pearson correlation coefficients with the plausible value or achievement scores. If you select more than one set of plausible values, it is not necessary to select any analysis variable.
Plausible Values	The plausible values to be used in the analysis. You can select more than one set of plausible values.
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

In this example, we will compute the Pearson correlations between the plausible values for reading for Literary Purpose (ASRLIT01-ARSLIT05), Reading for Information Purpose (ASRINF01-ASRING05), and the scales “Students like reading” (ASBGSLR) and “Students motivated to read” (ASBGSMR).

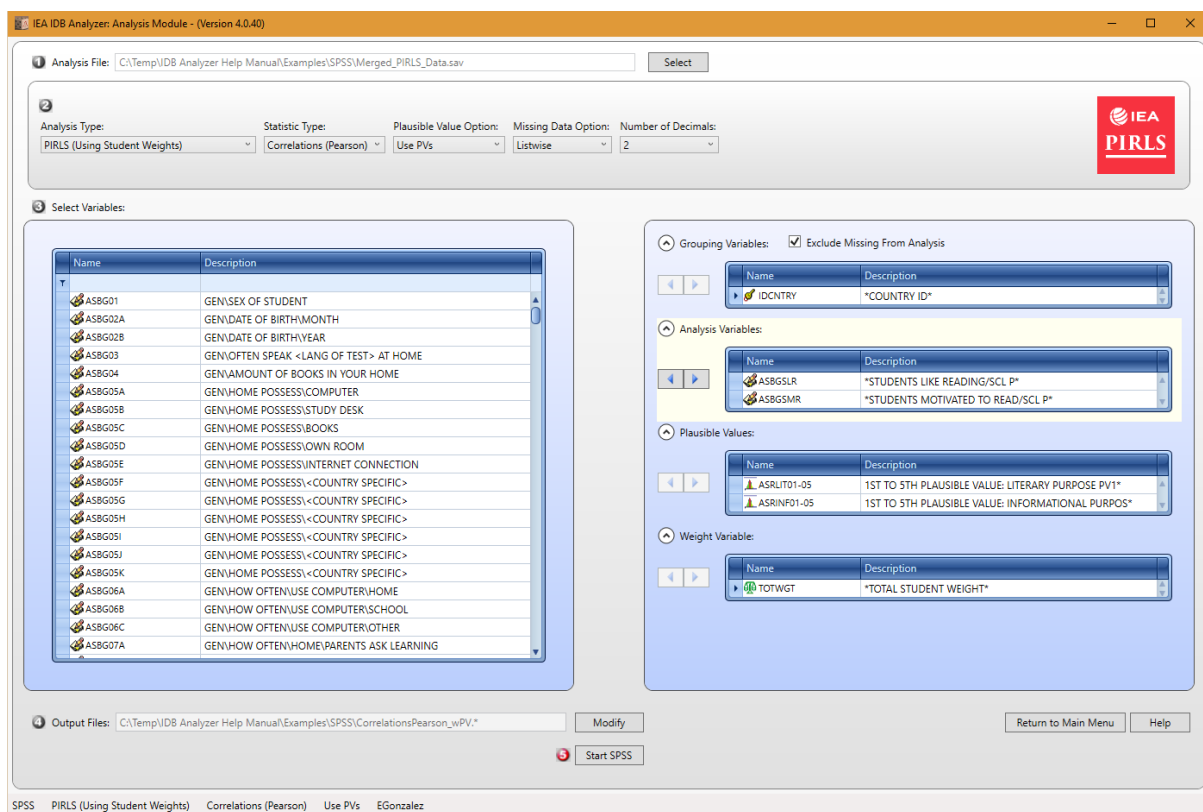
The steps in the IEA IDB Analyzer are as follows:

1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As type of the analysis select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally, the variance estimation variables are preselected by the software. For the correct weight and variance estimation variables, please refer to the technical documentation specific to the study.
4. From the **Statistic Type** dropdown menu, select **Correlations (Pearson)**. From the **Plausible Values Option** dropdown menu, choose **Use PVs**. Leave the other dropdown menus unchanged. Note that there are two options under the **Missing Data Option** dropdown menu – **Pairwise** and **Listwise**. Depending on how you want to treat the missing data, you might change it. For the time being, we will leave it with the default setting: **Listwise**. When choosing **Pairwise**, all available data are used in the analysis, when choosing **Listwise** only cases with complete data are used in the analysis.
5. In the next steps you will need to specify the variables:
 - As Grouping Variable, the software always selects variable IDCNTRY by default, or its equivalent depending on the study. No other variable needs to be added for this example.
 - Next, select the analysis variables. To activate this section, you will need to click in the area of the **Analysis Variables** field. Now you will need to select variables ASBGSLR and ASBGSMR from the list of variables and move them to the analysis variables window by pressing the right arrow button in this section.

- Next the plausible values need to be selected. To activate this section, you will need to click into the area of the **Plausible Values** field. Now you will need to select variable ASRLIT01-05 and ASRINF01-05 from the list of variables and move it to the **Plausible Values** field by pressing the right arrow button in this section.
6. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “CorrelationsPearson_wPV”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis. The suffixes _Desc, or _Corr are added to the filename to identify the statistics contained in the corresponding file. This name will also be used to create and name a new output window with the results from this analysis.
 7. Click on the **Start SPSS/SAS/R Start SPSS** or **Start SAS** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 16 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 16: Analysis Module Setup for Computing Pearson Correlations with Plausible Values



The columns in the “_Desc” Excel file and SPSS datasetoutput files are the following:

- **Grouping Variables:** Grouping variables used in defining the groups for the analysis (only IDCNTRY in this case)
- **VARIABLE:** The analysis variable specified in the analysis setup. Plausible values are listed without their sequential number. In our example instead of ASRLIT01...05 you will see “ASRLIT0_”.
- **MEAN:** Means for the analysis variables

- STDEV: Standard deviations of the analysis variables
- TOTWGT: Sum of the weights for cases in the groups defined by the Grouping Variables
- TOTWGT.SE: Standard error of the weights.
- NOBS: The number of cases used for this variable.
- MEAN.SE: Standard errors of the means of the analysis variables
- STDEV.SE: Standard errors of the standard deviations of the analysis variables
- XVAR: The list of variables in the analysis
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- SHORTCUT: whether the sampling variance was calculated using all plausible values (N), or just the first plausible value (Y).
- MISSOPTN: How the missing responses were handled (e.g., Pairwise, Listwise)
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- NPV: The number of plausible values used in the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

The columns in the “_Corr” output files Excel file and SPSS dataset are the following:

- Grouping Variables: Grouping variables used in defining the groups for the analysis (only IDCNTRY in this case)
- VARIABLE and ASBGSLR, ASBGSMR, ASRLITO_ and ASRINFO_: Reading across, these cells contain the Pearson correlation coefficient between the variable in the column VARIABLE and the variable listed in the column name.
- ASBGSLR.se, ASBGSMR.se, ASRLITO_.se and ASRINFO_.se: The standard errors for the Pearson correlation coefficients for the corresponding pair of variables.
- XVAR: The name of the variables in the analysis
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- SHORTCUT: Whether the sampling variance was calculated using all plausible values (N), or just the first plausible value (Y).
- MISSOPTN: How the missing responses were handled (e.g., Pairwise, Listwise)
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- NPV: The number of plausible values used in the analysis.
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

When using the parameter STRATVAR, an additional file will be created ending in “_NVS” (for Number of Variance Strata). This file contains one record per group defined by the grouping variables, and a

column labeled NVARSTRAT indicating the number of variance strata contributing cases to the corresponding group. Currently only configured for IEA studies.

Computing Spearman Rank-Order Correlation Coefficients

To compute Spearman rank-order correlation coefficients, you need to select “**Correlations (Spearman)**” from the **Statistic Type** dropdown menu.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables that are to be used to define the subgroups. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTY or its equivalent as the first grouping variable and there should always be at least one grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
Analysis Variables	The list of analysis variables that will be used to calculate the Spearman rank-order correlation coefficients. Since the Spearman rank-order correlation coefficient is intended to be computed between two ordinal variables, the analysis variables should all be ordinal. If they are not, you should use instead the Correlations (Pearson) option. The program does not check for this condition.
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

In this example, we will compute the Spearman rank-order correlation between the ordinal variables “Do you like being in school” (ASBG08A), “Do you feel safe at school” (ASBG08B) and “Do you feel like you belong at school” (ASBG08C). The response options to these variables are in a Likert scale format (Strongly Agree, Agree, Disagree and Strongly Disagree).

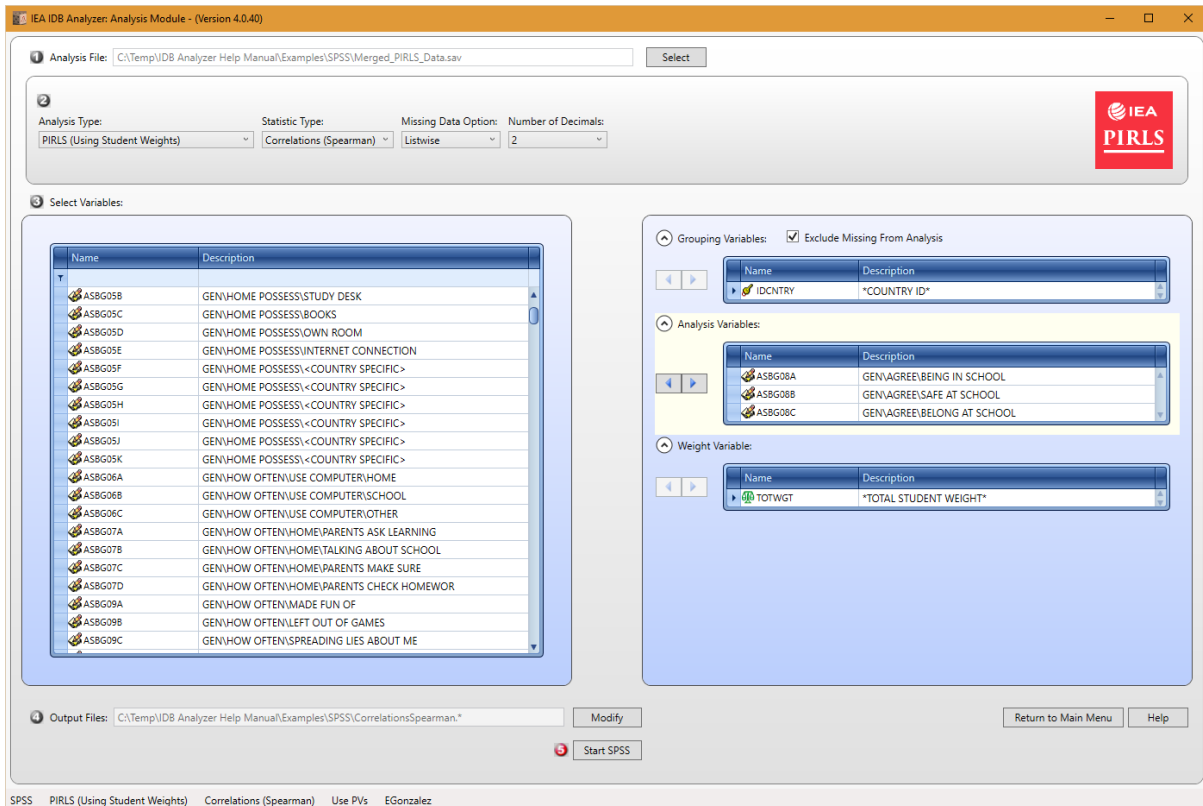
The steps in the IEA IDB Analyzer are as follows:

1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As Analysis Type, select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and variance estimation variables, please refer to the technical documentation specific to the study.
4. From the **Statistic Type** dropdown menu, select **Correlation (Spearman)**. Leave the other dropdown menus unchanged. Note that there are two options under the **Missing Data Option** dropdown menu – **Pairwise** and **Listwise**. Depending on how you want to treat the missing data, you might want to change it. For the time being, we will leave the default setting (**Listwise**). When choosing **Pairwise**, all available data are used in the analysis, when choosing **Listwise** only cases with complete data are used in the analysis.
5. In the next steps, all variables need to be specified:
 - As Grouping Variable, the software always selects variable IDCNTY by default, or its equivalent. No other variable needs to be added for this example.
 - The next step is to select the analysis variables. To activate this section, you will need to click into the area of the **Analysis Variables** field. Now you will need to select variables ASBG08A, ASBG08B and ASBG08C from the list of variables and move them to the analysis variables window by pressing the right arrow button in this section.

6. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “CorrelationsSpearman”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis. The suffixes `_Desc`, or `_Corr` are added to the filename to identify the statistics contained in the corresponding file. This name will also be used to create and name a new output window with the results from this analysis.
7. Click on the **Start SPSS/SAS/R Start SPSS** or **Start SAS** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 17 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 17: Analysis Module Setup for Computing Spearman Correlations



The columns in the “_Desc” output files are the following²⁷:

- Grouping Variables: Grouping variables used in defining the groups for the analysis (only IDCNTRY in this case)
- VARIABLE: The analysis variable specified in the analysis setup
- MEAN: Means for the analysis variables
- STDEV: Standard deviations of the analysis variables
- TOTWGT: Sum of the weights for cases in the groups defined by the Grouping Variables

²⁷ Note that the descriptive statistics are computed for the original variables after they are ranked. Therefore, we have added the letter R to the names of each of the analysis variables.

- TOTWGT.SE: Standard error of the weights
- NOBS: The number of cases used for this variable.
- MEAN.SE: Standard errors of the means of the analysis variables
- STDEV.SE: Standard errors of the standard deviations of the analysis variables
- XVAR: The list of variables in the analysis
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- MISSOPTN: How the missing responses were handled (e.g., Pairwise, Listwise)
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

The columns in the “_Corr” output files are the following²⁸:

- Grouping Variables: Grouping variables used in defining the groups for the analysis (only IDCNTY in this case)
- VARIABLE and ASBG08A, ASBG08B and ASBG08C: Reading across, these cells contain the Spearman rank-order correlation coefficient between the variable in the column VARIABLE and the variable listed in the column name.
- ASBG08A.se, ASBG08B.se, and ASBG08C.se: The standard errors for the Pearson correlation coefficients for the corresponding pair of variables.
- XVAR: The name of the variables in the analysis
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- MISSOPTN: How the missing responses were handled (e.g., Pairwise, Listwise)
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis.

When using the parameter STRATVAR, an additional file will be created ending in “_NVS” (for Number of Variance Strata). This file contains one record per group defined by the grouping variables, and a column labeled NVARSTRAT indicating the number of variance strata contributing cases to the corresponding group. Currently only configured for IEA studies.

²⁸ In order to emphasize that these results are Spearman Rank-order correlations, we have added the letter R to the variable names in this output to indicate we are correlating Ranks.

Computing Linear Regression Coefficients

To compute linear regression statistics with variables that do not involve plausible values, you need to select “**Linear Regression**” from the **Statistic Type** dropdown menu. The Appendix of this Help Manual describes additional uses and interpretation of linear regression coefficients when using dummy and effect coded variables.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables that are to be used to define the subgroups. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTRY or its equivalent as the first grouping variable and there should always be at least one grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
Independent Variables	<p>This is the list of analysis variables used as predictors in the linear regression model. The independent variables can be classified as categorical or continuous. Variables classified as categorical will be either dummy or effect contrast coded. Variables classified as continuous will be entered in the equation without further recoding. You can enter any combination of categorical or continuous variables.</p> <p>For each variable classified as categorical you will need to enter the number of categories and the reference category. Reference categories are selected by sort order of the values for the variable.²⁹ The program will automatically create dummy or effect coded variables for each of the non-reference categories. It will use the original variable name, plus a “D” or “E” followed by the category represented by the variable. For example, if you specify to effect code variable ASBG04, with 5 categories, and use the 3rd category as the reference category, the program will create the following variables: ASBG04_E1, ASBG04_E2, ASBG04_E4 and ASBG04_E5, and will use these in the analysis. Please note that ANY case with a missing value on any variable classified as categorical will be deleted from the analysis. If you want include these cases in the analysis you will need to recode the missing values to non-missing values.</p>
Dependent Variable	The dependent variable to be predicted by the list of independent variables. Only one dependent variable can be listed for each analysis specification.
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

As an example, we will compute a linear regression equation predicting how much students like reading (ASBGSLR) as a function of the number of books they have in the home (ASBG04), and how confident they are in their reading (ASBGSCR). The variable “books in the home” has 5 categories and it will be effect coded, using the 3rd category as the reference category. The resulting regression coefficients will tell us the difference between the mean of the 5 group means, and categories 1, 2, 4 and 5 for the variable books in the home.

The data will be read from the data file **Merged_PIRLS_Data** and the standard errors will be computed based on replicate weights.

²⁹ The reference category corresponds to the sequence of the sorted categories. For example, if you select 3 as your reference category, and the variable takes on values of A, B, D, E and F, reference category 3 corresponds to D. If the categories are -10, 0, 10, and 20, reference category 3 corresponds to category 10.

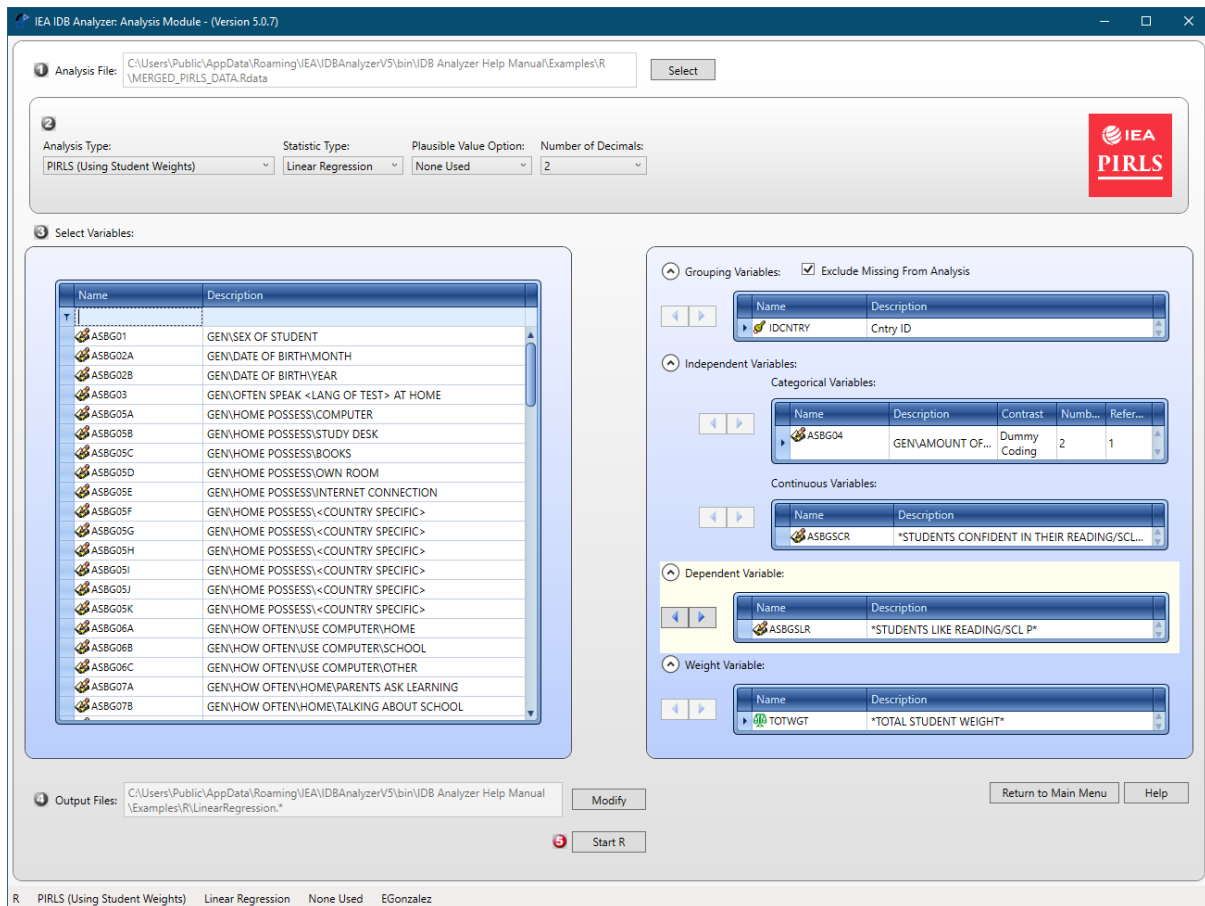
The steps in the IEA IDB Analyzer are as follows:

1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As type of the analysis, select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally, the variance estimation procedure is defined according to the technical specifications of the study.
4. From the **Statistic Type** dropdown menu, select **Linear Regression**.
5. Note that there are three options under the **Missing Data Option** dropdown menu – **Pairwise**, **Listwise** and **MeanSubstitution**³⁰. Depending on how you want to treat the missing data, you might change it. For the time being, we will leave it as default (listwise). When choosing **Pairwise**, all available data are used in the analysis, when choosing **Listwise** only cases with complete data are used in the analysis, when choosing **MeanSubstitution** missing data will be replaced with the mean for the variable. We do not recommend the use of **MeanSubstitution** when entering categorical variables in your analysis. This option is only used to select cases based on the continuous variables. Cases with missing values in any of the categorical variables are deleted from the analysis file.
6. In the next steps, all variables for the analysis are selected:
 - As Grouping Variables, the software always selects variable IDCNTY by default. No other variable needs to be added for this example.
 - Next the independent variables need to be identified. To activate this section, click into the area of the **Independent Variables** field. Now you will need to select variable ASBG04 as a categorical variable, select “Effect Coding”, 5 for the “Number of Categories”, and 3 as your reference category. As your continuous independent variable for the analysis choose ASBGSCR.
 - Click on the **Dependent Variable** field. Select variable ASBGSLR from the variable list and move it to the dependent variable field by pressing the right arrow button in this section or by double clicking on the variable name.
7. The weight variable is automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and jackknifing variables, please refer to the technical documentation specific to the study.
8. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “LinearRegression”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis. The suffixes **_Desc**, **_Model** and **_Coef** are added to the filename to identify the statistics contained in the corresponding file.
9. Click on the **Start SPSS/SAS/R** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 18 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

³⁰ Mean Substitution is only available with SPSS. Pairwise deletion is only available in SPSS and SAS.

Figure 18: Analysis Module Setup for Computing Linear Regression Coefficients



The columns in the “_Desc” output files are the following:

- Grouping Variables: Grouping variables used in defining the groups in the analysis (only IDCNTRY in this case)
- EQVAR: Variables included in the linear regression equation
- MEAN: Means of the variables included in the linear regression equation
- STDEV: Standard deviations of the variables included in the linear regression equation
- VAR: Variances of the variables included in the linear regression equation
- TOTWGT: Sum of the weights for cases in the groups defined by the Grouping Variables
- TOTWGT.se: Standard error of the sum of the weights
- NOBS: The number of cases used for this variable.
- MEAN.SE: Standard errors of the means of the variables included in the linear regression equation
- STDEV.SE: Standard errors of the standard deviations of the variables included in the linear regression equation
- VAR.SE: Standard errors of the variances of the variables included in the linear regression equation
- XVAR: The name of the independent variables in the analysis. Notice that for the categorical variables an index has been added indicating the type of contrast coding used (E for effect, D for Dummy) as well as the category represented by that variable.
- DVAR: The name of the dependent variable in the analysis

- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- MISSOPTN: Whether pairwise, listwise or mean substitution was used to deal with missing data
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

The columns in the “_Model” output files are the following:

- Grouping Variables: Grouping variables used in defining the groups in the analysis (only IDCNTRY in this case)
- RSQ: The multiple R squared coefficients
- ARSQ: Adjusted multiple R squared coefficients³¹
- RSQ.SE: Standard error of the multiple R squared coefficients
- ARSQ.SE: Standard error of the adjusted multiple R squared coefficients
- XVAR: The name of the independent variables in the analysis
- DVAR: The name of the dependent variable in the analysis
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- MISSOPTN: Whether pairwise, listwise or mean substitution was used to deal with missing data
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

The columns in the “_Coef” output files are the following:

- Grouping Variables: Grouping variables used in defining the groups in the analysis (only IDCNTRY in this case)
- EQVAR: Variables included in the linear regression equation
- B: Linear regression coefficients (constant for the model and coefficients for each variable in the equation)
- BETA: Standardized linear regression coefficients
- B.SE: Standard errors for the linear regression coefficients
- BETA.SE: Standard errors for the standardized linear regression coefficients

³¹ The Adjusted R Squared statistic is calculated as $[1 - (1 - R_Square) * (n - 1) / (n - p - 1)]$, where p is the number of regressors, and n is the sample size. While most statistical software use the actual number of cases for the value of n , SPSS and the R module of the IEA IDB Analyzer use the sum of the weights, thus resulting in different values for the Adjusted R Squared statistics when compared to those calculated by other software.

- B.T: t-statistics for the linear regression coefficients
- BETA.T: t-statistics for the standardized linear regression coefficients
- XVAR: The name of the independent variables in the analysis
- DVAR: The name of the dependent variable in the analysis
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- MISSOPTN: Whether pairwise, listwise or mean substitution was used to deal with missing data
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

When using the parameter STRATVAR, an additional file will be created ending in “_NVS” (for Number of Variance Strata). This file contains one record per group defined by the grouping variables, and a column labeled NVARSTRAT indicating the number of variance strata contributing cases to the corresponding group. Currently only configured for IEA studies.

Computing Linear Regression Coefficients with Plausible Values

To compute linear regression statistics with variables that include plausible values, you need to select “Linear Regression” from the **Statistic Type** dropdown menu, and under **Plausible Value Options** select “Use PVs”. When selecting “Use PVs”, you must select at least one set of plausible values for your dependent or independent variable list. The Appendix of this Help Manual describes additional uses and interpretation of linear regression coefficients when using dummy and effect coded variables.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables	This is the list of variables that are to be used to define the subgroups. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTY or its equivalent as the first grouping variable and there should always be at least one grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis. This is the default option.
Independent Variables	<p>This is the list of analysis variables used as predictors in the linear regression model. The independent variables can be classified as categorical, continuous or plausible values. Variables classified as categorical will be either dummy or effect contrast coded. Variables classified as continuous will be entered in the equation without further recoding. You can enter any combination of categorical or continuous variables.</p> <p>For each variable classified as categorical you will need to enter the number of categories and the reference category. Reference categories are selected by sort order of the values for the variable.³² The program will automatically create dummy or effect coded variables for each of the non-reference categories. It will use the original variable name, plus a “D” or “E” followed by the category represented by the variable. For example, if we specify to dummy code variable ASBG01, with 2 categories, and use the 1st category as the reference category, it will create the following variable ASBG01_D2 and use this in the analysis. Please note that ANY case with a missing value on any variable classified as categorical will be deleted from the analysis. If you want to include these cases in the analysis you will need to recode the missing values to non-missing values.</p> <p>As continuous variables you can choose any numeric variable in the files. While plausible values are treated as continuous variables, they have to be entered in a separate window.</p>
Dependent Variable	The dependent variable to use in the analysis. This can be a continuous variable, or a plausible value.
Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

Please note that when selecting “Use PVs” with linear regression, you **MUST** select at least one set of plausible values, either as a dependent or independent variable. You can also select plausible values for both: dependent and independent variable. If you do not select a set of plausible values for the

³² The reference category corresponds to the sequence of the sorted categories. For example, if you select 3 as your reference category, and the variable takes on values of A, B, D, E and F, reference category 3 corresponds to D. If the categories are -10, 0, 10, and 20, reference category 3 corresponds to category 10.

analysis, the program will not let you continue. You can select one or more plausible values as independent variable.

As an example, we will compute a linear regression equation predicting reading proficiency as a function of gender (ASBG01), and how confident students are in their reading (ASBGSCR). The variable ASBG01 has 2 categories, 1 for girls and 2 for boys, and it will be dummy coded, using the 1st category as the reference category. The resulting linear regression coefficient will tell us the difference between males and females in reading, after accounting for their confidence in reading.

The data will be read from the data file **Merged_PIRLS_Data** and the standard errors will be computed based on replicate weights and the plausible values.

The steps in the IEA IDB Analyzer are as follows:

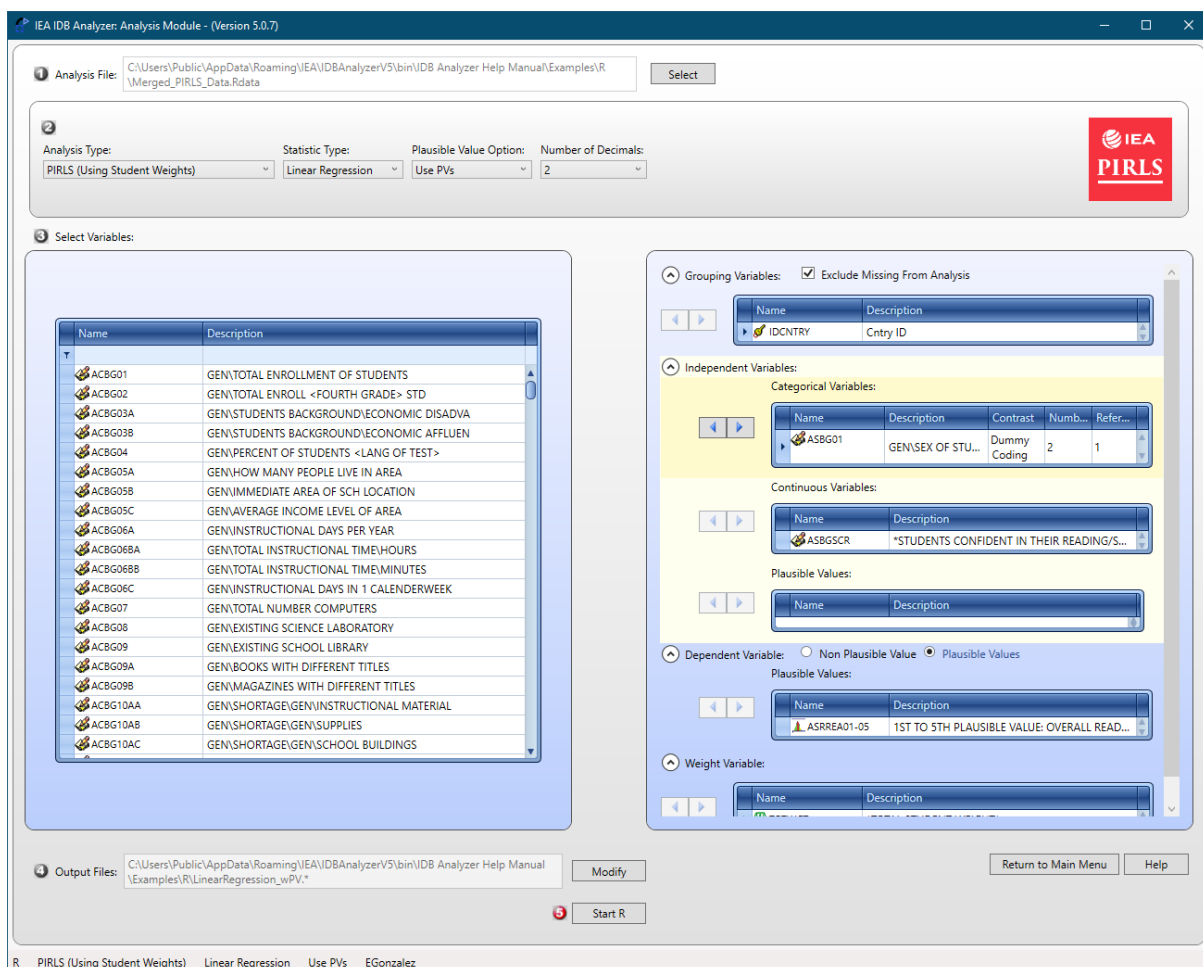
1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data** that was created with the **Merge Module**.
3. As type of the analysis, select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and jackknifing variables, please refer to the technical documentation specific to the study.
4. From the **Statistic Type** dropdown menu, select **Linear Regression**. From the **Plausible Values Option** dropdown menu, choose **Use PVs**. Leave the other dropdown menus unchanged.
5. Note that there are three options under the **Missing Data Option** dropdown menu – **Pairwise**, **Listwise** and **MeanSubstitution**³³. Depending on how you want to treat the missing data, you might change it. For the time being, we will leave it as default (listwise). When choosing **Pairwise**, all available data are used in the analysis, when choosing **Listwise** only cases with complete data are used in the analysis, when choosing **MeanSubstitution** missing data will be replaced with the mean for the variable. We do not recommend the use of **MeanSubstitution** when entering categorical variables in your analysis. This option is only used to select cases based on the missing values of continuous variables. Cases with missing values in any of the categorical variables are deleted from the analysis file.
6. In the next steps all variables for the analysis are selected:
 - As **Grouping Variable**, the software always selects variable IDCNTY by default. No other variable needs to be added for this example.
 - Next the independent variables need to be identified. To activate this section, click into the area of the **Independent Variables** field. Now you will need to select variable ASBG01 as a categorical variable, select “Dummy Coding”, 2 for the “Number of Categories”, and 1 as your reference category. As your continuous independent variable for the analysis choose ASBGSCR.
 - Next the dependent variable needs to be specified. To activate this section, you will need to click into the area of the **Dependent Variables** section and select the button for “Plausible Values”. This time you will need to select variable ASRREA01-05 from the list of variables and move it to the **Plausible Values** section of the dependent variables by pressing the right arrow button in this section.
7. The weight variable is automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and jackknifing variables, please refer to the technical documentation specific to the study.

³³ Mean Substitution is only available with SPSS.

8. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “LinearRegression_wPV”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis. The suffixes **_Desc**, **_Model** or **_Coef** are added to the filename to identify the statistics contained in the corresponding file. This name will also be used to create and name a new output window with the results from this analysis.
9. Click on the **Start SPSS/SAS/R** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 19 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 19: Analysis Module Setup for Computing Linear Regression with Plausible Values



The columns in the “_Desc” output files are the following:

- **Grouping Variables:** Grouping variables used in defining the groups in the analysis (only IDCNTRY in this case)
- **EQVAR:** Variables included in the analysis
- **TOTWGT:** Sum of the weights for cases in the groups defined by the Grouping Variables
- **TOTWGT.se:** Standard error of the sum of the weights

- NOBS: The number of cases used for this variable.
- MEAN: Means of the variables included in the linear regression equation
- STDEV: Standard deviations of the variables included in the linear regression equation
- VAR: Variances of the variables included in the linear regression equation
- MEAN.SE: Standard errors of the means of the variables included in the linear regression equation
- STDEV.SE: Standard errors of the standard deviations of the variables included in the linear regression equation
- VAR.SE: Standard errors of the variances of the variables included in the linear regression equation
- XVAR: The name of the independent variables in the analysis
- DVAR: The name of the dependent variable in the analysis
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- SHORTCUT: Whether the sampling variance was calculated using all plausible values (N), or just the first plausible value (Y).
- MISSOPTN: Whether pairwise, listwise or mean substitution was used to deal with missing data
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- NPV: Number of plausible values used in the analysis.
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

The columns in the “_Model” output files are the following:

- Grouping Variables: Grouping variables used in defining the groups in the analysis (only IDCNTRY in this case)
- RSQ: The multiple R squared coefficients
- ARSQ³⁴: Adjusted multiple R squared coefficients
- RSQ.SE: Standard error of the multiple R squared coefficients
- ARSQ.SE: Standard error of the adjusted multiple R squared coefficients
- XVAR: The name of the independent variables in the analysis
- DVAR: The name of the dependent variable in the analysis
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- SHORTCUT: Whether the sampling variance was calculated using all plausible values (N), or just the first plausible value (Y).

³⁴ The Adjusted R Squared statistic is calculated as $[1 - (1 - R_Square) * (n - 1) / (n - p - 1)]$, where p is the number of regressors and n is the sample size. While most statistical software used the actual number of cases for the value of n , SPSS uses the sum of the weights, thus resulting in different values for the Adjusted R Squared statistics when compared to those calculated by other software.

- MISSOPTN: Whether pairwise, listwise or mean substitution was used to deal with missing data
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- NPV: Number of plausible values used in the analysis.
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

The columns in the “_Coef” output files are the following:

- Grouping Variables: Grouping variables used in defining the groups in the analysis (only IDCNTRY in this case)
- EQVAR: Variables included in the linear regression equation
- B: Linear regression coefficients (constant for the model and coefficients for the separate variables)
- BETA: Standardized linear regression coefficients
- B.SE: Standard errors for the linear regression coefficients
- BETA.SE: Standard errors for the standardized linear regression coefficients
- B.T: t-statistics for the linear regression coefficients
- BETA.T: t-statistics for the standardized linear regression coefficients
- XVAR: The name of the independent variables in the analysis
- DVAR: The name of the dependent variable in the analysis
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- SHORTCUT: Whether the sampling variance was calculated using all plausible values (N), or just the first plausible value (Y).
- MISSOPTN: Whether pairwise, listwise or mean substitution was used to deal with missing data
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- NPV: Number of plausible values used in the analysis.
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

When using the parameter STRATVAR, an additional file will be created ending in “_NVS” (for Number of Variance Strata). This file contains one record per group defined by the grouping variables, and a column labeled NVARSTRAT indicating the number of variance strata contributing cases to the corresponding group. Currently only configured for IEA studies.

Computing Logistic Regression Coefficients³⁵

Logistic regression is used to predict a binary response based on one or more predictor variables. To compute logistic regression statistics with variables that do not include plausible values, you need to select “**Logistic Regression**” from the **Statistic Type** dropdown menu.

This analysis type requires the selection of the following variables for the analysis:

Grouping Variables³⁶ This is the list of variables that are to be used to define the subgroups. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTRY or its equivalent as the first grouping variable and there should always be at least one grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.

Independent Variables This is the list of analysis variables used as predictors in the logistic regression model. The independent variables can be classified as categorical or continuous. Variables classified as categorical will be contrast coded according to the specifications of the statistical engine used.

In SPSS you have the options of entering categorical variables using INDICATOR (default), DEVIATION, SIMPLE, DIFFERENCE, HELMERT or REPEATED contrast coding.

In SAS you have the options of entering categorical variables using EFFECT (default), ORDINAL, THERMOMETER, POLYNOMIAL, REFERENCE, ORTHEFFECT, ORTHORDINAL, ORTHOTERM, ORTHPOLY or ORTHRF contract coding.

For each categorical variable you will need to specify the reference category or use the default (highest one). For more information on the use of each of these contrast specifications please refer to the corresponding SPSS or SAS documentation.

Variables classified as continuous will be entered in the equation without further recoding. You can enter any combination of categorical or continuous variables. In addition, you can enter interaction effects between the variables.

Please note that ANY case with a missing value on any variable classified as categorical will be deleted from the analysis. If you want to use these cases, you will need to recode the missing values to non-missing values.

Dependent Variable The dependent variable to be predicted by the independent variables. Only one dependent variable can be listed for each analysis specification. The dependent variable must be dichotomous.

In SPSS, when estimating the equation, the program will predict membership to the highest category of the dependent variable.

In SAS you can specify the link function (LOGIT, CLOGLOG, GLOGIT, PROBIT), as well as the event category for the binary response model. You can specify FIRST or LAST. Please refer to the SAS documentation for information on these link functions.

³⁵ You need to have the Logistic Regression module installed in your system in order to use this functionality. This might require purchasing an additional module from your software vendor.

³⁶ Note that due to a bug in the SPSS Logistic Regression module, when the grouping variables defines a single group in the file, the program crashes. You will know this is the case by checking the descriptive statistics at the beginning of the output file. When this happens, you will need to set the parameter ONEGRP = Y.

Weight Variable	The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.
-----------------	--

As an example, we will compute a logistic regression predicting whether students respond “Agree a lot” when asked “What do you think about your school? Tell us how much you agree with the statement: I like being in school.” (Question G8a in the PIRLS 2011 Student Background Questionnaires, variable ASBG08A). Since the student had 4 response options, we recoded them into a binary variable where a 2 signifies “Agree a lot” and a 1 signifies “Does Not Agree a Lot.” You can see how this variable was recoded in this file “RecodeLikingSchool” by selecting the box **Access the Sample Files** from the main menu.

We will compute a logistic regression equation predicting the log of the odds of a student agreeing a lot that they like school (ASBG08A_r) as a function of their gender (ASBG01), and the scales on being bullied at school (ASBGSBS) and being engaged in reading lessons (ASBGERL). We will also calculate the interaction effects between gender and being bullied at school.

The data will be read from the data file **Merged_PIRLS_Data_r** which contains the recoded dependent variable, and the standard errors will be computed based on replicate weights. If you have been working through the examples in this manual, you will need to select this new file.

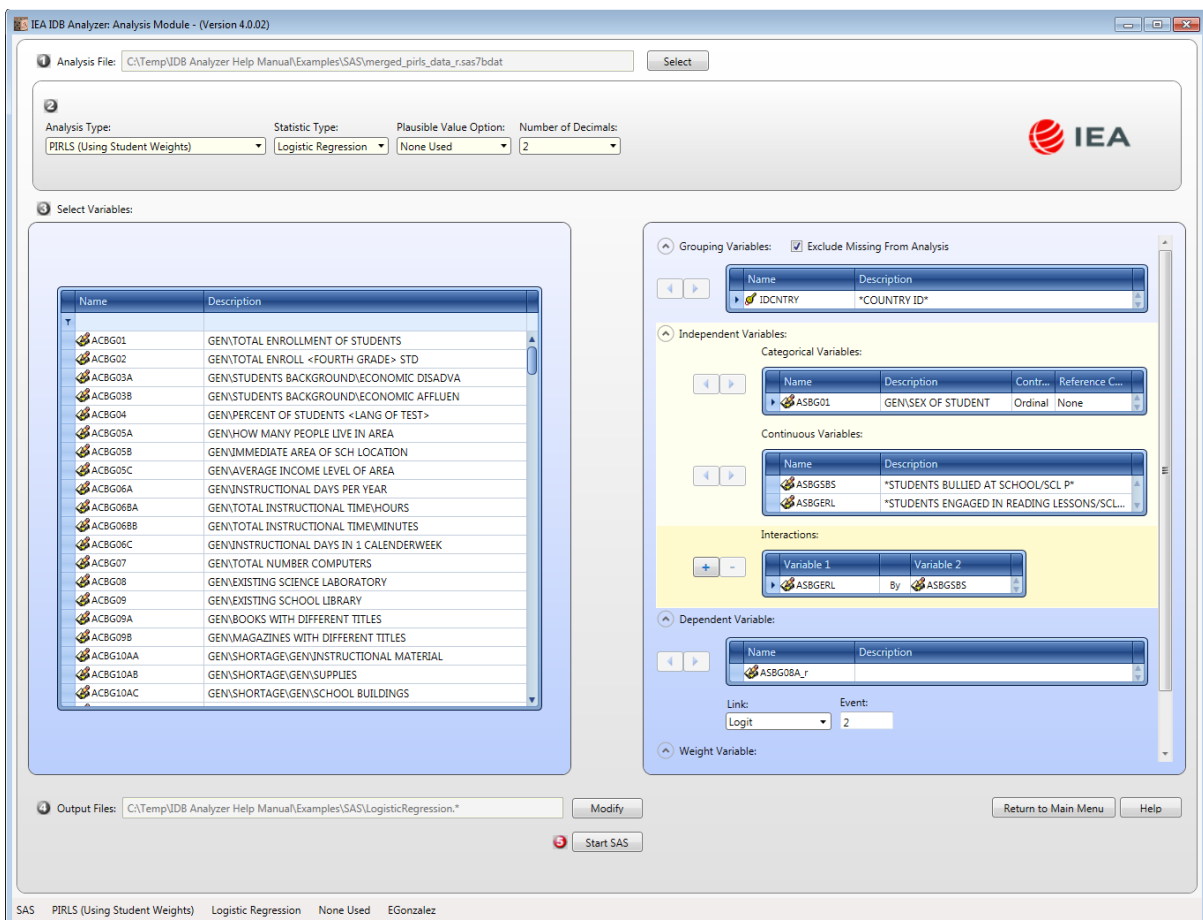
The steps in the IEA IDB Analyzer are as follows:

1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data_r** .
3. As type of the analysis, select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally, the variance estimation procedure is defined according to the technical specifications of the study.
4. From the **Statistic Type** dropdown menu, select **Logistic Regression**.
5. Note that there is no **Missing Data Option available for Logistic Regression**. Logistic regression excludes all cases with missing values on any of the independent variables.
6. In the next steps, all variables for the analysis are selected:
 - As Grouping Variables, the software always selects variable IDCNTY by default. No other variable needs to be added for this example.
 - Next the independent variables and interactions need to be identified. To activate this section, click into the area of the **Independent Variables** field. Now you will need to select variable ASBG01 as a categorical variable, select the type of contrast and reference category for the analysis. As your continuous independent variable for the analysis choose ASBGSBS and ASBGERL. For interaction effects select the variable ASBG01 by ASBGSBS. Variables used in the interaction must be first selected as categorical or continuous variables.
 - Click on the **Dependent Variable** field. Select variable ASBG08A_r from the variable list and move it to the dependent variable field by pressing the right arrow button in this section or by double clicking on the variable name.
7. The weight variable is automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and jackknifing variables, please refer to the technical documentation specific to the study.

8. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “LogisticRegression”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis. The suffixes `_Model` or `_Coef` are added to the filename to identify the statistics contained in the corresponding file.
9. Click on the **Start SPSS/SAS/R** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file.

Figure 20 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 20: Analysis Module Setup for Computing Logistic Regression Coefficients



The columns in the “_Model” output files are the following:

- Grouping Variables: Grouping variables used in defining the groups in the analysis (only IDCNTRY in this case)
- LL: -2 Log Likelihood Coefficient
- CSR: Cox and Snell R Squared coefficient
- NKR: Nagelkerke R Squared coefficient
- LL_SE: The standard error of the -2 Log Likelihood coefficient
- CSR_SE: The standard error of the Cox and Snell R Squared coefficient

- NKR_SE: The standard error of the Nagelkerke R Squared coefficient
- DVAR: The name of the dependent variable in the analysis
- CONVAR: The name of the continuous independent variable in the analysis
- CATVAR: The name of the categorical independent variables in the analysis
- CONTRAST and REFCAT: In the sequence of CATVAR, the contrast type and reference category used for each of the categorical variables
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

The columns in the “_Coef” output files are the following:

- Grouping Variables: Grouping variables used in defining the groups in the analysis (only IDCNTRY in this case)
- EQVAR: Variables included in the logistic regression equation
- B: Logistic regression coefficients (constant for the model and coefficients for each variable in the equation)
- DF: The degrees of freedom for the variable
- B.SE: Standard errors for the logistic regression coefficients
- B.WALD: The Wald statistic for each of the variables in the equation
- B.SIG: The significance of the Wald statistic
- ExpB: The exponent of the B or logistic regression coefficient
- ExpB.SE: The standard error of the exponent of the B or logistic regression coefficient
- DVAR: The name of the dependent variable in the analysis
- CONVAR: The name of the continuous independent variable in the analysis
- CATVAR: The name of the categorical independent variables in the analysis
- CONTRAST and REFCAT: In the sequence of CATVAR, the contrast type and reference category used for each of the categorical variables
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

When using the parameter STRATVAR, an additional file will be created ending in “_NVS” (for Number of Variance Strata). This file contains one record per group defined by the grouping variables, and a column labeled NVARSTRAT indicating the number of variance strata contributing cases to the corresponding group. Currently only configured for IEA studies.

Computing Logistic Regression Coefficients with Plausible Values

Logistic regression is used to predict a binary response based on one or more predictor variables. To compute logistic regression statistics with variables that include plausible values, you need to select “**Logistic Regression**” from the **Statistic Type** dropdown menu, and under **Plausible Value Options** select “**Use PVs**”. When selecting “**Use PVs**”, you must select at least one set of plausible values for your independent variable list.

This analysis type requires the selection of the following variables for the analysis:

- Grouping Variables**³⁷ This is the list of variables that are to be used to define the subgroups. The list can consist of one or more variables. The IEA IDB Analyzer always includes IDCNTY or its equivalent as the first grouping variable and there should always be at least one grouping variable. If the option “Exclude Missing from Analysis” is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.
- Independent Variables** This is the list of analysis variables used as predictors in the logistic regression model. The independent variables can be classified as categorical or continuous. Variables classified as categorical will be contrast coded according to the specifications of the statistical engine used.
- In SPSS you have the option of entering categorical variables using INDICATOR (default), DEVIATION, SIMPLE, DIFFERENCE, HELMERT or REPEATED contrast coding.
- In SAS you have the option of entering categorical variables using EFFECT (default), ORDINAL, THERMOMETER, POLYNOMIAL, REFERENCE, ORTHEFFECT, ORTHORDINAL, ORTHOTERM, ORTHPOLY or ORTHRF contract coding.
- For each categorical variable you will need to specify the reference category or use the default (highest one). For more information on the use of each of these contrast specifications please refer to the corresponding SPSS or SAS documentation.
- Variables classified as continuous will be entered in the equation without further recoding. You can enter any combination of categorical or continuous variables. In addition, you can enter interaction effects between the variables.
- Please note that ANY case with a missing value on any variable classified as categorical will be deleted from the analysis. If you want to use these cases, you will need to recode the missing values to non-missing values.
- Dependent Variable** The dependent variable to be predicted by the independent variables. Only one dependent variable can be listed for each analysis specification. The dependent variable must be dichotomous.
- In SPSS the program will predict membership to the highest category of the dependent variable.
- In SAS you can specify the link function (LOGIT, CLOGLOG, GLOGIT, PROBIT), as well as the event category for the binary response model. You can specify FIRST or LAST. Please refer to the SAS documentation for information on these different link functions.

³⁷ Note that due to a bug in the SPSS Logistic Regression module, when the grouping variables defines a single group in the file, the program crashes. You will know this is the case by checking the descriptive statistics at the beginning of the output file. When this happens, you will need to set the parameter ONEGRP = Y.

Should you use plausible values as a dependent variable you will need to provide a cut point for the distribution of scores. For example, if you want to predict membership to the Advanced proficiency level, you would enter 625 as the cut point for the plausible values. The program will use this cut point to divide the distribution of each of the plausible values into two groups: at or above 625, and below 625, and use this categorized distribution as the dependent variable. The cut point provided is treated as the minimum score of the higher category.

Weight Variable The sampling weight that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight and replication variables for the analysis.

As an example, we will compute a logistic regression predicting whether students respond “Agree a lot” when asked “What do you think about your school? Tell us how much you agree with the statement: I like being in school.” (Question G8a in the PIRLS 2011 Student Background Questionnaires, variable ASBG08A). Since the student had 4 response options, we recoded them into a binary variable where a 2 signifies “Agree a lot” and a 1 signifies “Does Not Agree a Lot.” You can see how this variable was recoded in this file “RecodeLikingSchool” by selecting the box **Access the Sample Files** from the main menu.

We will compute a logistic regression equation predicting the log of the odds of a student agreeing a lot that they like school (ASBG08A_r) as a function of their gender (ASBG01), the scales on being bullied at school (ASBGSBS) and being engaged in reading lessons (ASBGERL), and their reading plausible value (ASRREA01-05). We will also calculate the interaction effects between gender and being bullied at school, as well as the interaction between engagement in reading lessons and the reading score.

The data will be read from the data file **Merged_PIRLS_Data_r** which contains the recoded dependent variable ASBG08A_r, and the standard errors will be computed based on replicate weights.

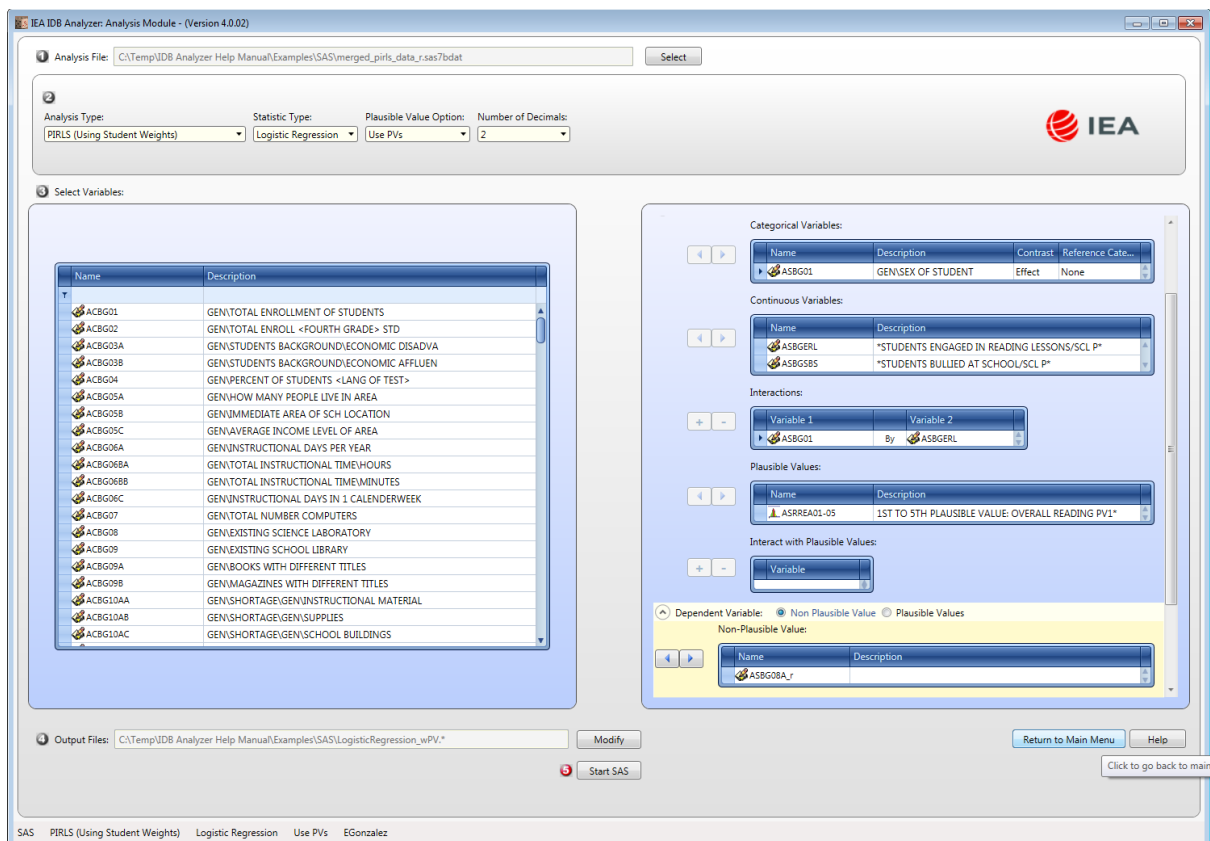
The steps in the IEA IDB Analyzer are as follows:

1. Open the Analysis Module of the IEA IDB Analyzer.
2. Select the data file named **Merged_PIRLS_Data_r**.
3. As type of the analysis, select **PIRLS (Using Student Weights)**. The weight variable is automatically selected by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. Additionally, the variance estimation procedure is defined according to the technical specifications of the study.
4. From the **Statistic Type** dropdown menu, select **Logistic Regression**.
5. Note that there is no **Missing Data Option available for Logistic Regression**. Logistic regression excludes all cases with missing values on any of the independent.
6. In the next steps, all variables for the analysis are selected:
 - As Grouping Variables, the software always selects variable IDCNTY by default. No other variable needs to be added for this example.
 - Next the independent variables and interactions need to be identified. To activate this section, click into the area of the **Independent Variables** field. Now you will need to select variable ASBG01 as a categorical variable, select the contrast and reference category for your analysis. As your continuous independent variable for the analysis choose ASBGSBS and ASBGERL. As your plausible values select ASRREA01-05. As interaction effects select the variable ASBG01 by ASBGSBS, and for interaction with the plausible values select the variable for engagement in reading lessons. When you select a variable to interact with plausible values, this will be combined with all plausible values selected for the analysis. Variables used in the interaction must be first selected as categorical or continuous variables.

- Click on the **Dependent Variable** field. Select variable ASBG08A_r from the variable list and move it to the dependent variable field by pressing the right arrow button in this section or by double clicking on the variable name.
7. The weight variable is automatically defined by the software. As this is an example for analysis on student level, the weight TOTWGT is selected by default. For the correct weight and jackknifing variables, please refer to the technical documentation specific to the study.
 8. Click on the **Define/Modify** button next to **Output Files** and specify the name of the output files. For our example we will use the name “LogisticRegression_wPV”. This filename will be used to create the syntax to perform the analysis, and a set of output files with statistics from the analysis. This name will also be used to create and name a new output window with the results from this analysis. The suffixes _Model or _Coef are added to the filename to identify the statistics contained in the corresponding file.
 9. Click on the **Start SPSS/SAS/R** button to create the syntax file and open it in a syntax window ready for execution. The IEA IDB Analyzer will give a warning if it is about to overwrite an existing file in the specified folder. The syntax file must then be submitted for processing using the corresponding command sequence indicated at the top of the syntax file..

Figure 21 shows the IEA IDB Analyzer Setup Screen for this analysis. The corresponding syntax and output files with the results from this example analysis can be viewed by selecting the button **Access the Sample Files** from the main menu.

Figure 21: Analysis Module Setup for Computing Logistic Regression Coefficients



The columns in the “_Model” output files are the following:

- Grouping Variables: Grouping variables used in defining the groups in the analysis (only IDCNTY in this case)
- LL: -2 Log Likelihood coefficient
- CSR: Cox and Snell R Squared coefficient

- NKR: Nagelkerke R Squared coefficient
- LL_SE: The standard error of the -2 Log Likelihood coefficient
- CSR_SE: The standard error of the Cox and Snell R Squared coefficient
- NKR_SE: The standard error of the Nagelkerke R Squared coefficient
- DVAR: The name of the dependent variable in the analysis
- CONVAR: The name of the continuous independent variable in the analysis
- CATVAR: The name of the categorical independent variables in the analysis
- CONTRAST and REFCAT: In the sequence of CATVAR, the contrast type and reference category used for each of the categorical variables
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- SHORTCUT: Whether only one plausible value was used to calculate the sampling error (Y) or all of them were used (N)
- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- NPV: The number of plausible values used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

The columns in the “_Coef” output files are the following:

- Grouping Variables: Grouping variables used in defining the groups in the analysis (only IDCNTRY in this case)
- EQVAR: Variables included in the logistic regression equation
- B: Logistic regression coefficients (constant for the model and coefficients for each variable in the equation)
- DF: The degrees of freedom for the variable
- B.SE: Standard errors for the logistic regression coefficients
- B.WALD: The Wald statistic for each of the variables in the equation
- B.SIG: The significance of the Wald statistic
- ExpB: The exponent of the B or logistic regression coefficient
- ExpB.SE: The standard error of the exponent of the B or logistic regression coefficient
- DVAR: The name of the dependent variable in the analysis
- CONVAR: The name of the continuous independent variable in the analysis
- CATVAR: The name of the categorical independent variables in the analysis
- CONTRAST and REFCAT: In the sequence of CATVAR, the contrast type and reference category used for each of the categorical variables
- WEIGHT: The weighting variable used for the analysis
- METHOD: The method of replication used for the analysis
- SHORTCUT: Whether only one plausible value was used to calculate the sampling error (Y) or all of them were used (N)

- DATE: The date the analysis was conducted
- TIME: The time the analysis was conducted
- REPS: The number of replicates used for the analysis
- NPV: The number of plausible values used for the analysis
- INFILE: Data used for the analysis
- SELCRIT: Selection criteria used for the analysis

When using the parameter STRATVAR, an additional file will be created ending in “_NVS” (for Number of Variance Strata). This file contains one record per group defined by the grouping variables, and a column labeled NVARSTRAT indicating the number of variance strata contributing cases to the corresponding group. Currently only configured for IEA studies.

Troubleshooting

If you have any problems using the IDB Analyzer, please send an e-mail to: ldb-analyzer@iea-hamburg.de. Please include a description of the problem you are facing with the IEA IDB Analyzer, the operating system on the computer you are using, and the input and output files related to the problems you are encountering.

Appendix A: Working with IEA Data from Multiple Years

The IEA IDB Analyzer is currently not capable of merging IEA data from across multiple years. To create a dataset from multiple years, you will need to follow these steps:

1. Create the corresponding dataset for each of the years using the Merge Module.
2. Open each dataset in SPSS, SAS, or R, and add a variable called YEAR with the corresponding value of the assessment year and save the dataset.³⁸
3. Append the datasets using the corresponding commands in SPSS, SAS, or R.
4. Save the combined dataset with a new name.
5. Open the Analysis Module of the IEA IDB Analyzer and select the file that contains the combined datasets. When conducting the analysis select YEAR as your second grouping variable. The results from the IEA IDB Analyzer will then be by Country and by Year.

Please note that when merging data from across different years you will need to verify that the variable names and variable content correspond across the different years.

Also note that, when following this approach, the statistical significance results comparing results between years will be treated by the IDB Analyzer as comparisons between dependent samples, whereas these should be treated as independent samples. When comparing independent samples, the standard error of the difference is calculated as the square root of the sum of the variances of the two statistics.

³⁸ Alternatively, you can use the corresponding „IN= <varname>“ statement when you append the files from the different years.

Appendix B: Cut points for International Proficiency Levels³⁹

ICCS – Civics and Citizenship Scale

Cut point	Descriptor
563	Level 3
479	Level 2
395	Level 1

ICILS – Computer and Information Literacy Scale

Cut point	Descriptor
661.001	Level 4
576.001	Level 3
492.001	Level 2
407.001	Level 1

LaNA – Mathematics and Reading Scales

Cut point	Descriptor
625	Advanced
550	High
475	Intermediate
400	Low
325	Basic

PIAAC – Literacy Scale

Cut point	Descriptor
376	Level 5
326	Level 4
276	Level 3
226	Level 2

³⁹ Values in the table are the lower bound for the corresponding level or benchmark. For example, in TIMSS, the High level begins with 550 and ends below 625. The Advanced level begins with 625. Enter these values in the IEA IDB Analyzer to obtain percentages in the population at or above the level, or within a level.

Cut point	Descriptor
176	Level 1

PIAAC – Numeracy Scale

Cut point	Descriptor
376	Level 5
326	Level 4
276	Level 3
226	Level 2
176	Level 1

PIAAC – Problem Solving in Technology Rich Environments Scale

Cut point	Descriptor
341	Level 3
291	Level 2
241	Level 1

PIAAC – Adaptive Problem Solving

Cut point	Descriptor
326	Level 4
276	Level 3
226	Level 2
176	Level 1

PIRLS – Reading Literacy Scale

Cut point	Descriptor
625	Advanced
550	High
475	Intermediate
400	Low

PISA – Collaborative Problem-Solving Scale⁴⁰

Cut point	Descriptor
640	Level 4
540	Level 3
440	Level 2
340	Level 1

PISA – Financial Literacy Scale

Cut point	Descriptor
624.63	Level 5
549.86	Level 4
475.10	Level 3
400.33	Level 2
325.57	Level 1

PISA – Mathematics Literacy Scale

Cut point	Descriptor
669.30	Level 6
606.99	Level 5
544.68	Level 4
482.38	Level 3
420.07	Level 2
357.77	Level 1a
295.47	Level 1b
233.17	Level 1c

PISA – Problem Solving Scale⁴¹

Cut point	Descriptor
683.14	Level 6

⁴⁰ Not to be confused with the “PISA Problem Solving” scale.

⁴¹ Not to be confused with the “PISA Collaborative Problem Solving” scale

Cut point	Descriptor
618.21	Level 5
553.28	Level 4
488.35	Level 3
423.42	Level 2
358.49	Level 1

PISA – Reading Literacy Scale

Cut point	Descriptor
698.32	Level 6
625.61	Level 5
552.89	Level 4
480.18	Level 3
407.47	Level 2
334.75	Level 1a
262.04	Level 1b
189.33	Level 1c

PISA – Science Literacy Scale

Cut point	Descriptor
707.93	Level 6
633.33	Level 5
558.73	Level 4
484.14	Level 3
409.54	Level 2
334.94	Level 1a
260.54	Level 1b
185.94	Level 1c

TIMSS – Mathematics and Science Scales (Grades 4 and 8)

Cut point	Descriptor
625	Advanced
550	High
475	Intermediate
400	Low

TIMSS Advanced – Advanced Mathematics and Physics Scales

Cut point	Descriptor
625	Advanced
550	High
475	Intermediate

Appendix C: Interpreting Contrast Coded Variables in Linear Regression

The IEA IDB Analyzer now gives the user the option to create and use contrast coded variables in a linear regression analysis. Use of contrast coding can be very useful for calculating differences between group means, and their corresponding standard errors.

When an independent variable is classified as categorical for the purpose of a linear regression analysis, it will be coded to produce either dummy or effect contrasts. For each variable classified as categorical you will need to specify the number of categories and the reference category. Reference categories are selected by sort order of the values for the variable. For example, if a variable has the following 3 categories: "1", "3", and "5", and you choose 2 as the reference category, those with a value of "3" will constitute this reference group.

The program will automatically create a dummy or effect coded variable for each of the non-reference categories. It will use the original variable name, plus a "D" or "E" followed by the category represented by the variable. For example, if you specify to effect code a 5-category variable called ASBE, and use the 3rd category as the reference category, the program will create the following variables: ASBE_E1, ASBE_E2, ASBE_E4 and ASBE_E5, and will use these in the analysis. If you specify to dummy code a 3-category variable called ASBD, and use the 1st category as the reference category, the program will create the following variables: ASBD_D2 and ASBD_D3, and will use these in the analysis.

The contrast coded variables are computed only for the purpose of the analysis and are then discarded. If you want to use these variables in other analysis types, you will need to manually create contrast coded variables within SPSS and save them with the analysis file. The output contains cross tables between the original variables and the recoded variables so you can check the recording has been done to your specifications and expectations.

Dummy Coding

Using dummy coded variables in linear regression is useful to compare each group against a reference group. For example, compare average achievement of boys to that of the girls; compare average achievement between each category of books in the home to the achievement of those with no books; etc. The IEA IDB Analyzer allows you to choose the reference category, or the group against you will compare each of the other groups.

The IEA IDB Analyzer names the dummy coded variables using the name of the original variable and adding the suffix '_D#' to indicate the comparison group used to create the variable, where # is the sequential number of each category. If you dummy code the variable ITSEX, which has 2 categories (1 for Females and 2 for Males), and you specify 1 as the reference category, it will create the variable ITSEX_D2 which is coded 1 for males, and zero otherwise. If you dummy code the variable ITBOOK, which has 3 categories (1 for "no books", 2 for "some books", and 3 for "lots of books"), and you specify 3 as the reference category, it will create the variables ITBOOK_D1 (coded 1 for those with "no books", and zero otherwise) and ITBOOK_D2 (coded 1 for those with "some books", and zero otherwise).

Cases with the categorical variable coded as missing are excluded from the analysis. Check the section "Check Coding of Coded Variables" in the output window to verify proper coding of the categorical variable has been achieved.

When a dummy coded variable is used in a regression, the intercept or constant is the mean of the reference group (first category), and the slope or regression coefficient is the difference between the mean of the reference group and the group identified (coded 1) with the dummy coded variable. Since the regression coefficients are presented with a standard error and a *t* value, these can be used to test whether a difference between means is statistically significant.

Effect Coding

Effect coding is useful to compare groups against the overall mean of different groups. For example, compare average achievement between each of the categories defined by number of books in the home with the mean of the groups means based on this variable. Effect coded variables are created by

assigning -1, 0 or +1 to cases according to their group membership. There can be as many as $(k-1)$ effect coded variables, k being the number of categories.

The IEA IDB Analyzer allows you to choose the reference category for the effect coding. This corresponds to the group that will be assigned a value of “-1” for each of the effect coded variables. When effect coded variables are used in regression, regression coefficients will be computed for all but the reference category. The coefficient for the reference category will be the sum of the coefficients for all the categories multiplied by minus 1.

The IEA IDB Analyzer names the effect coded variables using the name of the original variable and adding the suffix ‘_E#’ to indicate the comparison group used to create the variable, where # is the sequential number of each category. Effect coding is most useful when working with variables that have 3 or more categories.

If you effect code the variable ITBOOK, which has 3 categories (1 for “no books”, 2 for “some books”, and 3 for “lots of books”), and you specify 3 as the reference category, it will create the variables ITBOOK_E1 (coded 1 for those with “no books”, zero for those with “some book” and -1 for those with “lots of books”) and ITBOOK_E2 (coded zero for those with “no books” coded 1 for those with “some books” and -1 for those with “lots of books”). As you see, those with “lots of books” are coded -1 in both effect coded variables.

Cases with the categorical variable coded as missing are excluded from the analysis. Check the section “Check Coding of Coded Variables” in the SPSS output window to verify proper coding of the categorical variable has been achieved.

When an effect coded variable is used in a regression, the intercept or constant is the mean of the group means, and the slope or regression coefficients are the difference between the mean of the group means, and the group represented by the effect coded variable. Since the regression coefficients are presented with a standard error and a t value, these can be used to test whether a difference between means is statistically significant.

Other Types of Coding

While the linear regression module of the IEA IDB Analyzer can only create dummy or effect coded variables, you can always create additional contrast coded variables and use in the analysis. If you are not familiar with the use of contrast coded variables in regression, we suggest you become familiar with the use of these variables before using them.

Appendix D: Working with TALIS Data from across ISCED Levels

In 2013, TALIS administered the instruments to several ISCED levels. The IEA IDB Analyzer is currently not capable of merging TALIS data from across ISCED levels. To create a dataset from multiple ISCED levels, you will need to follow these steps:

1. Create the corresponding dataset for each ISCED level using the Merge Module. You should save these datasets using a unique name for each of them.
2. Append the datasets from the different ISCED levels.
3. Save the combined dataset with a new name.
4. Open the Analysis Module of the IEA IDB Analyzer and select the file that contains the combined datasets.
5. Under Analysis Type, select the option: 'TALIS 2013 Multiple Level/Population'. This will force the use of IDCNTPOP as the grouping variable which will produce results by ISCED level within country.

Appendix E: Calculating Standard Errors in TIMSS and PIRLS

Since the inception of TIMSS and PIRLS, the sampling variance has been computed using what is described as the “shortcut” method. As of calendar year 2016, all official TIMSS and PIRLS publications will be using the full method for computing the sampling variances and the IEA IDB Analyzer will be computing these consistent with these publications.

In the case of TIMSS and PIRLS, when using the full method for computing the sampling variance, we assume there are two clusters within each of 75 strata. For each stratum, we create 2 sets of replicate weights: the first set doubles the contribution of the first cluster and zeroes out the contribution of the second cluster. The second set does the reverse: it zeroes out the contribution of the first cluster and doubles the contribution of the second cluster. The sampling variance is computed summing across the resulting 150 replicates and multiplying the results by $\frac{1}{2}$. Under the shortcut method, we created one replicate weight per stratum where a cluster was selected at random to have its contribution doubled, and the contribution of the remaining cluster was zeroed out. The sampling variance was computed summing across the 75 replicates and multiplying the results by 1.

In addition, when using plausible values, the full method will compute the sampling variance using each of the five plausible values in the analysis and will use and report the average of these. Under the shortcut method the sampling variance for statistics involving plausible values was computed using only the first plausible value.

Should you want to replicate TIMSS and PIRLS variance estimates published prior to calendar year 2016, using the shortcut method, you will need to make the following changes to the code generated by the IEA IDB Analyzer:

- From “JK2TYPE = FULL / ” to “JK2TYPE = HALF/ ”
- From “NRWGT = 150 / ” to “NRWGT = 75/ ”

If you are using plausible values in your analysis, you will also need to make the following change:

- From “SHRTCUT = N / ” to “SHRTCUT = Y / ”.

Appendix F: Analysis Notes for Specific Studies

OECD - PISA

Beginning in 2015, the OECD-PISA will be producing and using 10 plausible values for the analysis of its data. The IEA IDB Analyzer provides the user the option to work with pre-2015 data, using 5 plausible values, or with 2015 data and beyond, using 10 plausible values. The user will need to make the corresponding selection using the “Analysis Type” drop-down menu within the Analysis Module.

U.S. NAEP

The U.S. National Assessment of Educational Progress (U.S.-NAEP) has been conducted since the 1960's and has a long list of published assessment data. Starting in 2013, the U.S.-NAEP started producing and using 20 plausible values for the proficiency scales. Prior to 2013 the U.S.-NAEP produced and used only 5 plausible values. Therefore, when working with the U.S.-NAEP data you will first need to specify whether you are using pre-2013 data or not, and then select the appropriate analysis type using the “Analysis Type” drop-down menu within the Analysis Module. You will notice that some assessment subjects and years have their own entry. Scan the list carefully to make sure to make the correct selection. You will have the following options:

NAEP

- NAEP District
- NAEP National
- NAEP State

NAEP (pre-2013)

- NAEP District (pre-2013)
- NAEP National (Geography & History 2001)
- NAEP National (Math & Science 1996 & 2000 - Grades 4 & 8)
- NAEP National (Math 1996 & 2000, Grade 12)
- NAEP National (pre-2013)
- NAEP National (Reading 1998 & 2000 - Grade 4)
- NAEP National (Reading 1998, Grades 8 & 12)
- NAEP National (Vocabulary 2009)
- NAEP State (Math & Science 2000)
- NAEP State (pre-2013)
- NAEP State (Reading 1998)

Appendix G: Calculating Design Effects⁴²

Surveys in education rarely sample students by simply selecting a random sample of students (known as a simple random sample) from the population. In its simplest form, a sampling design is used where schools (clusters) are first selected and, within each selected school, classes (clusters) or students are randomly selected. Sometimes, geographic areas are first selected before sampling schools and students. The resulting samples are often the result of multi-stage and stratified selection as well.

Students selected from the same school cannot be considered as independent observations as assumed with a simple random sample because they are usually more similar to one another than to students attending other schools.

A simple random sample of 4 000 students is thus likely to cover the diversity of the population better than a sample of 100 schools with 40 students selected from within each school. It follows that the uncertainty associated with any population parameter estimate will be larger for a clustered sample estimate than for a simple random sample estimate of the same size.

In the case of a simple random sample, the variance of a mean estimate $\hat{\mu}$ is calculated as:

$$Var_{SRS}(\hat{\mu}) = \frac{\sigma^2}{n}$$

where σ^2 denotes the variance of the whole student population and n is the student sample size. The variance for the mean from a simple random sample is inversely proportional to the number of selected students.

It is usual to express the effect of the sampling design on the standard errors by a statistic called the „design effect“. The design effect is calculated as the ratio of the variance of the statistic obtained assuming a complex sample design, to the variance of the same estimate assuming a simple random sample selection of the same number of sampling units (Var_{SRS}). The estimate of the variance for the complex sample is typically calculated using a replication procedure like jackknife or balanced repeated replication ($Var_{REP}(t)$).

Therefore, a design effect for a statistic t can be computed:

$$Def f(t) = \frac{Var_{REP}(t)}{Var_{SRS}(t)}$$

where $Var_{REP}(t)$ is the sampling variance for the statistic t computed by the corresponding replication method, and $Var_{SRS}(t)$ is the sampling variance for the same statistic t on the same data but considering the sample as a simple random sample. The IEA's IEA IDB Analyzer calculates this design effect for any non-plausible value.

When computing design effects for statistics that involve plausible values, the notion of design effect as given earlier needs to be extended to incorporate the imputation variance. This gives rise to five possible design effect formulae that can be used to describe the influence of the sampling and test designs on the standard errors for a statistic.

The variance for statistics that involves plausible values consists of two components: sampling variance (Var_{REP}), calculated using the corresponding replication method, and imputation variance (Var_{IMP}), calculated as:

$$Var_{IMP} = \left(1 + \frac{1}{P}\right) * Var(t_p)$$

where P is the number of plausible values, and $Var(t_p)$ is the variance across the P statistics computed using each of the plausible values in the analysis.

The variance of a statistic calculated using plausible values is then calculated as the sum of the sampling and the imputation variances, or $Var_{REP} + Var_{IMP}$.

Given these two components, design effects can be defined and calculated as follows:

⁴² Adapted from „Chapter 12: Sampling Outcomes“, by Christian Monseur, Keith Rust and Sheila Krawchuk; in the *PISA 2000 Technical Report*, by Ray Adams and Margaret Wu (Eds.).

1. Design Effect 1: Shows the inflation of the total variance that would have occurred due to measurement error if in the sample was considered as a simple random sample.

$$DEff_1(t) = \frac{Var_{SRS}(t) + Var_{IMP}(t)}{Var_{SRS}(t)}$$

2. Design Effect 2: Shows the inflation of the total variance due only to the use of the complex sampling design.

$$DEff_2(t) = \frac{Var_{REP}(t) + Var_{IMP}(t)}{Var_{SRS}(t) + Var_{IMP}(t)}$$

3. Design Effect 3: Shows the inflation of the sampling variance due to the use of the complex sample design.

$$DEff_3(t) = \frac{Var_{REP}(t)}{Var_{SRS}(t)}$$

4. Design Effect 4: Shows the inflation of the total variance due to imputation variance.

$$DEff_4(t) = \frac{Var_{REP}(t) + Var_{IMP}(t)}{Var_{REP}(t)}$$

5. Design Effect 5: Shows the inflation of the total variance due to the imputation variance and due to the complex sampling design.

$$DEff_5(t) = \frac{Var_{REP}(t) + Var_{IMP}(t)}{Var_{SRS}(t)}$$

The product of the first and second design effects equals the product of the third and fourth design effects, and both products are equal to the fifth design effect.

The IEA IDB Analyzer computes each of these design effect statistics.

Appendix H: Using the IDB Analyzer with SAS OnDemand for Academics

About SAS OnDemand for Academics

SAS OnDemand for Academics (SAS ODA) provides **free online access to the SAS software**. It is designed as a delivery model for teaching and learning statistical analysis, data mining and forecasting. Whether you are an independent learner, a student, or an instructor, you can access SAS software via the cloud free of charge.

To access SAS ODA, please visit <https://support.sas.com/en/software/ondemand-for-academics-support.html> to sign up and open an account. A frequently asked questions page is currently available from https://support.sas.com/ondemand/caq_new.html.

As a free resource, it does have some limitations. Since SAS ODA supports a large population of SAS users, they focus considerable attention on ensuring the site is available and reliable to all users. To help meet this goal, they rely on users to also be mindful of resources consumed by their jobs, and the resulting impact on their fellow users as well as the site. As a safeguard, resource consumption by user sessions is continuously monitored. This includes, for instance, the number of user processes, duration of the session, processing time consumed, process memory consumed, and the number of bytes transferred between files or network resources. Sessions found to be consuming excessive resources are subject to termination without notice.

SAS recommends avoiding large long-running jobs where possible. Such jobs tend to eventually report high resource usage, and hence are more likely to be terminated. In addition, these are more likely to be terminated due to the duration of the session. Sessions are subject to termination after 6 hours and after 3 hours of inactivity.

All users of SAS ODA have 5 GB of personal data storage. This data storage includes any SAS programs and projects that you create with software such as SAS Studio as well as any personal data that you upload. There are also limits on the size of the files that can be transferred to the cloud, and the overall size of several files combined for each transfer.

The IEA and SAS are not affiliated in any way. The IEA provides these instructions to facilitate using the IDB Analyzer in conjunction with SAS ODA, but the IEA makes no warranties that all functions will work with SAS ODA, or about the availability of SAS ODA to users in the future. SAS ODA is owned and maintained by SAS Institute.

Using the IDB Analyzer with SAS ODA

The IDB Analyzer is an application developed by the International Association for the Evaluation of Educational Achievement (IEA) that can be used to combine and analyze data from all IEA's large-scale assessments, as well as data from most major large-scale assessments surveys. Originally developed for IEA's international large-scale assessments, it has also been configured to work with many national and international large-scale assessment databases. For more information about the functionality and capabilities of the IDB Analyzer, and study compatibility, please refer to the Help Manual available when you install the IDB Analyzer. The IDB Analyzer is available from <https://ww.iea.nl>.

The IDB Analyzer works by reading information from SAS or SPSS databases located in a local computer and providing the user with options to merge or analyze the data. The IDB Analyzer is not able to access data files in the SAS ODA cloud. Therefore, when using the IDB Analyzer in conjunction with SAS ODA, you will need to have mirror copies of the SAS datasets in your local machine, and your SAS ODA directory. You will then use the IDB Analyzer to create SAS syntax that you will need to copy over to SAS ODA, edit, and submit from within your SAS ODA account.

The first time you will be using the IDB Analyzer with SAS ODA, you will need to follow the following steps:

1. **Uninstall the IDB Analyzer.** Even if you have the latest version, we recommend uninstalling it, and installing a fresh new copy of the IDB Analyzer as this will ensure the capitalization of the filenames is what SAS ODA expects to find.

2. Install the IDB Analyzer in your local machine. Make sure it is Version 5.0 or later.
3. In your local machine, create a directory where you will have all your SAS files from the study you want to work with. While any directory will do, we recommend using a simple name. For the purpose of these instructions, we will use "c:\temp\Data".
4. Copy or download all the relevant SAS datasets to "c:\temp\Data", including format libraries.
5. If you do not have a copy of SAS installed in your machine, you will need to install the SAS Local Data Provider application. To do this, go to www.sas.com/apps/demosdownloads/setupintro.jsp and select "SAS Providers for OLE DB". Install the latest and follow the prompts. This will allow the IDB Analyzer to read the variable information within the SAS datasets.
6. Log-in to your SAS ODA account by going to <https://welcome.oda.sas.com>. If you do not have a SAS profile, you will need to create one by selecting the "Don't have a SAS Profile?" option.
7. When you log-in, select the SAS Studio option. It might take a few seconds to get you there, so please be patient. Depending on your browser, you might also need to allow pop-ups for the SAS ODA session to start.
8. Once your SAS ODA starts, create the following 4 folders in your home directory (To create a folder, right-click over 'Files (Home)' and select 'NEW' -> 'Folder'):
 - a. IDB Analyzer Macros
 - b. Data
 - c. Programs
 - d. Work

Note that all file and directory references in SAS ODA are case sensitive, so you will need to be attentive to the capitalization of these.

To get the complete reference name for each folder, right-click over the name, and select 'Properties'. Under the location you will see, for example, "/home/u#####/Data". The "u#####" refers to your unique SAS user ID within the SAS ODA cloud.

9. Upload the SAS datasets you want to work with and corresponding format libraries to the '/Data' directory. Depending on how many files you are uploading, you might have to do this in more than one step to avoid reaching the upload limits. To upload these files, right-click on the name of the folder, select "Upload files..." and follow the prompts. Confirm all the files were copied into the '/Data' directory.
10. You then need to upload the SAS macros installed by the IDB Analyzer. To accomplish this, right-click on the '/IDB Analyzer Macros' folder, select "Upload Files...", and follow the prompts. You will likely find the SAS macros in your local machine in a directory called: C:\Users\\AppData\Roaming\IEA\IDBAnalyzerV5\bin\Data\Templates\SAS_Macros. Confirm all the macros were copied to the '/IDB Analyzer Macros' directory.
11. Check that the filenames are all lowercase. If they are not, you will need to edit the filenames and make them all lowercase.

Once you have completed these steps, you are ready to use the IDB Analyzer to create code usable within SAS ODA.

There are two types of tasks you will want to accomplish using the IDB Analyzer in conjunction with SAS ODA. The first is to merge files, the second is to conduct a selected statistical analysis. We will cover each of these separately.

As a reminder, you will use the IDB Analyzer in your local machine to create SAS syntax, to then copy and paste the syntax in your SAS ODA session and run the syntax from within your SAS ODA session.

Merging Files

You will be merging files if you want to combine data from different countries or combine data from different sources (i.e., combining students with school and teacher data). When merging files, you will be creating syntax with the IDB Analyzer, moving it to the SAS ODA platform, and running the code within the SAS ODA platform.

When merging files, you will need to follow these steps:

1. Start your SAS ODA session and press F4 to open a new program window.
2. Open the IDB Analyzer, select SAS as your statistical software and then select 'Merge Module'
3. Make the corresponding selections in the Merge Module
4. After making your selections, click on the "Start SAS" button. If you do not have SAS installed in your machine, and this is the first time doing this, you will be prompted to select the application to use to open the syntax file. Select Notepad, or any other text editor you are comfortable with.
5. A copy of the syntax is saved to the location you indicated. With the syntax open, select all of it, and copy it.
6. Go to your SAS ODA session and paste the syntax in the Program CODE window.
7. Once copied, edit the following lines using the corresponding directories in your SAS ODA directory structure so they read as follows (the "u#####" corresponds to your SAS user ID)

```
/* This is where the macros are located */
%let mdir = /home/u#####/IDB Analyzer Macros;

/* This is where the data will be read from */
%let idir = /home/u#####/Data;

/* This is where the output will be saved */
%let odir = /home/u#####/Work;
```

8. Save the CODE to your Programs directory.
9. Before running the code, also copy the file "idbanalyzercountries.sas" to the "/home/u#####/Work" directory.
10. Run the code
11. Check your results and make sure there are no errors or warnings.

Once you have merged the file, you will need to move the file created in "home/<u#####>/Work" to your computer so it can be used by the IDB Analyzer.

Conducting Analysis

As when you are merging files, you will be creating syntax in the IDB Analyzer with the files in your computer, moving the syntax to the SAS ODA platform, and running the code from within the SAS ODA platform.

When setting up an analysis, you will need to follow these steps:

1. Start your SAS ODA session and press F4 to open a new program window.
2. Open the IDB Analyzer, select SAS as your statistical software and then select 'Analysis Module'
3. Make the corresponding selections in the Analysis Module

4. After making your selections, click on the "Start SAS" button. If you do not have SAS installed in your machine, and this is the first time doing this, you will be prompted to select the application to use to open the syntax file. Select Notepad, or any other text editor you are comfortable with.
5. A copy of the syntax is saved to the location you indicated. With the syntax open, select all of it, and copy it.
6. Go to your SAS ODA session and paste the syntax in the Program CODE window.
7. Once copied, edit the following lines using the corresponding directories in your SAS ODA directory structure so they read as follows (the "u#####" corresponds to your SAS user ID):

```
/* This is where the macros are located */
%let mdir = /home/u#####/IDB Analyzer Macros;

/* This is where the data will be read from */
%let idir = %let odir = /home/u#####/Work;

/* This is where the output will be saved */
%let odir = %let odir = /home/u#####/Work;
```

8. Save the CODE to your Programs directory.
9. Run the code
10. Check your results

Appendix I: Analysis Settings

The following are the parameter settings used by the IEA's IDB Analyzer for each of the data files that it works with. These correspond to the analysis settings specified in the technical documentation of the studies at the time this was written. For more information about the parameter settings, please consult the corresponding technical documentation of the study.

The key to reading the parameters is the following:

METHOD: This indicates the method used in the calculation of the sampling variance. Options are JK1 for jackknife1, JK2 for jackknife2, BRR for balanced repeated replicates, and PIAAC. When option PIAAC is used, the program looks for the variables VEMETHOD, VENREPS and VEFAYFAC to complete the analysis specifications.

KFAC: The Fay factor to use in the calculation of the sampling variance when METHOD=BRR.

JK2TYPE: When using METHOD=JK2, you have the option of using replicate weights where only one of the two replicates (as specified by JKR) are dropped from within each stratum (JKZ), or that both replicates be dropped.

SHRTCUT: This indicates where to use all plausible values in the calculation of the sampling variance (SHRTCUT=N), or to take a shortcut and use only the first plausible value in the calculation of the sampling variance (SHRTCUT=Y).

WGT: name of the sampling weight variable used for the analysis.

RWGT: Prefix of the replicate weight variables. If replicate weights are calculated on the fly, this entry is left blank. When replicate weights are included in the file, these are expected to be in the format RWGT1 to RWGT#, where # corresponds to the value of NRWGT below.

NPV: The number of plausible values used for the analysis.

NRWGT: The number of replicate weights used to calculate the sampling variance.

Analysis Type	METHOD	KFAC	JK2TYPE	SHRTCUT	WGT	RWGT	NPV	NRWGT
CivED (Using Student Weights)	JRR		HALF	Y	TOTWGT		5	75
CivED (Using Teacher Weights)	JRR		HALF	Y	TCHWGT		5	75
IALS/ALLS (Original data)	JK2		HALF	Y	POPWT	Replic	5	30
IALS/ALLS (Rescaled data - 2013)	JK2		HALF	N	POPWT	Replic	10	30
ICCS (Using School Weights)	JRR		HALF	N	TOTWGT		5	75
ICCS (Using Student Weights)	JRR		HALF	N	TOTWGT		5	75
ICCS (Using Teacher Weights)	JRR		HALF	N	TOTWGT		5	75
ICILS (Using Teacher Weights)	JRR		HALF	N	TOTWGT	TRWGT	5	75
ICILS (Using School Weights)	JRR		HALF	N	TOTWGT	CRWGT	5	75
ICILS (Using Student Weights)	JRR		HALF	N	TOTWGT	SRWGT	5	75
NAEP District	JK2		HALF	Y	ORIGWT	SRWT	20	62
NAEP District (pre-2013)	JK2		HALF	Y	ORIGWT	SRWT	5	62
NAEP National	JK2		HALF	Y	ORIGWT	SRWT	20	62
NAEP National (Geography & History 2001)	JK2		HALF	Y	AORIGWT	ASRWT	5	62
NAEP National (Math & Science 1996 & 2000 - Grades 4 & 8)	JK2		HALF	Y	NORIGWT	NSRWT	5	62
NAEP National (Math 1996 & 2000, Grade 12)	JK2		HALF	Y	AORIGWT	ASRWT	5	62
NAEP National (pre-2013)	JK2		HALF	Y	ORIGWT	SRWT	5	62
NAEP National (Reading 1998 & 2000 - Grade 4)	JK2		HALF	Y	NORIGWT	NSRWT	5	62
NAEP National (Reading 1998, Grades 8 & 12)	JK2		HALF	Y	AORIGWT	ASRWT	5	62
NAEP National (Vocabulary 2009)	JK2		HALF	Y	VORIGWT	VSRWT	5	62
NAEP State	JK2		HALF	Y	ORIGWT	SRWT	20	62
NAEP State (Math & Science 2000)	JK2		HALF	Y	AORIGWT	ASRWT	5	62
NAEP State (pre-2013)	JK2		HALF	Y	ORIGWT	SRWT	5	62
NAEP State (Reading 1998)	JK2		HALF	Y	AORIGWT	ASRWT	5	62
PIAAC (using final full sample weight)	PIAAC		HALF	N	SPFWT	SPFWT	10	80
PIRLS (Using School Weights)	JRR		FULL	N	SCHWGT		5	150
PIRLS (Using Student Weights)	JRR		FULL	N	TOTWGT		5	150
PIRLS (Using Teacher Weights)	JRR		FULL	N	TCHWGT		5	150
PISA (Student level data)	BRR	0.5		N	W_FSTUWT	W_FSTURWT	10	80
PISA pre-2015 (Student level data)	BRR	0.5		N	W_FSTUWT	W_FSTR	5	80
prePIRLS (Using School Weights)	JRR		FULL	N	SCHWGT		5	150
prePIRLS (Using Student Weights)	JRR		FULL	N	TOTWGT		5	150
prePIRLS (Using Teacher Weights)	JRR		FULL	N	TCHWGT		5	150
PRIDI (Using Child Weights)	JRR		HALF		TOTWGT			100
SITES (Using Math Teacher Weights)	JRR		HALF		MTOTWGT			100
SITES (Using School Weights)	JRR		HALF		SCHWGT			100

Analysis Type	METHOD	KFAC	JK2TYPE	SHRTCUT	WGT	RWGT	NPV	NRWGT
SITES (Using Science Teacher Weights)	JRR		HALF		STOTWGT			100
TALIS (Using School Weights)	BRR	0.5			SCHWGT	SRWGT		100
TALIS (Using Teacher Weights)	BRR	0.5			TCHWGT	TRWGT		100
TEDS (Using Educator Weights)	BRR	0.5			FINWGTE	FINRWE		32
TEDS (Using Institution Weights)	BRR	0.5			FINWGTI	FINRWI		32
TEDS (Using Primary Teacher Weights)	BRR	0.5			FINWGTP	FINRWP		32
TEDS (Using Secondary Teacher Weights)	BRR	0.5			FINWGTS	FINRWS		32
TIMSS (Using Math Teacher Weights)	JRR		FULL	N	MATWGT		5	150
TIMSS (Using School Weights)	JRR		FULL	N	SCHWGT		5	150
TIMSS (Using Science Teacher Weights)	JRR		FULL	N	SCIWGT		5	150
TIMSS (Using Student Weights)	JRR		FULL	N	TOTWGT		5	150
TIMSS/PIRLS (Using Math Teacher Weights)	JRR		FULL	N	MATWGT		5	150
TIMSS/PIRLS (Using Reading Teacher Weights)	JRR		FULL	N	REAWGT		5	150
TIMSS/PIRLS (Using School Weights)	JRR		FULL	N	SCHWGT		5	150
TIMSS/PIRLS (Using Science Teacher Weights)	JRR		FULL	N	SCIWGT		5	150
TIMSS/PIRLS (Using Student Weights)	JRR		FULL	N	TOTWGT		5	150
TIMSS Advanced (Using Math Teacher Weights)	JRR		FULL	N	MATWGT		5	150
TIMSS Advanced (Using Physics Teacher Weights)	JRR		FULL	N	PHYWGT		5	150
TIMSS Advanced (Using School Weights)	JRR		FULL	N	SCHWGT		5	150
TIMSS Advanced (Using Student Weights)	JRR		FULL	N	TOTWGT		5	150
TIMSS Numeracy (Using Math Teacher Weights)	JRR		FULL	N	MATWGT		5	150
TIMSS Numeracy (Using School Weights)	JRR		FULL	N	SCHWGT		5	150
TIMSS Numeracy (Using Student Weights)	JRR		FULL	N	TOTWGT		5	150

Appendix J: IEA's IDB Analyzer License Agreement

IEA's IEA IDB Analyzer

NON-COMMERCIAL RESEARCH SOFTWARE LICENSE AGREEMENT

READ BEFORE INSTALLING THE IEA's IEA IDB Analyzer

NOTICE TO USER: THIS IS A CONTRACT. BY CLICKING "I AGREE" AND INSTALLING THIS SOFTWARE YOU ACCEPT ALL THE TERMS AND CONDITIONS OF THIS AGREEMENT. IF YOU DO NOT AGREE WITH THE TERMS AND CONDITIONS SET FORTH BELOW, DO NOT INSTALL THE IEA's IEA IDB Analyzer SOFTWARE AND IMMEDIATELY DELETE ALL COPIES OF IT.

This Software License Agreement (the "Agreement") accompanies IEA's executable version of its proprietary IEA's IEA IDB Analyzer Software, a PC-based statistical software program with accompanying macros ("Licensed Software") and related explanatory materials ("User Documentation") (collectively, the "Licensed Product"). Please read this Agreement carefully. If you accept and agree to the terms of this Agreement, you may install the Licensed Software and use in accordance with the terms and conditions hereunder; if you do not agree to any of the terms or conditions of this Agreement, you are not permitted to use the Licensed Product and should return the Licensed Software and the User Documentation to the address above.

1. Definitions.

"Licensee" means, you, the sole user that is authorized to use the Licensed Product in accordance with this Agreement.

2. License and Scope of Use.

- a) Software License. Upon accepting the terms of this Agreement by clicking "I AGREE" below and the installation of the Licensed Software, IEA grants to Licensee, and Licensee accepts, a limited, revocable, non-exclusive, nontransferable license to, store, load, execute and display (collectively, "Use") the Licensed Product solely at the Installation Site listed below (the "Software License"). Licensee's Use of the Licensed Product shall be solely for individual non-commercial research purposes only, by the Licensee. The Licensed Product may not be used to perform any services for any third party or for any commercial use whatsoever. The Software License shall be effective upon installation of the Licensed Software and shall continue unless terminated pursuant to Section 7. Neither the Software License nor any other provision of this Agreement shall be construed as creating, expressly or by implication, any right or interest to access, use, modify, copy or own the source code for the Licensed Software.
- b) Transfer and Other Restrictions. Except as specifically authorized by another provision of this Agreement, Licensee may not copy, relocate, move, sublicense, rent, loan, lease or otherwise distribute or transfer the Licensed Product without IEA's prior written consent and any attempt to the contrary shall be void and of no legal effect. Further, Licensee agrees not to embed, integrate or bundle the Licensed Software with any other technology, system or software.
- c) Exception. Notwithstanding Section 2(b), Licensee may make one copy of the Licensed Software solely for archival/backup purposes.

3. User Documentation.

IEA shall provide to Licensee, one (1) copy of the User Documentation on computer readable media. Licensee's rights in and to the User Documentation are limited to reproducing the User Documentation solely for the purposes of any Use of the Licensed Software by Licensee.

4. Ownership; Confidential Information.

- a) Licensee acknowledges that the Licensed Product is and shall remain the exclusive property of IEA and that Licensee has no right, title, or interest in or to the Licensed Product except as expressly granted in this Agreement.

- b) Acknowledgement. Licensee acknowledges that the Licensed Product (including the User Documentation, translations, compilations, partial copies and derivative works) and additional confidential and proprietary information received from or on behalf of IEA shall belong exclusively to IEA (“Confidential and Proprietary Information”). Confidential and Proprietary Information does not include information that Licensee can show is already known by Licensee prior to receipt of the Licensed Product in the public domain through no wrongful act of Licensee; or received by Licensee from a third party who was free to disclose such information.
- c) Covenants. With respect to the Confidential and Proprietary Information, and as expressly authorized herein, Licensee shall not use, commercialize or disclose the Confidential and Proprietary Information to any person or entity. Licensee shall not (i) alter or remove from any Licensed Product any proprietary, copyright, trademark or trade secret legend; or (ii) attempt to decompile, decompose, disassemble, reconstruct or reverse engineer the Licensed Product or other Confidential and Proprietary Information (and any information derived in violation of such covenant shall automatically be deemed Confidential and Proprietary Information owned exclusively by IEA); or (iii) use the Licensed Product to develop functionally-similar computer applications or allow, authorize or cause any third party to do so. Licensee shall use at least the same degree of care in safeguarding the Confidential and Proprietary Information as Licensee uses in safeguarding its own confidential information, but in no event less than reasonable due diligence and care.

5. Disclaimer.

LICENSEE ACKNOWLEDGES AND AGREES THAT THE LICENSED PRODUCT IS PROVIDED TO LICENSEE “AS-IS”. IEA EXPRESSLY DISCLAIMS ANY AND ALL WARRANTIES AND REPRESENTATIONS, WHETHER EXPRESS OR IMPLIED, WHETHER ARISING BY STATUTE OR OTHERWISE IN LAW OR FROM A COURSE OF DEALING OR USAGE OF TRADE, INCLUDING BUT NOT LIMITED TO ANY WARRANTY OF ACCURACY, COMPLETENESS, PERFORMANCE, RESULTS, COMPUTER STOPPAGE OR FAILURE, NONINFRINGEMENT, TITLE, AND THE IMPLIED WARRANTIES OF MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE.

6. Limitations of Remedies and Liabilities.

- a) IN NO EVENT SHALL IEA BE RESPONSIBLE OR LIABLE FOR ANY DAMAGES ARISING OUT OF THE USE OF THE LICENSED PRODUCT OR ARISING UNDER THIS AGREEMENT, WHETHER IN TORT, CONTRACT OR OTHERWISE, TO LICENSEE, OR TO ANY OTHER PERSON OR ENTITY.
- b) IEA SHALL NOT BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, CONSEQUENTIAL, PUNITIVE, SPECIAL, OR SIMILAR DAMAGES, INCLUDING WITHOUT LIMITATION, DAMAGES OR LOSS OF PROFITS, LOSS OF GOODWILL, WORK STOPPAGE, COMPUTER/HARDWARE FAILURE OR MALFUNCTION, LOSS OF WORK PRODUCT, OR ANY AND ALL OTHER COMMERCIAL OR NONCOMMERCIAL DAMAGES OR LOSSES WHETHER DIRECTLY OR INDIRECTLY CAUSED, WHETHER IN TORT, CONTRACT, OR OTHERWISE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

7. Term and Termination.

- a) Term. This Agreement shall commence upon installation of the Licensed Software and shall continue unless otherwise terminated in accordance with this Section 7.
- b) Termination. Either party may terminate this Agreement at any time upon notice to the other party.
- c) Effect of Termination. Upon termination of this Agreement: (i) the license to the Licensed Software immediately terminates; and (ii) Licensee shall cease all use of the Licensed Product, and return or destroy, all Confidential and Proprietary Information in its possession or control. Licensee shall certify to IEA in writing, within fifteen (15) days after such termination, that Licensee has complied with the foregoing sentence.

8. Survival.

The provisions contained in Sections 4, 5, 6, 7, and 10 shall survive the termination of this Agreement for any reason in accordance with their respective terms.

9. Export Regulations.

The transfer of technology across national boundaries is regulated by some Governments. Licensee shall not acquire, ship, transport, export, or re-export the Licensed Product, directly or indirectly, into any country in violation of any applicable law nor will Licensee use the Licensed Product for any purpose prohibited by such laws.

10. Miscellaneous.

- a) Severability. If any provision of this Agreement shall be determined to be void, invalid, unenforceable, or illegal for any reason, then the validity and enforceability of all of the remaining provisions hereof shall not be affected thereby. Furthermore, if any particular provision of this Agreement shall be adjudicated to be invalid or unenforceable, then such provision shall be deemed amended by limiting and reducing it so to be as close to the parties' intent while remaining valid and enforceable to the maximum extent compatible with the applicable laws of such jurisdiction, such amendment only to apply with respect to the operation of such amendment only to apply with respect to the operation of such provision in the applicable jurisdiction in which the adjudication is made.
- b) Waiver and Remedies. No waiver by either of the parties hereto of any failure by the other party to keep or perform any covenant or condition of this Agreement shall be deemed a waiver of any preceding or succeeding breach of the same or any other covenant or condition. Except for those remedies denominated as sole and exclusive remedies in this Agreement, the remedies herein provided shall be deemed cumulative, and the exercise of one shall not preclude the exercise of any other remedy nor shall the specifications of remedies herein exclude any rights or remedies at law or in equity which may be available.
- c) Transferability. Licensee shall not assign, transfer or encumber the rights granted under this Agreement, in whole or in part, without obtaining the prior written consent of IEA.
- d) Governing Law. This Agreement shall be construed and interpreted, and its performance shall be governed by the laws of the Netherlands, without regard to conflicts of law principles of any jurisdiction in any country throughout the world.
- e) Acknowledgement in Publications. Licensee shall include acknowledgement of IEA's ownership of the IEA'S IEA IDB ANALYZER Software in all publications.
- f) Amendments. This Agreement may not be modified or amended except in a writing executed by authorized representatives of both parties whereupon such new documentation shall become a supplement to this Agreement.
- g) Entire Agreement. This Agreement contains the entire agreement between the parties with respect to the subject matter hereof, and supersedes all other oral or written representations, statements, promises, agreements and letters or other expressions of intent of any kind with respect to the subject matter hereof between them.

BY INSTALLING THE SOFTWARE AND CLICKING THE "I AGREE" BUTTON BELOW, YOU ACKNOWLEDGE THAT YOU HAVE READ THIS AGREEMENT, UNDERSTAND IT, AND AGREE TO BE BOUND BY ITS TERMS AND CONDITIONS. IF YOU DO NOT AGREE TO THE TERMS AND CONDITIONS OF THIS AGREEMENT, CLICK "I DO NOT AGREE" AND RETURN THE LICENSED SOFTWARE TO THE ADDRESS LISTED ABOVE.