IEA TIMSS 2023

TIMSS 2023

USER GUIDE FOR THE INTERNATIONAL DATABASE

Bethany Fishbein
Milena Taneva
Kamil Kowolik

© IEA, 2025

IEA
TIMSS & PIRLS
BOSTON COLLEGE

Copyright © 2025
International Association for the Evaluation of Educational Achievement (IEA)

TIMSS 2023 User Guide for the International Database
Bethany Fishbein, Milena Taneva, and Kamil Kowolik

Suggested citation:

Fishbein, B., Taneva, M., & Kowolik, K. (2025). *TIMSS 2023 User Guide for the International Database*. Boston College, TIMSS & PIRLS International Study Center. https://timss2023.org/data

Publishers:
TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College, and International Association for the Evaluation of Educational Achievement (IEA)

For more information about TIMSS, contact:
IEA's TIMSS & PIRLS International Study Center
Lynch School of Education and Human Development
Boston College
Chestnut Hill, MA 02467
United States
tel: +1-617-552-1600
e-mail: timssandpirls@bc.edu
timssandpirls.bc.edu

Boston College is an equal opportunity, affirmative action employer.

The authors are grateful to these individuals for their contributions to this User Guide and supplemental database material:
Susan Flicop, Eugenio Gonzalez, Svenja Kalmbach, Jurij Lenar, Katherine Reynolds, Lillian Tyack, Vanisa Vongphanakhone, and Liqun Yin

Contents

Introduction

Overview

This User Guide is a companion to the TIMSS 2023 International Database, presenting example analyses using the data and describing the content and format of the data files and supplemental materials within the database. This User Guide should be used in conjunction with *TIMSS 2023 Technical Report (Methods and Procedures)*, documenting the methods and procedures involved in the development, implementation, and reporting of TIMSS 2023, as well as the *TIMSS 2023 Assessment Frameworks*, describing the foundation for developing the assessments. The results of the TIMSS 2023 assessments are reported in *TIMSS 2023 International Results in Mathematics and Science*. The *TIMSS 2023 Encyclopedia* describes the broader national education contexts in each of the TIMSS 2023 participating countries and benchmarking systems.

TIMSS (the Trends in International Mathematics and Science Study) is an international comparative study of students' mathematics and science achievement. Conducted on a regular four-year cycle since 1995, TIMSS assesses the mathematics and science achievement of fourth- and eighth-grade students and has accumulated 28 years of trend measurements. In addition to student assessment results, TIMSS collects a rich array of information about students' attitudes and experiences as well as about the home, school, and classroom contexts in which they learn mathematics and science, through questionnaires administered to students, their parents, their teachers, and their school principals. Data collected from TIMSS National Research Coordinators highlight the national contexts of participating countries, with a focus on education policies and mathematics and science curricula. These contextual data provide internationally comparable evidence about the educational factors related to mathematics and science achievement. In addition, process data collected through computer-based assessment allows for analyzing students' test navigation behaviors.

The TIMSS 2023 assessment cycle completed the TIMSS transition to computer-based assessment, with all new content developed only in digital format. Nearly all participating countries and benchmarking systems administered TIMSS in the digital format. For a small number of countries, a paper version of the assessment was offered containing only assessment material from TIMSS 2019. At the fourth grade, a less difficult paper assessment option was offered with easier mathematics material for countries where fourth-grade students were still developing fundamental mathematics skills.

Between the previous cycle, TIMSS 2019, and TIMSS 2023, the COVID-19 pandemic caused significant disruptions to education systems worldwide. As a cross-sectional study, TIMSS is not designed to provide estimates of the causal effects of the pandemic on

student learning outcomes. However, the depth and breadth of the data available through TIMSS offers opportunities to examine differences in achievement outcomes and contexts before and after the onset of the pandemic.

The TIMSS 2023 International Database makes the data collected and analyzed by the TIMSS 2023 project available to researchers, analysts, and other users to support and promote secondary analysis aimed at improving mathematics and science education at the fourth- and eighth-grade levels. Across participating countries and benchmarking systems, the database includes response records for over 359,000 fourth-grade students within more than 12,000 schools and nearly 300,000 eighth-grade students within almost 9,000 schools. The national data contained in the TIMSS 2023 International Database has been authorized to be made publicly available by the participating countries.

For countries that participated in previous assessments, TIMSS 2023 provides trends for up to eight cycles—1995, 1999 (eighth grade only), 2003, 2007, 2011, 2015, 2019, and 2023. For all countries participating in TIMSS 2023, the 2023 results can help policymakers and practitioners assess their comparative standing and gauge the rigor and effectiveness of their mathematics and science programs.

About the TIMSS 2023 User Guide

This User Guide describes the content and format of the data and supplemental material in the TIMSS 2023 International Database and presents example analyses with the data. Following this introduction, the User Guide includes the following chapters:

- **Chapter 1** introduces the IEA International Database (IDB) Analyzer Software (IEA, 2025) and presents examples of analyses with the TIMSS 2023 data using this software in conjunction with R/RStudio (R Development Core Team, 2025; Posit Software, 2025).
- **Chapter 2** serves as a reference for details about the structure and contents of the TIMSS 2023 International Database, including detailed descriptions of the various data files, conventions for naming data files and variables, and descriptions of all the supporting documentation provided with the International Database.
- **Chapter 3** describes special R, SPSS, and SAS programs needed to make full use of the TIMSS 2023 International Database, including programs to score the achievement items according to the assigned item response codes.

The primary purpose of this User Guide is to introduce users to the TIMSS 2023 International Database and demonstrate the basic functionality of the IEA IDB Analyzer through simple examples of results published in *TIMSS 2023 International Results in Mathematics and Science*. The IEA IDB Analyzer comes with its own manual, available

through the Help Module, which describes the full functionality and features of the IEA IDB Analyzer. This User Guide also provides references to other TIMSS 2023 publications and documentation to facilitate proper interpretation of data analysis results.

TIMSS 2023 involved complex procedures for drawing student samples, assessing students' achievement, analyzing the data, and reporting the results. To work effectively with TIMSS data, it is important to understand the characteristics of the study. This User Guide should be used in conjunction with the corresponding technical documentation described fully in *TIMSS 2023 Technical Report (Methods and Procedures)*. While the User Guide describes the organization and contents of the database, the chapters of the technical report provide the rationale for the techniques used and for the measures created in the process of data collection and compilation. Throughout this User Guide, references are provided to specific technical report chapters and to other TIMSS 2023 documentation and materials where relevant.

About the TIMSS 2023 International Database

The TIMSS 2023 International Database, including the User Guide and supplemental material, is available on the Boston College, TIMSS & PIRLS International Study Center's website and through IEA's Data Repository. The International Database contains the TIMSS 2023 student achievement data files, process data files, student, home, school, and teacher context data files, curriculum data files, national item selections from the TIMSS 2023 Test-Curriculum Matching Analysis (TCMA), and supplemental materials. Exhibit 1 describes the general structure of the International Database, with a brief description of the support materials available for download. Detailed descriptions of the contents are provided in Chapter 2.

Exhibit 1: Summary of Contents of the TIMSS 2023 International Database

User Guide	TIMSS 2023 User Guide for the International Database
Data Files	
Data (R, SPSS, SAS)	TIMSS 2023 student, process, home, school, and teacher data files
Curriculum Data	TIMSS 2023 Curriculum Questionnaire data Excel files
TCMA Data	TIMSS 2023 Test-Curriculum Matching Analysis data Excel files
Supplemental Material	
Codebooks	List describing all variables in the TIMSS 2023 R, SPSS, and SAS data files
Data Almanacs	Summary statistics for all TIMSS 2023 achievement items and context variables
Achievement Item Information	TIMSS 2023 achievement item information
Achievement IRT Parameters	IRT item parameters for TIMSS 2023 achievement items
Context Questionnaire Variables	International versions of the TIMSS 2023 Context Questionnaires
National Adaptations Database	National adaptations to the TIMSS 2023 Context Questionnaires
Derived Context Variables	Variables derived from the student, home, teacher, and school context data
Special Programs	R, SPSS, and SAS programs to score achievement items

Public-Use and Restricted-Use Versions of the TIMSS 2023 International Database

The TIMSS International Database is available in two versions: a public-use version and a restricted-use version. The public-use version excludes some variables to minimize the risk of disclosing confidential information. The list of variables removed from the public use version is given in Chapter 2 of this User Guide. The public-use version is available for immediate access from the TIMSS 2023 International Database webpage. Regardless of which database is used, users should be able to replicate all published TIMSS 2023 results with this version of the TIMSS 2023 International Database. Users who require any of the removed variables to conduct their analyses should contact IEA through the IEA Study Data Repository to obtain permission and access to the restricted-use version of the TIMSS 2023 International Database.

Acknowledgments

TIMSS 2023 was a collaborative effort involving numerous organizations and individuals around the world, including the TIMSS & PIRLS International Study Center, IEA Amsterdam, IEA Hamburg, RTI International, RM Results, and the National Research Coordinators and their teams in the participating countries. A list of organizations and individuals responsible for TIMSS 2023 is in Appendix F of *TIMSS 2023 International Results in Mathematics and Science*.

References

International Association for the Evaluation of Educational Achievement (IEA). (2025). IEA IDB Analyzer (version 5.0) [Computer software]. Hamburg, Germany: IEA Hamburg. https://www.iea.nl/data

Posit Software, PBC formerly RStudio (2025). RStudio: Integrated Development Environment for R (version 2024.09.1+394). Boston, MA, USA: PBC. https://posit.co/products/open-source/rstudio/

R Development Core Team (2025). R: A language and environment for statistical computing (version 4.4.2). Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org

CHAPTER 1

Analyzing TIMSS 2023 Data with the IEA IDB Analyzer

Overview

This chapter describes the use of the IEA's IDB Analyzer software (IEA, 2025) for analyzing the TIMSS 2023 data, with a focus on replicating results in *TIMSS 2023 International Results in Mathematics and Science*. Used in conjunction with R (R Development Core Team, 2025), IBM SPSS Statistics (IBM Corporation, 2024), or SAS (SAS Institute, 2023), the IEA IDB Analyzer provides a user-friendly interface to easily merge and analyze the various data file types of the TIMSS 2023 International Database. The software seamlessly accounts for the sampling information and the multiple imputed achievement scores to produce accurate statistical results.

Seven example analyses presented in this chapter illustrate some of the capabilities of the IEA IDB Analyzer (Version 5.0), including the computation of means and percentages of students in specified subgroups, mean student achievement in specified subgroups, regression coefficients, and percentages of students reaching the TIMSS 2023 International Benchmarks of Achievement. The examples use student, home, school, and teacher context data files to replicate some of the TIMSS 2023 results.

Users should be able to perform statistical analyses with the IEA IDB Analyzer with a basic knowledge of the TIMSS 2023 International Database. Chapter 2 gives a detailed description of the data files contained in the International Database, including their structure and contents, conventions for naming data files and variables, and descriptions of all the supporting material provided with the International Database.

About the IEA IDB Analyzer

Developed by IEA Hamburg, the IEA IDB Analyzer is a graphical user interface that reads the contents of IEA study data files and generates syntax for R, SPSS, or SAS to combine data files and conduct analysis with these data. The IEA IDB Analyzer enables users to combine data files from IEA's large-scale assessments and conduct analyses without writing programming code. The generated R, SPSS, or SAS syntax accounts for information from the sampling design in the computation of statistics and their standard

errors. In addition, it makes appropriate use of plausible values for calculating estimates of achievement scores and their standard errors, combining both sampling variance and imputation variance. Chapter 13 of *TIMSS 2023 Technical Report (Methods and Procedures)* provides details about estimating standard errors in the TIMSS results.

The IEA IDB Analyzer consists of two main modules—the Merge Module and the Analysis Module. The Merge Module is used to create analysis datasets by combining data files of different types (e.g., student and teacher context data files) and from different countries and selecting subsets of variables for analysis. The Analysis Module provides procedures for computing various statistics and their standard errors. A third module converts SPSS files to R format for merging or analyzing in R.

R and RStudio can be downloaded and installed at no cost from r-project.org and posit.co, respectively (R Development Core Team, 2025; Posit Software, 2025). The latest version of R should be used with the IEA IDB Analyzer. When running a script produced by the IDB Analyzer, RStudio will print in the console a list of necessary packages.

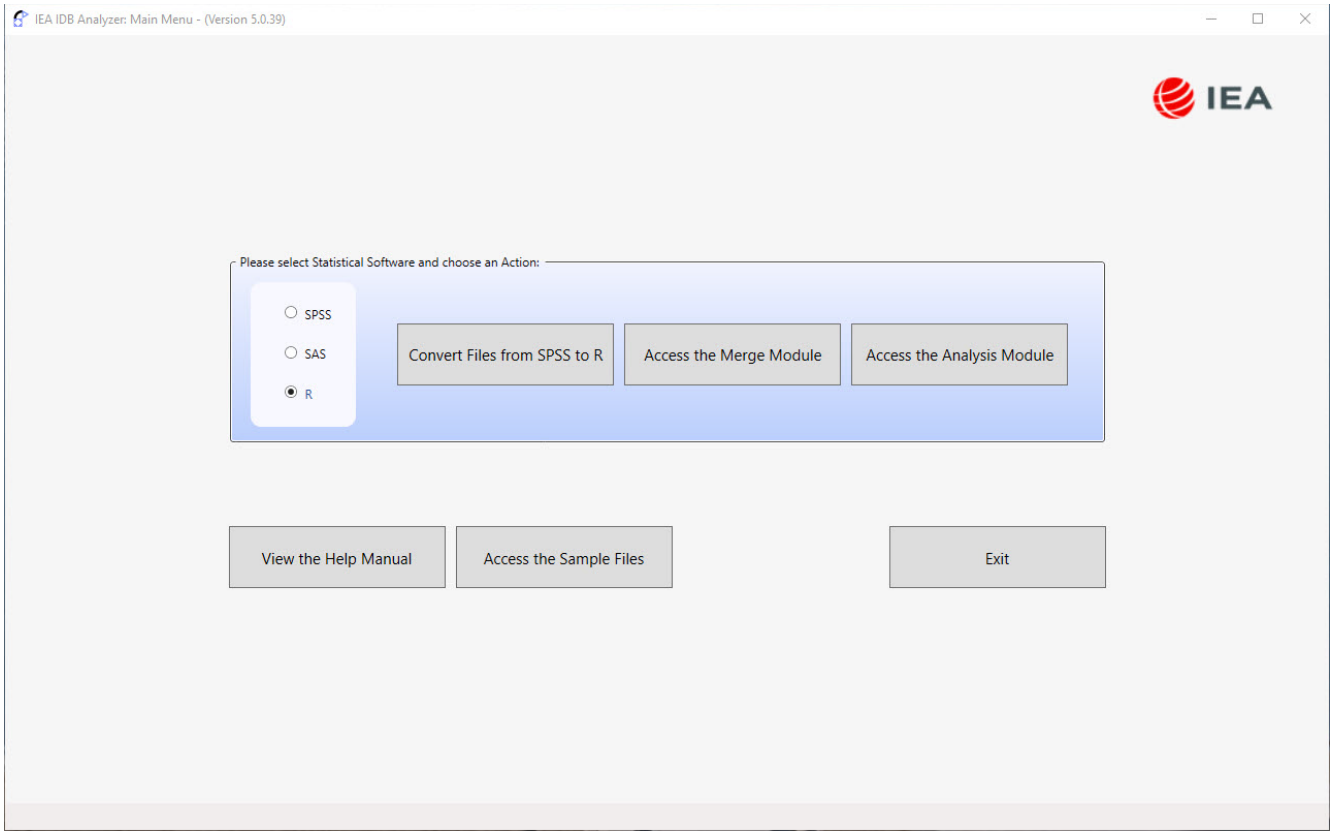
Installing and Launching the IEA IDB Analyzer

The latest version of the IEA IDB Analyzer is available for download from the IEA Data and Tools website. When the IEA IDB Analyzer application is launched, the main window will appear, as shown in Exhibit 1.1. Users are first directed to choose SPSS, SAS, or R as their statistical software of choice. The examples in this chapter primarily use R and RStudio. However, the IDB Analyzer interface is the same for all software options. Differences between the R, SPSS, and SAS outputs are noted where applicable.

The main window presents users with the options to **Convert Files from SPSS to R**, **Access the Merge Module**, **Access the Analysis Module**, **View the Help Manual**, **Access the Sample Files**, or simply **Exit** the application.

The IEA IDB Analyzer has an extensive manual, accessible through the Help button, which users are encouraged to consult for full details on all the functionalities and features of the IEA IDB Analyzer.

Exhibit 1.1: IEA IDB Analyzer Main Window



Merging TIMSS 2023 Data Files with the IEA IDB Analyzer

The IEA IDB Analyzer uses the data files available from the TIMSS & PIRLS International Study Center's TIMSS 2023 International Database webpage and the IEA Data Repository. The TIMSS 2023 data files are disseminated separately by file type (i.e., data source) and by country. In addition to allowing users to combine datasets from more than one country for cross-country analyses, the Merge Module allows for the combination of data from different sources (e.g., student, home, school, and teacher) into one R, SPSS, or SAS dataset for subsequent analysis.

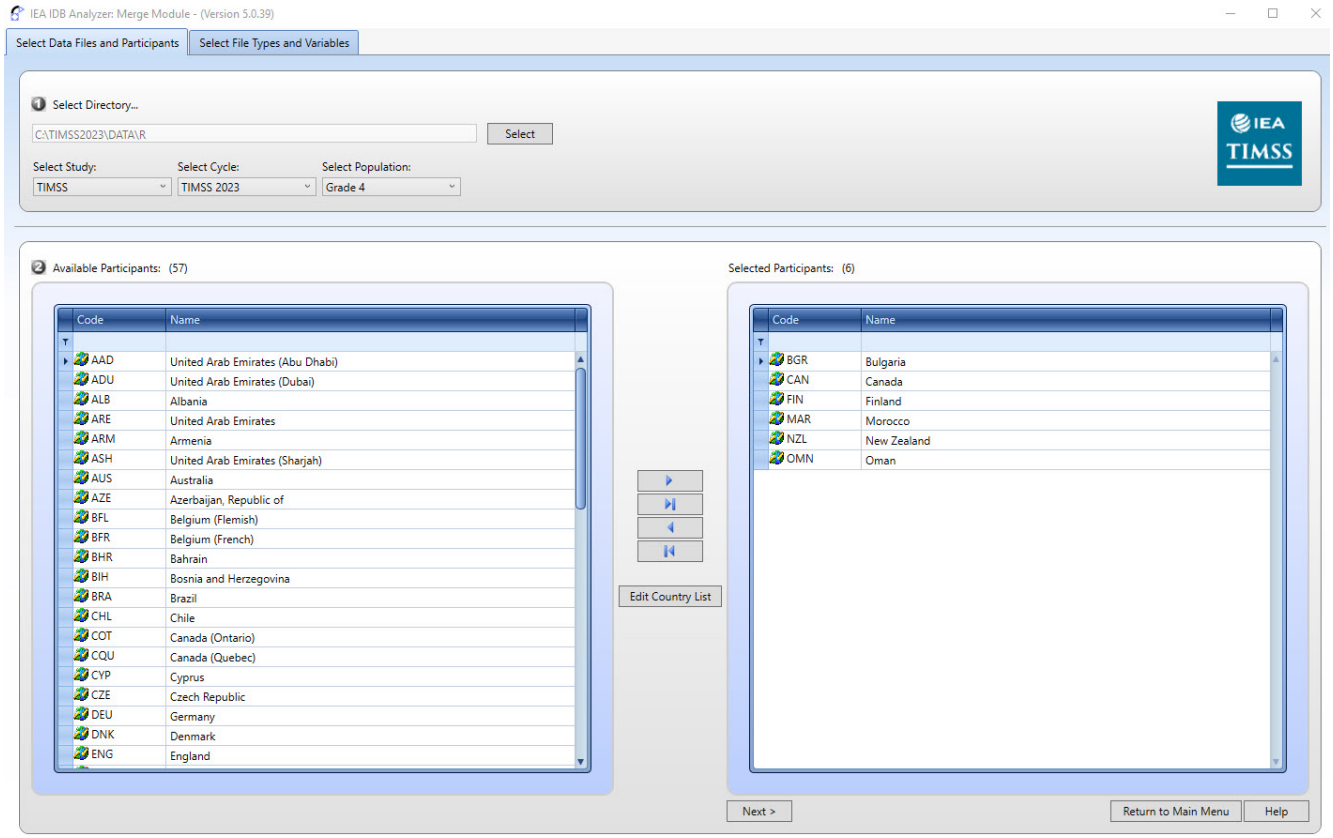
Before doing any statistical analysis with the TIMSS 2023 International Database, users should download and copy all of the data files into a single folder either on their computer or on a server. All files should be within a single folder. For the examples in this chapter, all data files are copied within the folder C:\TIMSS2023\DATA\R.

The following steps will create a data file with data from multiple countries and/or multiple file types:

1. Start the IEA IDB Analyzer and click the **Merge Module** button.
2. Under the **Select Data Files and Participants** tab and in the **Select Directory** field, browse to the folder where all data files are located. For example, in

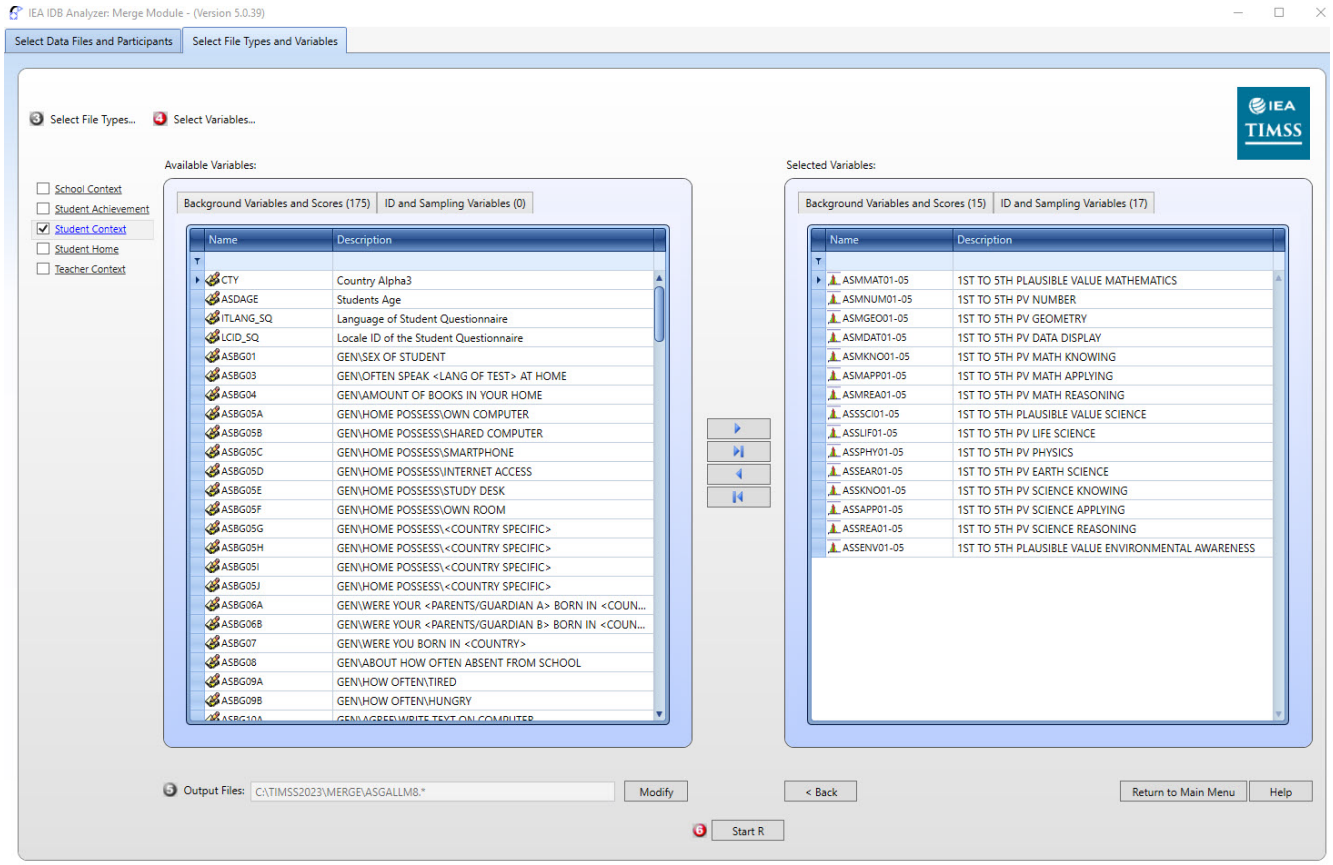
Exhibit 1.2, all R data files are located in the folder C:\TIMSS2023\DATA\R. The program will automatically recognize and complete the **Select Study**, **Select Cycle**, and **Select Population** fields and list all countries available in this folder as possible to merge. If the folder contains data from more than one IEA study (e.g., TIMSS, PIRLS), cycle (e.g., TIMSS 2019, TIMSS 2023), or population (e.g., Grade 4, Grade 8), users should select the desired combination. TIMSS 2023 Grade 4 is selected in Exhibit 1.2.

Exhibit 1.2: IEA IDB Analyzer Merge Module—Select Data Files and Participants



3. Click a country of interest from the **Available Participants** list and click the **right arrow** (▸) button to move it to the **Selected Participants** panel. Individual countries can be moved directly to the **Selected Participants** panel by double-clicking on the row. To select multiple countries, hold the **Ctrl** key on the keyboard when clicking countries. Click the **tab-right arrow** (▸|) button to move all countries to the **Selected Participants** panel. In Exhibit 1.2, Bulgaria, Canada, Finland, Morocco, New Zealand, and Oman are selected.
4. Click the **Next >** button to proceed to the next step. The software will open the **Select File Types and Variables** tab of the Merge Module, as shown in Exhibit 1.3, to select the file types and the variables to be included in the merged data file.

Exhibit 1.3: IEA IDB Analyzer Merge Module—Select File Types and Variables



5. Select the files for merging by checking the appropriate boxes to the left of the window. For example, in Exhibit 1.3, the box next to **Student Background** is checked, indicating the TIMSS student context data files are selected.[1]
6. Select the variables of interest from the **Available Variables** list in the left panel. As described in Chapter 2 of this User Guide, the Codebook files and the Context Questionnaire Variable downloads provide the variable names for storing the data from all questions in the TIMSS 2023 Context Questionnaires. Variables are selected by clicking on them, then moving them to the **Selected Variables** list by clicking the **right arrow** (▶) button. Clicking the **tab-right arrow** (▶|) button selects and moves all variables to the **Selected Variables** list. Note that there are two tabs under the **Selected Variables** list: **Background Variables and Scores** and **ID and Sampling Variables**. All achievement scores and all identification, tracking, and sampling variables are selected by default.
7. Specify the desired name for the merged data file and the folder where it will be stored in the **Output Files** field by clicking the **Define** (or **Modify**) button. The IEA IDB Analyzer will create an R script (*.R), SPSS syntax file (*.SPS), or SAS syntax

[1] The IEA IDB Analyzer uses the term "background" when referring to context data files or variables.

file (*.SAS) of the same name and in the same folder, with the code necessary to perform the merge. In the example shown in Exhibit 1.3, the R script file ASGALLM8.R and the merged data file ASGALLM8.Rdata both will be created and stored in the folder C:\TIMSS2023\MERGE. The merged data file will contain all the variables listed in the **Selected Variables** panel on the right.

8. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** ( ) button (or select **Submit** in the **Run** menu).

Once R, SPSS, or SAS has completed its execution, it is important to check the software output window or log file for possible warnings. If warnings appear, they should be examined carefully, as they might indicate that the merge process was not performed properly and that the resulting merged data file might not be as expected.[2]

Merging Student and Home Context Data Files

The parents of the fourth-grade students participating in TIMSS 2023 responded to the home questionnaire. Their responses are included in the fourth-grade home context data files (named beginning with "ASH"). Although home context variables are located in their own files, they are in essence attributes of the students and must be analyzed in the same manner as student context variables. This will require users to merge the home context data files with the student context data files by selecting both the **Home Background** and **Student Background** file types in the Merge Module of the IEA IDB Analyzer. This is an important step to ensure the proper weights and achievement variables are included for analyses. Variables of interest to be included in the merged data file are chosen separately by file type, as described step-by-step below in the section on Merging Student and Teacher Context Data Files.

Merging Student and School Context Data Files

Because TIMSS includes representative samples of schools, it is possible to compute appropriate statistics with schools as units of analysis. However, the school samples were designed to optimize the student samples and the student-level results. For this reason, it is preferable to analyze school context variables as attributes of the students by merging the school information with the student records.

To merge the student and school context data files, select both the **Student Background** and **School Background** file types in the **Select File Types and Variables**

[2] For more information on how to use the IEA IDB Analyzer, and for troubleshooting, users should consult the Help manual.

tab of the IEA IDB Analyzer Merge Module. This is an important step to ensure the student weights and achievement variables are included for analyses. The variables of interest to be included in the merged data file are selected separately by file type, as described step-by-step below in the next section.

Merging Student and Teacher Context Data Files

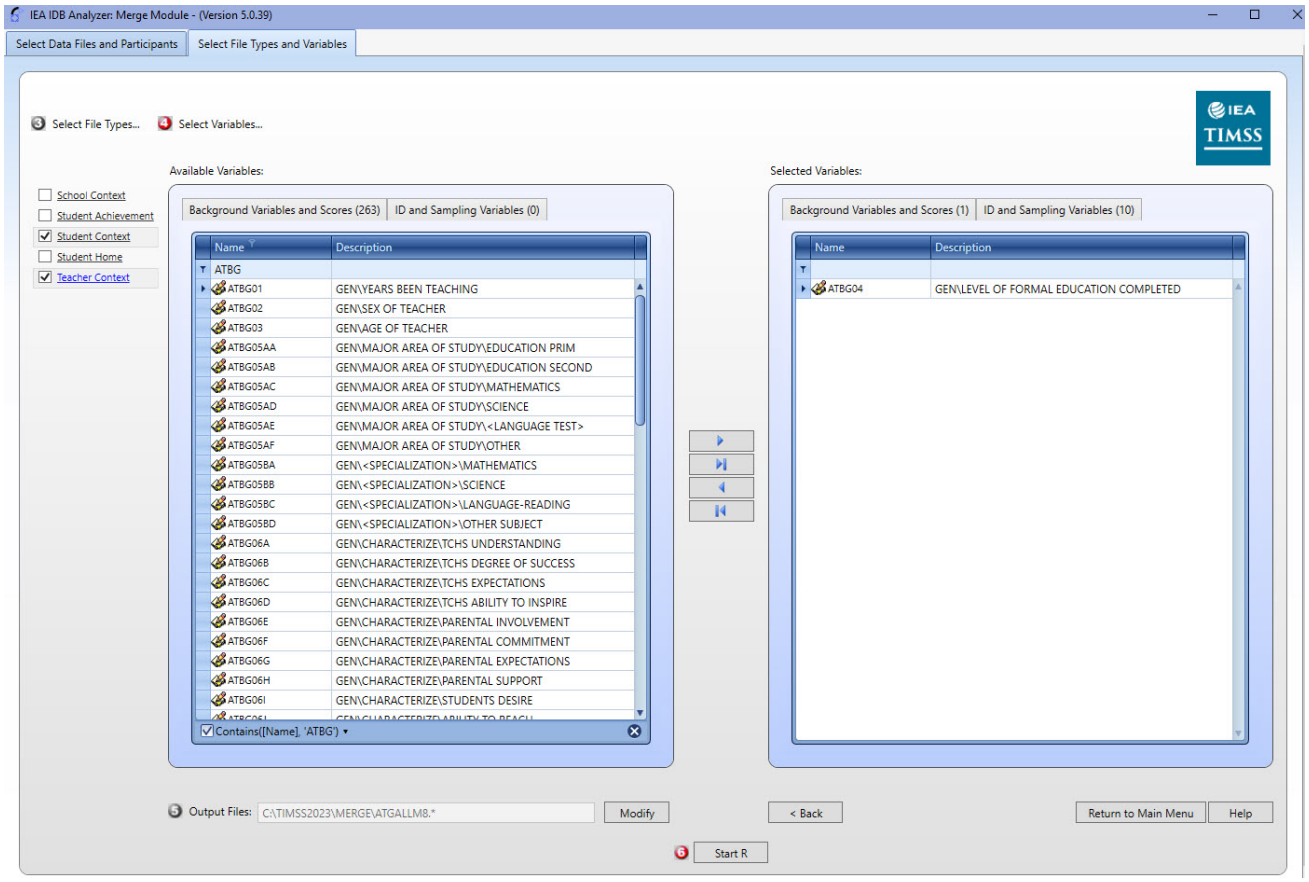
The teachers in the TIMSS 2023 sample do not constitute representative samples of teachers in the participating countries. Rather, they are the teachers of nationally representative samples of students. Therefore, analyses with teacher data should always be made with students as the units of analysis and reported in terms of students who are taught by teachers with a particular attribute.

TIMSS teacher context data are analyzed by linking the student records to their teachers by using the student-teacher linkage data files. The IEA IDB Analyzer does this automatically. To merge teacher data for analysis, it is sufficient to select the **Teacher Background** file type in the **Select File Types and Variables** tab of the IEA IDB Analyzer Merge Module. To analyze student and teacher context data simultaneously, however, both the **Student Background** and **Teacher Background** file types must be selected in the **Select File Types and Variables** tab (see Exhibit 1.4).[3] The variables of interest need to be selected separately for both file types, as follows:

1. Click the checkbox next to the **Student Background** file type so that it appears checked and highlighted. The **Background Variables and Scores** listed in the left-hand **Available Variables** panel will list all variables from the student context data files. This is an important step to ensure the proper weights and achievement variables are included for analyses.
2. By default, all student achievement plausible values, identification, and tracking variables are added to the data file. Select any additional student variables of interest from the left panel and click the **right arrow** (▸) button to move these variables to the **Selected Variables** panel on the right. Click the **tab-right arrow** (▸|) button to select all available variables.
3. Click the checkbox next to the **Teacher Background** file type, and select the variables of interest from the **Background Variables and Scores** panel on the left in the same manner as in Steps 1 and 2, as shown in Exhibit 1.4.

[3] When working with eighth-grade teacher data, users can choose either mathematics teachers or science teachers, but not both.

Exhibit 1.4: IEA IDB Analyzer Merge Module—Select File Types and Variables for Merging Student and Teacher Context Data



4. Specify the desired name for the merged data file and the folder where it will be stored in the **Output Files** field by clicking the **Define/Modify** button. The IEA IDB Analyzer also will create an R script (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the merge. In the example shown in Exhibit 1.4, the R script file ATGALLM8.R and the merged data file ATGALLM8.Rdata both will be created and stored in the folder C:\TIMSS2023\MERGE. The merged data file will contain all the variables listed in the **Selected Variables** panel on the right.
5. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** ( ) button (or select **Submit** in the **Run** menu).

Steps 1 and 2 above are required only if student data (achievement or context) and teacher context data are analyzed simultaneously. It is not recommended to combine both types of files and analyze only student data, as the results may not be correct. Because of

the way in which weights are allocated to teachers, a file with teacher data should only be used to analyze teacher variables, with or without student or school level variables. For analyses that do not make use of teacher variables, you should not use a file with teachers merged into it.

Merged Data Files for the User Guide Examples

To conduct the analysis examples presented in this chapter, a number of merged data files were created following the instructions provided above. Because the examples presented in this User Guide are all about TIMSS 2023 Grade 4, merged data files were produced with countries that participated in TIMSS 2023 at the fourth grade. A full list of countries included in the TIMSS 2023 International Database is provided in Chapter 2 (↓ Exhibit 2.2).

The following merged data files were created with all available context variables and achievement scores selected for the example countries:

ASGALLM8	Merged student context data files
ASHALLM8	Merged home and student context data files
ACGALLM8	Merged school and student context data files
ATGALLM8	Merged teacher and student context data files

Conducting Analyses with the IEA IDB Analyzer

This chapter presents examples of actual analyses used to produce exhibits in the *TIMSS 2023 International Results in Mathematics and Science* report. The IEA IDB Analyzer can create syntax for statistical analyses on any files created using the Merge Module. The **Analysis Module** of the IEA IDB Analyzer allows users to specify the type of analysis and select variables from a merged data file as analysis variables. To conduct analyses using plausible values (PVs) for mathematics or science achievement, after selecting a **Statistic Type**, users should select the **Use PVs** option from the **Plausible Value Option** drop-down menu.

All statistical procedures offered in the Analysis Module of the IEA IDB Analyzer make appropriate use of sampling weights, and standard errors are computed using the jackknife repeated replication (JRR) method (see Chapter 13 of *TIMSS 2023 Technical Report*). When PVs are used, the analyses are performed five times (once for each plausible value) and the results are aggregated to produce accurate estimates of achievement and standard errors that incorporate both sampling and imputation errors.

Statistical Procedures in the IEA IDB Analyzer

Many types of analyses can be conducted using the data files from the TIMSS 2023 International Database with the Analysis Module of the IEA IDB Analyzer. The following statistical procedures can be used to produce exhibits in the *TIMSS 2023 International Results in Mathematics and Science* report, including percentages and means, linear regression, and International Benchmark analyses. See the IEA IDB Analyzer Help manual for a full list of statistical procedures.

Percentages and Means
Compute percentages, means, and standard deviations for selected analysis variables by subgroups defined by grouping variable(s). Plausible values can be included as analysis variables. This procedure is used in Examples 1, 2, 5, 6, and 7 of this chapter.

Percentages Only
Compute percentages by subgroups defined by grouping variable(s).

Linear Regression
Compute linear regression coefficients for selected independent variables to predict a continuous dependent variable by subgroups defined by grouping variable(s). Plausible values can be included as dependent or independent variables. This procedure is used in Example 3 of this chapter.

Logistic Regression
Compute logistic regression coefficients for selected independent variables to predict a dichotomous dependent variable by subgroups defined by grouping variable(s). Plausible values can be included as dependent or independent variables. When used as a dependent variable, plausible values will be dichotomized using a specified cutpoint, such as one of the TIMSS International Benchmarks. This procedure is available only for use with SPSS or SAS.

Correlations
Compute means, standard deviations, and correlation coefficients for selected analysis variables by subgroups defined by grouping variable(s). Plausible values can be included as analysis variables.

Benchmarks
Compute percentages of students meeting a set of user-specified achievement benchmarks, in particular the TIMSS International Benchmarks, by subgroups defined by grouping variable(s). This procedure is used in Example 4 of this chapter.

Percentiles
Compute the score points that separate a given proportion of the distribution of a continuous analysis variable by subgroups defined by the grouping variable(s). Plausible values can be included as analysis variables.

Definitions of Analysis Variables in the IEA IDB Analyzer

The various variables required to conduct an analysis are input into specific variable fields according to their purpose. All available features of the IEA IDB Analyzer are described extensively in its Help manual.

Grouping Variables

This is a list of variables to define subgroups of interest. The list must consist of at least one grouping variable. By default, the IEA IDB Analyzer includes the variable IDCNTRY used to distinguish the participating countries. Additional variables can be selected from the available list. If the **Exclude Missing From Analysis** option is checked, only cases that have non-missing values in the grouping variables will be used in the analysis. If it is not checked, missing values become reporting categories.

Analysis Variables

This is a list of variables for which means, percentages, correlations, or percentiles are to be computed. Usually, more than one analysis variable can be selected. To compute statistics based on achievement scores, after choosing the **Statistic Type**, it is necessary to select the **Use PVs** option in the **Plausible Value Option** drop-down menu and select the achievement scores of interest in the **Plausible Values** field.

Plausible Values (PVs)

This section is used to identify the set of plausible values to be used when achievement scores are the analysis variable for computing statistics. After choosing the **Statistic Type**, select the **Use PVs** option in the **Plausible Value Option** drop-down menu before specifying the achievement scores of interest in the **Plausible Values** field.

Independent Variables

This is a list of variables to be treated as independent variables for a linear or logistic regression analysis. More than one independent variable can be selected. Categorical variables and continuous variables can be specified as independent variables. When specifying categorical variables as independent variables, they can be treated either as “effect coding” or “dummy coding” using the **Contrast** drop-down menu (dummy coding is used in Example 3). Achievement scores also can be included as an independent variable. To specify achievement scores as an independent variable, it is necessary to select the **Use PVs** option in the **Plausible Value Option** drop-down menu and select the achievement scores of interest in the **Plausible Values** field.

Dependent Variable

This is the variable to be used as the dependent variable when a linear or logistic regression analysis is specified. Only one dependent variable can be listed and can be either a context variable or achievement variables (PVs). To use achievement as the dependent variable, select the **Use PVs** option in the **Plausible Value Option** drop-down

menu, click on the **Plausible Values** radio button in the **Dependent Variable** section, and select the achievement scale of interest in the **Plausible Values** field.

Achievement Benchmarks
These are the values that will be used as cutpoints on an achievement scale, selected in the **Plausible Values** section, for computing the percentages of students meeting the specified benchmarks. Multiple cutpoints can be specified, each separated by a blank space. A drop-down menu is available to select the four TIMSS International Benchmarks.

Percentiles
These are the percentiles that will be calculated from the distribution of a continuous analysis variable selected in the **Analysis Variables** section. Achievement PVs can be selected as analysis variables. Select the **Use PVs** option in the **Plausible Value Option** drop-down menu and select the achievement scale of interest in the **Plausible Values** field. Multiple percentiles can be specified, each separated by a blank space.

Weight Variable
This is the sampling weight variable that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight variable for analysis based on the file types included in the merged data file. Generally, this will be TOTWGT, but SENWGT and HOUWGT also are available for student-level analyses with student, home, or school context data. For analyzing teacher data, MATWGT should be used to analyze mathematics teacher data and SCIWGT should be used to analyze science teacher data. The Sampling and Weighting Variables section in Chapter 2 of this User Guide provides more information on the TIMSS sampling weights.

Conducting Analyses with TIMSS Student Achievement and Context Data

The examples in this section use the merged TIMSS 2023 fourth-grade student context data file ASGALLM8 described earlier in the section on Merging Data Files with the IEA IDB Analyzer. Example 1 computes average achievement by country, whereas Example 2 computes national average achievement separately for girls and boys. Example 3 expands on the second example by performing a test of statistical significance on the gender difference using linear regression. Lastly, Example 4 computes the percentages of fourth-grade students reaching each of the TIMSS 2023 International Benchmarks of Mathematics Achievement.

Example 1—Analysis of Average Achievement

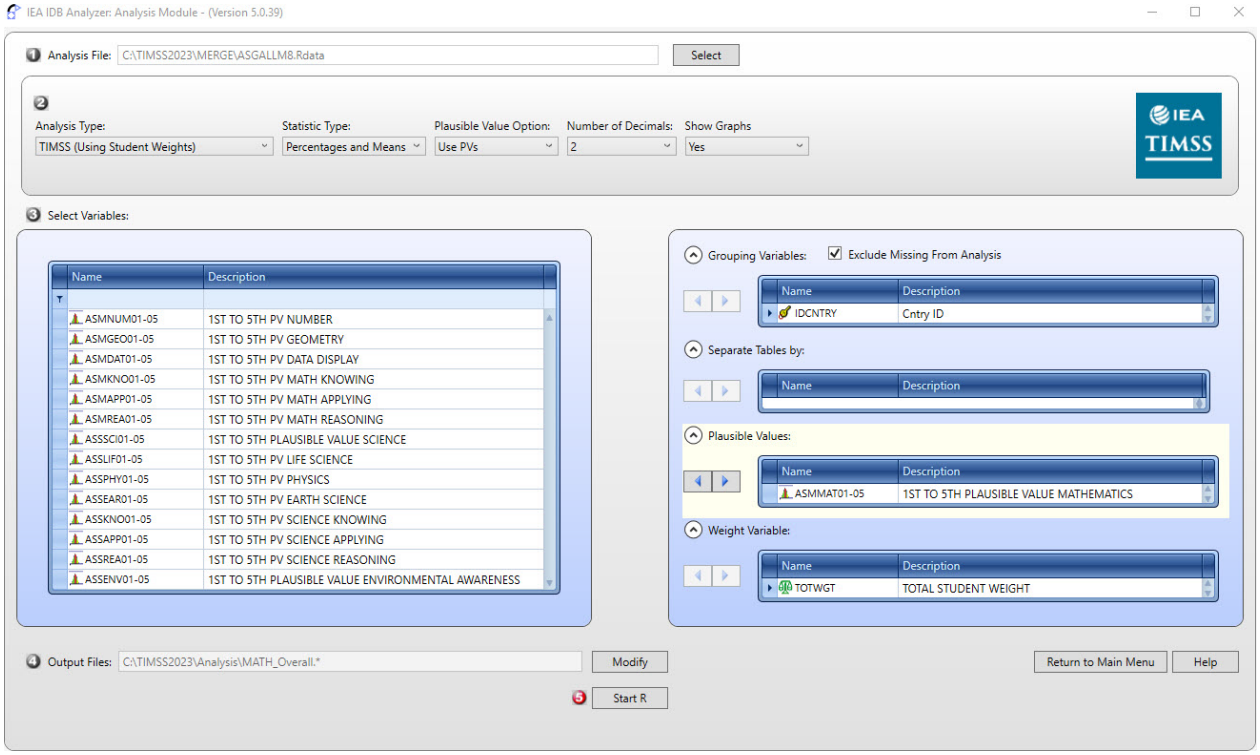
This first example replicates the analysis of the overall distribution of mathematics achievement, presented in Exhibit 1.1.1 of *TIMSS 2023 International Results in Mathematics and Science,* repeated below in Exhibit 1.5.

↓ Exhibit 1.5: International Results Exhibit of Example 1—Analysis of Average Achievement

Because the results in Exhibit 1.5 are based on plausible values, users should make sure they are included as selected variables when creating the file using the Merge Module and also indicate that the analysis will make use of achievement scores with the **Use PVs** option. The **Percentages and Means** statistic type with the **Use PVs** option selected will compute percentages and average achievement scores based on plausible values and their respective standard errors.

After creating the merged data file ASGALLM8, the **Analysis Module** of the IEA IDB Analyzer is used to conduct the analysis in the following steps. The completed Analysis Module for this example is shown in Exhibit 1.6.

Exhibit 1.6: IEA IDB Analyzer Analysis Module Setup for Example 1—Analysis of Average Achievement



1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ASGALLM8 as the **Analysis File** by clicking the **Select** button.
3. Select **TIMSS (Using Student Weights)** as the **Analysis Type**.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**.

6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce two graphs in the output file: one graph showing average achievement by country (bar graph in R and SPSS; line graph in SAS), and one bar graph for the weighted percentage of the total students in each country.
8. The IDB Analyzer automatically selects the variable IDCNTRY for the **Grouping Variables**. No additional grouping variables are needed for this analysis. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (non-achievement) variables. See the IEA IDB Analyzer Help manual for more information.
10. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASMMAT01–05 from the list of available variables in the left panel and move it to the right **Plausible Values** field by clicking the **right arrow** (▸) button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis uses student data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define** (or **Modify**) button in the **Output Files** field. The IEA IDB Analyzer also will create an R script (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the analysis. In Exhibit 1.6, the syntax file MATH_Overall.R and the output files with the same name will be created and stored in the C:\TIMSS2023\Analysis folder.
13. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** (🏃) button (or select **Submit** in the **Run** menu).

The IDB Analyzer produces and saves the results output in three file formats within the same folder specified in Step 12—an HTML output file (or output in SPSS/SAS), R data file (*.Rdata), and Microsoft Excel Worksheet (*.xlsx). The output files are named using the same name specified for the syntax file in Step 12. The HTML reports produced by R are named with the suffix "_ASMMAT" indicating the outcome variable. Graphs are included

only in the HTML (or SPSS/SAS) output files. The IEA IDB Analyzer produces an additional results file in Rdata and xlsx formats, named with the suffix "_sig," that indicates the significance of differences in the outcome variable (achievement) by the grouping variable (IDCNTRY). For this example, the "_sig" output indicates the significance of the differences in achievement between each possible combination of two countries.

Exhibit 1.7 displays the results in the R output with six example countries: Bulgaria, Canada, Finland, Morocco, Oman, and New Zealand. The results are presented in the "Report" section of the HTML output produced by R.

Exhibit 1.7: R Output for Example 1—Analysis of Average Achievement

Report

Analysis for ASMMAT by IDCNTRY

Cntry ID	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASMMAT (Mean)	ASMMAT (s.e.)	Confidence Interval (95)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	4103	54771	617.87	4.34	0.08	529.90	3.60	523 to 537	90.94	3.11	0.00	77
Canada	13716	272593	5019.27	21.61	0.46	503.81	2.01	500 to 508	80.87	1.10	0.00	125
Finland	5803	61145	1215.94	4.85	0.11	529.26	2.50	524 to 534	79.60	1.27	0.00	87
Morocco	7846	735023	20177.83	58.28	0.69	393.25	4.60	384 to 402	102.47	2.45	0.00	125
Oman	8068	78476	690.83	6.22	0.11	421.36	4.03	413 to 429	98.58	2.31	0.00	119
New Zealand	4947	59188	886.95	4.69	0.10	490.28	2.63	485 to 495	90.74	1.20	0.00	78

Each country's results are presented on a single line, with countries ordered sequentially according to their numeric ISO code (↓ Exhibit 2.2). Results for "Table Average" may be produced (not shown), based on all countries included in the data file, including any benchmarking participants. The countries are identified in the first column (Cntry ID) and the second column reports the number of valid cases (N of Cases). The third column reports the sum of weights of the sampled students (Sum of TOTWGT), indicating the estimated total fourth-grade population. The fourth column is the standard error of the sum of weights (Sum of TOTWGT (s.e.)). The next two columns report the weighted percentage of students by the grouping variable (Percent), which for this analysis is the percentage of all students in each country out of the total, and its standard error (Percent (s.e.)). The next two columns report the estimated average for the outcome variable, in this case mathematics achievement (ASMMAT (Mean)) and its standard error (ASMMAT (s.e.)) ("mnpv" and "mnpv_se" in Excel). The "Confidence Interval (95)" column reports the 95% confidence interval for the mean. The confidence interval provides a score range around the estimated mean that illustrates the uncertainty in this estimate, which is based on a sample of students in a selection or subset of classrooms within a

sample of schools in each country. Intervals constructed in this way around the estimated average have a 95% chance to include the true average of the country's achievement.

The standard deviation of the achievement scores (Std.Dev.) and its standard error (Std.Dev. (s.e.)) are reported in the next two columns. The last two columns report the percentage of cases with missing data (Percent Missing) and the number of jackknife zones used for computing standard errors (Number of Variance Strata), respectively (see Chapter 13 of *TIMSS 2023 Technical Report* for information about jackknifing).

Example 2—Analysis of Average Achievement for Girls and Boys

The second example using student context data replicates another set of results presented in the *TIMSS 2023 International Results in Mathematics and Science* report. This analysis investigates the differences in average mathematics achievement between fourth-grade girls and boys. These results, presented in Exhibit 1.1.2 of *TIMSS 2023 International Results in Mathematics and Science*, are repeated below in Exhibit 1.8.

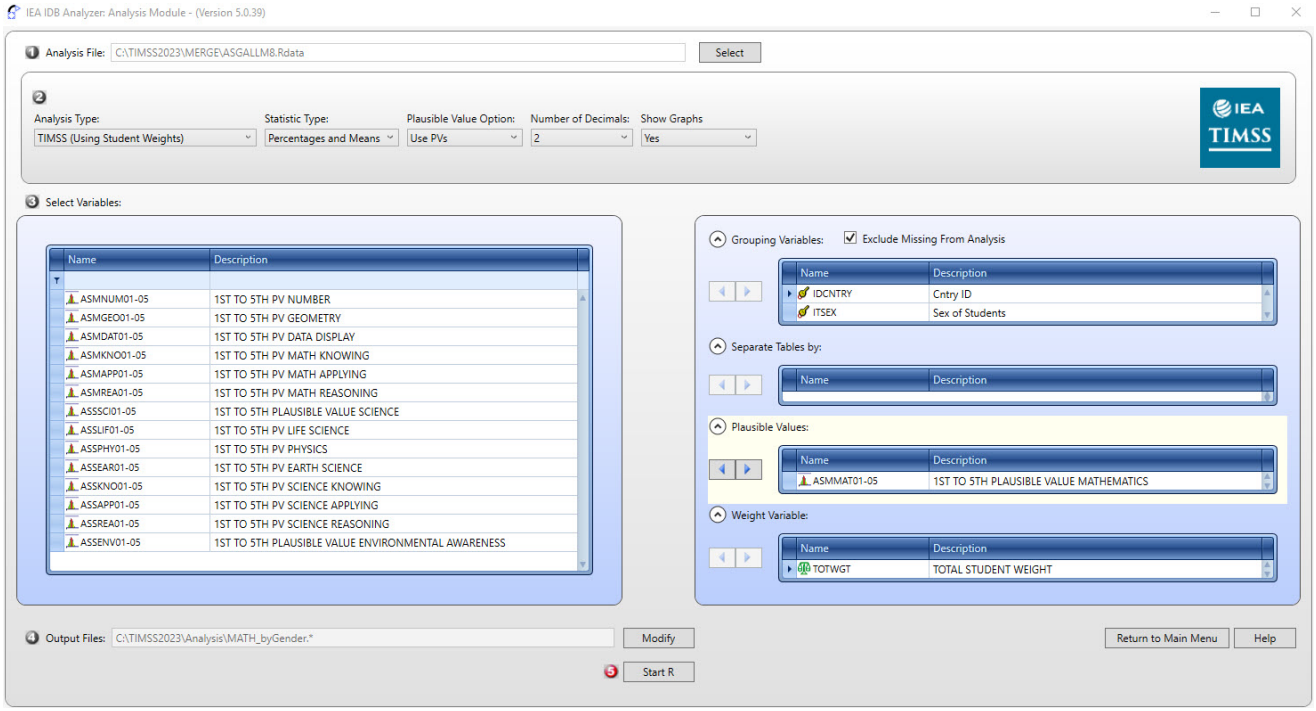
↓ Exhibit 1.8: International Results Exhibit of Example 2—Analysis of Average Achievement for Girls and Boys

The results of this analysis are based on characteristics of students. In general, before conducting analyses using TIMSS contextual variables, users should refer to the relevant codebook for the data file to identify the appropriate variables and understand the coding scheme. From the TIMSS 2023 International Database webpage, the Context Questionnaire Variables downloads present all the context questionnaires administered in TIMSS 2023 and the associated variable names under which the data are saved. The National Adaptations Database downloads should also be checked for any national adaptations made to the questionnaire items that may impact international comparability.

The codebook for student context data files indicates that the tracking variable ITSEX contains categorical information on the gender of students (and is typically preferred for analysis instead of the student-reported questionnaire variable). The **Percentages and Means** statistic type and the **Use PVs** plausible value option will allow us to compute the percentages of students in each gender group and their average achievement based on plausible values and their respective standard errors.

The **Analysis Module** of the IEA IDB Analyzer is used to conduct this analysis for six example countries using the following steps. Exhibit 1.9 presents the completed Analysis Module for this example.

Exhibit 1.9: IEA IDB Analyzer Analysis Module Setup for Example 2—Analysis of Average Achievement for Girls and Boys



1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ASGALLM8 as the **Analysis File** by clicking the **Select** button.
3. Select **TIMSS (Using Student Weights)** as the **Analysis Type**.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce three graphs in the output file: a line graph of the average achievement for each gender by country, a clustered bar graph of average achievement for each gender by country, and a stacked bar graph of average percent of students for each gender by country. R also produces graphs separately for each country.
8. Specify the variable ITSEX as a second grouping variable by first clicking the **Grouping Variables** field to activate it. Then, select ITSEX from the list of variables in the left panel, and move it to the **Grouping Variables** field by clicking the **right arrow** (▸) button**.** The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.

9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (not achievement) variables. See the Help manual for more information.
10. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASMMAT01–05 from the list of available variables in the left panel, and move it to the right **Plausible Values** field by clicking the **right arrow** (▸) button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis uses student context data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer also will create an R script (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the analysis. In Exhibit 1.9, the syntax file MATH_byGender.R and the output files with the same name will be created and stored in the C:\TIMSS2023\Analysis folder.
13. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** (🏃) button (or select **Submit** in the **Run** menu).

The IDB Analyzer produces and saves the results output in three file formats within the folder specified in Step 12—an HTML output file (or output in SPSS/SAS), R data file (*.Rdata), and Microsoft Excel Worksheet (*.xlsx). The output files are named using the same name specified for the syntax file in Step 12. The HTML reports produced by R are named with the suffix "_ASMMAT" indicating the outcome variable. Graphs are included only in the HTML output files. For the Percentages and Means statistic using a second grouping variable (i.e., in addition to IDCNTRY), the IEA IDB Analyzer produces two additional results files in Rdata and xlsx formats. The output file named with the suffix "_sig" reports the significance of the differences between analysis groups—in this case girls and boys—for each country. The output file named with the suffix "_sig2" reports the significance of differences between countries within each of the gender groups.

The results of Example 2 as shown in the R output file are presented in Exhibit 1.10 with example countries Bulgaria, Canada, Finland, Morocco, New Zealand, and Oman. The results are presented in the "Report" section of the R output.

Exhibit 1.10: R Output for Example 2—Analysis of Average Achievement for Girls and Boys

Report

Analysis for ASMMAT by IDCNTRY ITSEX

Cntry ID	Sex of Students	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASMMAT (Mean)	ASMMAT (s.e.)	Confidence Interval (95)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	Girl	1968	25773	525.19	47.05	0.82	528.09	3.83	521 to 536	88.71	3.39	0.00	77
Bulgaria	Boy	2135	28999	565.91	52.95	0.82	531.50	4.16	523 to 540	92.85	3.26	0.00	77
Canada	Girl	6966	137904	3010.95	50.69	0.57	493.86	2.06	490 to 498	78.86	1.48	0.00	125
Canada	Boy	6725	134124	2881.09	49.31	0.57	514.05	2.51	509 to 519	81.56	1.35	0.00	125
Finland	Girl	2875	30118	840.63	49.26	0.91	526.22	2.77	521 to 532	75.88	1.62	0.00	87
Finland	Boy	2928	31027	806.98	50.74	0.91	532.20	2.93	526 to 538	82.94	1.60	0.00	87
Morocco	Girl	3829	354116	12189.40	48.18	0.94	392.46	4.86	383 to 402	98.30	2.75	0.00	125
Morocco	Boy	4017	380907	12263.00	51.82	0.94	393.97	5.19	384 to 404	106.18	2.82	0.00	125
Oman	Girl	4043	39322	633.35	50.11	0.57	420.78	3.94	413 to 428	96.33	2.56	0.00	119
Oman	Boy	4025	39154	491.37	49.89	0.57	421.94	4.43	413 to 431	100.78	2.51	0.00	119
New Zealand	Girl	2438	29094	592.09	49.20	0.90	479.42	2.95	474 to 485	87.89	1.82	0.00	78
New Zealand	Boy	2503	30037	780.77	50.80	0.90	500.78	3.25	494 to 507	92.18	1.65	0.00	78

Countries are ordered sequentially according to their numeric ISO code. Each country's results are displayed on two lines, one for each value of the grouping variable (ITSEX). The country is identified in the first column (Cntry ID) and the second column (Sex of Students) indicates the category of the grouping variable ITSEX being reported according to the value labels (1: Girl, 2: Boy). The third column reports the number of valid cases (N of Cases), the fourth column reports the sum of weights of the sampled students (Sum of TOTWGT), indicating the estimated total students in the population represented by the sample, and the fifth column is the standard error of the sum of weights (Sum of TOTWGT (s.e.)).

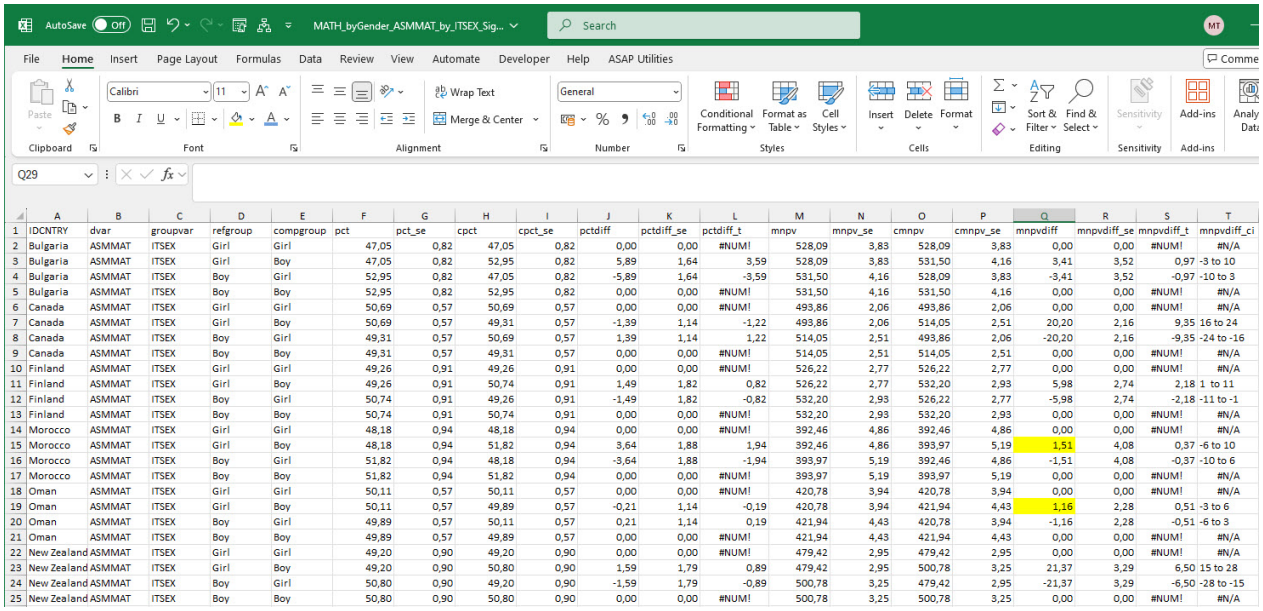
The next two columns report the weighted percentage of students in the particular category of the second grouping variable (Percent), which for this analysis is the percent of students in each category of ITSEX within IDCNTRY, and its standard error (Percent (s.e.)). The next two columns report the estimated average for the outcome variable for the group, in this case average mathematics achievement (ASMMAT (Mean)) and its standard error (ASMMAT (s.e.)). The "Confidence Interval (95)" column reports the 95% confidence interval for the mean for the group. The standard deviation of the achievement scores (Std.Dev.) and its standard error (Std.Dev. (s.e.)) are reported in the next two columns. The last two columns report the percentage of cases with missing data (Percent Missing)

and the number of jackknife zones used for computing standard errors (Number of Variance Strata), respectively.

The results for Oman are interpreted here as an example. In Exhibit 1.8, Oman is one of several countries that did not show a statistically significant difference in mathematics achievement between girls and boys. From the two lines of results for Oman in Exhibit 1.10, Oman's sample had approximately equal distribution of girls and boys: 50.11% of students were girls (standard error of 0.57) and 49.89% were boys (standard error of 0.57). The average mathematics achievement of girls was 420.78 (standard error of 3.94) and it was 421.94 for boys (standard error of 4.43).

The statistical significance of the gender differences, reported in the "Difference" column of Exhibit 1.8, can be determined by examining the output file named with the suffix "_sig" (MATH_byGender_ASMMAT_by_ITSEX_sig in this example) provided in R data (*.Rdata) and Excel (*.xlsx) file formats. This example refers to the xlsx version, which is the same for all software, shown in Exhibit 1.11 with differences in achievement between girls and boys highlighted for Morrocco and Oman.

Exhibit 1.11: Excel "Sig" Output for Example 2—Analysis of Average Achievement for Girls and Boys



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	IDCNTRY	dvar	groupvar	refgroup	compgroup	pct	pct_se	cpct	cpct_se	pctdiff	pctdiff_se	pctdiff_t	mnpv	mnpv_se	cmnpv	cmnpv_se	mnpvdiff	mnpvdiff_se	mnpvdiff_t	mnpvdiff_ci
2	Bulgaria	ASMMAT	ITSEX	Girl	Girl	47,05	0,82	47,05	0,82	0,00	0,00	#NUM!	528,09	3,83	528,09	3,83	0,00	0,00	#NUM!	#N/A
3	Bulgaria	ASMMAT	ITSEX	Girl	Boy	47,05	0,82	52,95	0,82	5,89	1,64	3,59	528,09	3,83	531,50	4,16	3,41	3,52	0,97	-3 to 10
4	Bulgaria	ASMMAT	ITSEX	Boy	Girl	52,95	0,82	47,05	0,82	-5,89	1,64	-3,59	531,50	4,16	528,09	3,83	-3,41	3,52	-0,97	-10 to 3
5	Bulgaria	ASMMAT	ITSEX	Boy	Boy	52,95	0,82	52,95	0,82	0,00	0,00	#NUM!	531,50	4,16	531,50	4,16	0,00	0,00	#NUM!	#N/A
6	Canada	ASMMAT	ITSEX	Girl	Girl	50,69	0,57	50,69	0,57	0,00	0,00	#NUM!	493,86	2,06	493,86	2,06	0,00	0,00	#NUM!	#N/A
7	Canada	ASMMAT	ITSEX	Girl	Boy	50,69	0,57	49,31	0,57	-1,39	1,14	-1,22	493,86	2,06	514,05	2,51	20,20	2,16	9,35	16 to 24
8	Canada	ASMMAT	ITSEX	Boy	Girl	49,31	0,57	50,69	0,57	1,39	1,14	1,22	514,05	2,51	493,86	2,06	-20,20	2,16	-9,35	-24 to -16
9	Canada	ASMMAT	ITSEX	Boy	Boy	49,31	0,57	49,31	0,57	0,00	0,00	#NUM!	514,05	2,51	514,05	2,51	0,00	0,00	#NUM!	#N/A
10	Finland	ASMMAT	ITSEX	Girl	Girl	49,26	0,91	49,26	0,91	0,00	0,00	#NUM!	526,22	2,77	526,22	2,77	0,00	0,00	#NUM!	#N/A
11	Finland	ASMMAT	ITSEX	Girl	Boy	49,26	0,91	50,74	0,91	1,49	1,82	0,82	526,22	2,77	532,20	2,93	5,98	2,74	2,18	1 to 11
12	Finland	ASMMAT	ITSEX	Boy	Girl	50,74	0,91	49,26	0,91	-1,49	1,82	-0,82	532,20	2,93	526,22	2,77	-5,98	2,74	-2,18	-11 to -1
13	Finland	ASMMAT	ITSEX	Boy	Boy	50,74	0,91	50,74	0,91	0,00	0,00	#NUM!	532,20	2,93	532,20	2,93	0,00	0,00	#NUM!	#N/A
14	Morocco	ASMMAT	ITSEX	Girl	Girl	48,18	0,94	48,18	0,94	0,00	0,00	#NUM!	392,46	4,86	392,46	4,86	0,00	0,00	#NUM!	#N/A
15	Morocco	ASMMAT	ITSEX	Girl	Boy	48,18	0,94	51,82	0,94	3,64	1,88	1,94	392,46	4,86	393,97	5,19	1,51	4,08	0,37	-6 to 10
16	Morocco	ASMMAT	ITSEX	Boy	Girl	51,82	0,94	48,18	0,94	-3,64	1,88	-1,94	393,97	5,19	392,46	4,86	-1,51	4,08	-0,37	-10 to 6
17	Morocco	ASMMAT	ITSEX	Boy	Boy	51,82	0,94	51,82	0,94	0,00	0,00	#NUM!	393,97	5,19	393,97	5,19	0,00	0,00	#NUM!	#N/A
18	Oman	ASMMAT	ITSEX	Girl	Girl	50,11	0,57	50,11	0,57	0,00	0,00	#NUM!	420,78	3,94	420,78	3,94	0,00	0,00	#NUM!	#N/A
19	Oman	ASMMAT	ITSEX	Girl	Boy	50,11	0,57	49,89	0,57	-0,21	1,14	-0,19	420,78	3,94	421,94	4,43	1,16	2,28	0,51	-3 to 6
20	Oman	ASMMAT	ITSEX	Boy	Girl	49,89	0,57	50,11	0,57	0,21	1,14	0,19	421,94	4,43	420,78	3,94	-1,16	2,28	-0,51	-6 to 3
21	Oman	ASMMAT	ITSEX	Boy	Boy	49,89	0,57	49,89	0,57	0,00	0,00	#NUM!	421,94	4,43	421,94	4,43	0,00	0,00	#NUM!	#N/A
22	New Zealand	ASMMAT	ITSEX	Girl	Girl	49,20	0,90	49,20	0,90	0,00	0,00	#NUM!	479,42	2,95	479,42	2,95	0,00	0,00	#NUM!	#N/A
23	New Zealand	ASMMAT	ITSEX	Girl	Boy	49,20	0,90	50,80	0,90	1,59	1,79	0,89	479,42	2,95	500,78	3,25	21,37	3,29	6,50	15 to 28
24	New Zealand	ASMMAT	ITSEX	Boy	Girl	50,80	0,90	49,20	0,90	-1,59	1,79	-0,89	500,78	3,25	479,42	2,95	-21,37	3,29	-6,50	-28 to -15
25	New Zealand	ASMMAT	ITSEX	Boy	Boy	50,80	0,90	50,80	0,90	0,00	0,00	#NUM!	500,78	3,25	500,78	3,25	0,00	0,00	#NUM!	#N/A

TIMSS 2023 International Results in Mathematics and Science labels differences between girls and boys as statistically significant based on two-tailed null hypothesis significance tests. For each country, the "sig" output reports the average achievement difference between the reference group (column D) and the comparison group (column E) in column Q, labeled "mnpvdiff." Dividing this value by its standard error ("mnpvdiff_se" in column R) gives a *t*-statistic ("mnpvdiff_t" in column S) for evaluating whether the estimated difference is significantly different from zero. For a *t*-test with infinite degrees of

freedom and an error level (α) of 5%, values greater than +1.96 (the upper critical value) or less than –1.96 (the lower critical value) indicate that the difference between the reference group (girls) average and the comparison group (boys) average is significantly different from zero. Values between –1.96 and +1.96 (the lower and upper critical values for α = 0.05) indicate the achievement difference between the two groups is not significantly different from zero. However, due to smaller degrees of freedom encountered when using estimated variances, the critical value may be larger than +1.96 or smaller than –1.96, (Johnson & Rust, 1993; Satterthwaite, 1941, 1946; Student, 1908; von Davier, 2025; Welch, 1947), so that a test with these as critical values may flag differences at an error rate larger than 5%.

The *t*-value for the achievement difference between girls and boys in Oman is 0.51, which is between the lower and upper critical *t*-values for an α level of 0.05. The (null) hypothesis was not rejected, indicating the achievement difference is not statistically significant.

It can be observed in Exhibit 1.8 that differences of about 10 points or more tend to be flagged as statistically significant with an alpha level of 0.05. In contrast, smaller differences tend not to be flagged as significantly different from 0, unless their standard error (given in parentheses) is unusually small. Statistical significance does not necessarily imply that a difference is practically meaningful and achievement differences between girls and boys should always be considered in broader educational contexts within countries.

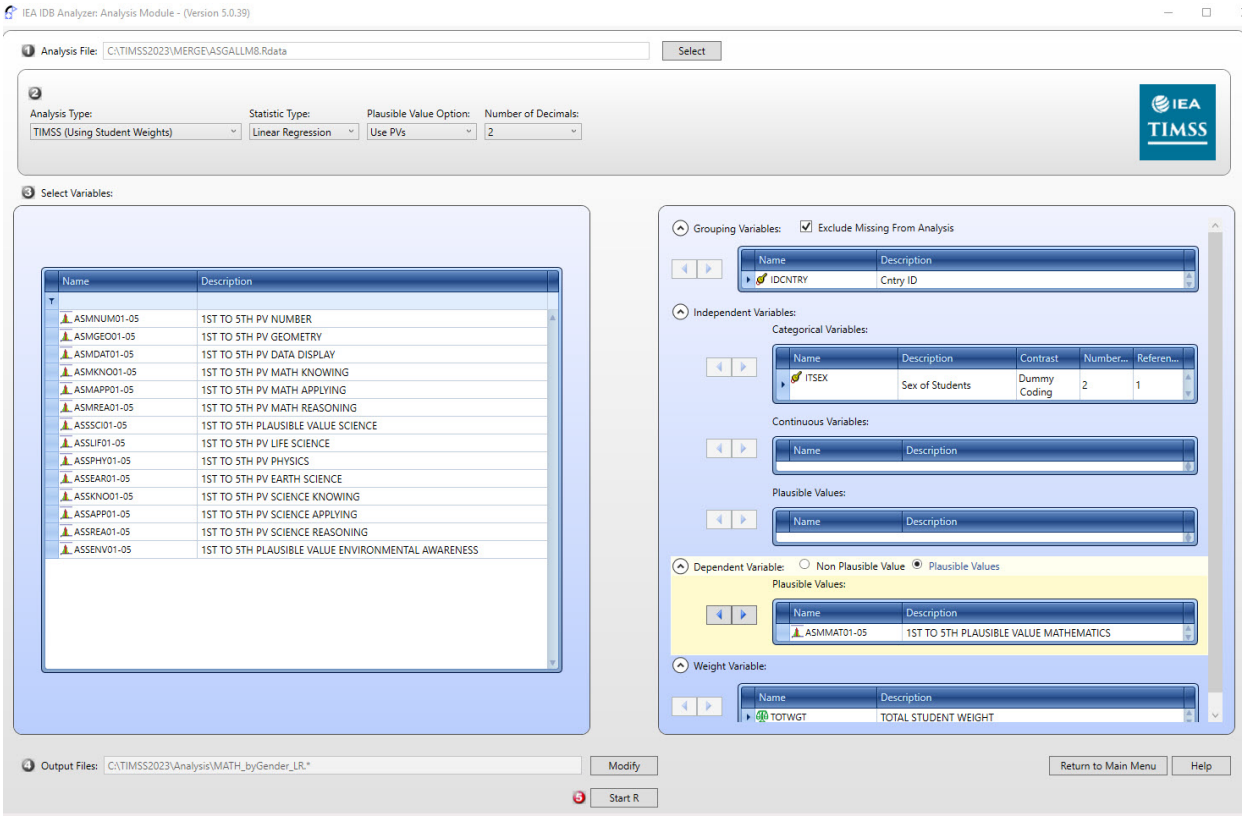
Example 3—Linear Regression Analysis with Student Context Data

The third example is an extension of the previous (Example 2) and describes an alternative method to examine the difference in fourth-grade mathematics achievement between girls and boys and determine if it is statistically significant. This example also demonstrates the **Dummy Coding** feature of the IEA IDB Analyzer. Like Example 2, the results of this example are presented in Exhibit 1.1.2 of *TIMSS 2023 International Results in Mathematics and Science* and are shown above in Exhibit 1.8 in the column labeled “Difference.”

The ITSEX variable has a value of one (1) for girls and two (2) for boys. By using ITSEX as a categorical variable in the IEA IDB Analyzer with **Dummy Coding**, and defining category 1 (girls) as the reference category, the regression intercept estimate is the average mathematics achievement of girls, and the regression slope is the estimated change in average mathematics achievement for boys.

The **Analysis Module** of the IEA IDB Analyzer is used to conduct the analysis, with **Linear Regression** defined as the statistic type in the following steps. Exhibit 1.12 shows the completed Analysis Module for this example.

Exhibit 1.12: IEA IDB Analyzer Analysis Module Setup for Example 3—Linear Regression Analysis with Student Data



1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ASGALLM8 as the **Analysis File** by clicking the **Select** button.
3. Select **TIMSS (Using Student Weights)** as the **Analysis Type**.
4. Select **Linear Regression** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The box for **Exclude Missing From Analysis** should be checked for this analysis. This option uses listwise deletion, excluding records with missing values on any of the analysis variables.
8. The IDB Analyzer automatically selects the variable IDCNTRY for the **Grouping Variables**. No additional grouping variables are needed for this analysis.
9. Specify ITSEX as a **Categorical Variable** in the **Independent Variables** section, first by clicking the **Categorical Variables** field to activate it. Then, select ITSEX from the list of available variables in the left panel, and move it to the right **Categorical Variables** field by clicking the **right arrow** (▸) button. Next, click the

Contrast field of ITSEX, and its drop-down menu will appear. **Dummy Coding** is selected by default, and the IEA IDB Analyzer determines the **Number of Categories** for the variable ITSEX (2). By default, category 1 (girls) will be selected as the **Reference Category**. These settings should not be changed.

10. In the **Dependent Variable** section, click the **Plausible Values** radio button. Specify the achievement scores to be used as the **Dependent Variable** by first clicking the **Plausible Values** field to activate it. Then, select ASMMAT01–05 from the list of available variables in the left panel, and move it to the right **Plausible Values** field by clicking the **right arrow** (▸) button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis uses student context data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer also will create a syntax file of the same name and in the same folder, with the code necessary to perform the analysis. In Exhibit 1.12, the syntax file MATH_byGender_LR.R and the output files with the same name will be created and stored in the C:\TIMSS2023\Analysis folder.
13. Click the **Start R** button to create the R script and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard.

Conducting a linear regression analysis with the IEA IDB Analyzer produces several results output files. The main results for this example are the regression coefficients, reported in the file named with the suffix "_Coef." Separate output files are also produced with descriptive statistics by country, named with the suffix "_Desc" for the intercept (girls' average achievement) and the regression coefficients (change in achievement from girls to boys), and with estimated *R*-square values for the regression models, named with the suffix "_Model" (MATH_byGender_LR_Model). All results are included in the HTML output produced by R.

Exhibit 1.13 displays the main results for this example analysis—the regression coefficients—in the R output file for the example countries. Countries are ordered numerically according to their numeric ISO code, with their results each displayed on two lines: the first for the intercept (CONSTANT) and the second for the ITSEX coefficient (ITSEX_D2). For all regression analyses, there will be as many lines per country as there are regression coefficients, including the intercept.

Exhibit 1.13: R Regression Coefficient Output for Example 3—Linear Regression Analysis with Student Data

Regression Coefficients

EqVar	Cntry ID	Variable	Regression Coefficient	Regression Coefficient (s.e.)	Regression Coefficient (t-value)	Stndrdzd. Coefficient	Stndrdzd. Coefficient (s.e.)	Stndrdzd. Coefficient (t-value)
(CONSTANT)	Bulgaria	(CONSTANT)	528.09	3.83	137.71	NaN	NaN	NaN
ITSEX_D2	Bulgaria	ITSEX_D2	3.41	3.52	0.97	0.02	0.02	0.97
(CONSTANT)	Canada	(CONSTANT)	493.86	2.06	240.18	NaN	NaN	NaN
ITSEX_D2	Canada	ITSEX_D2	20.20	2.16	9.35	0.12	0.01	9.71
(CONSTANT)	Finland	(CONSTANT)	526.22	2.77	189.89	NaN	NaN	NaN
ITSEX_D2	Finland	ITSEX_D2	5.98	2.74	2.18	0.04	0.02	2.17
(CONSTANT)	Morocco	(CONSTANT)	392.46	4.86	80.81	NaN	NaN	NaN
ITSEX_D2	Morocco	ITSEX_D2	1.51	4.08	0.37	0.01	0.02	0.37
(CONSTANT)	Oman	(CONSTANT)	420.78	3.94	106.90	NaN	NaN	NaN
ITSEX_D2	Oman	ITSEX_D2	1.16	2.28	0.51	0.01	0.01	0.51
(CONSTANT)	New Zealand	(CONSTANT)	479.42	2.95	162.42	NaN	NaN	NaN
ITSEX_D2	New Zealand	ITSEX_D2	21.37	3.29	6.50	0.12	0.02	6.49

The countries are identified in the second column (Cntry ID) and the third column (Variable) indicates the intercept (CONSTANT) or the regression coefficient being reported. The fourth column reports the "Regression Coefficient" ("b" in Excel), indicating, for the intercept, the average value of the dependent variable for the reference group (girls in this case), and for the regression coefficients, the average difference in the dependent variable from the intercept. The fifth column is the standard error of the regression coefficient (Regression Coefficient (s.e.)). The sixth column reports the value of the *t*-statistic for the regression coefficient (Regression Coefficient (*t*-value)). The IEA IDB Analyzer also computes standardized regression coefficients in the last three columns, corresponding to the fourth, fifth, and sixth columns, whereby the dependent and independent variables are standardized to have a mean of zero (0) and standard deviation of one (1).

In Exhibit 1.13, the first line of results for Oman, labeled "(CONSTANT)" ("Intercept" in SAS), indicates the estimated average mathematics achievement of fourth-grade girls in Oman: 420.78 with a standard error of 3.94. This estimate concurs with the results obtained in the previous example (Exhibit 1.10). From the second line of results labeled "ITSEX_D2," the fourth-grade boys in Oman had a positive average mathematics achievement difference from girls of 1.16 with an estimated standard error of 2.28. The *t*-

value for the coefficient is 0.51, which is less than 1.96 (the upper critical value for α = 0.05), indicating this achievement difference is not statistically significant. Counting the two regression coefficients together (420.78 + 1.16) yields the estimated average mathematics achievement of fourth-grade boys in Oman, which was 421.94 in Exhibit 1.10.

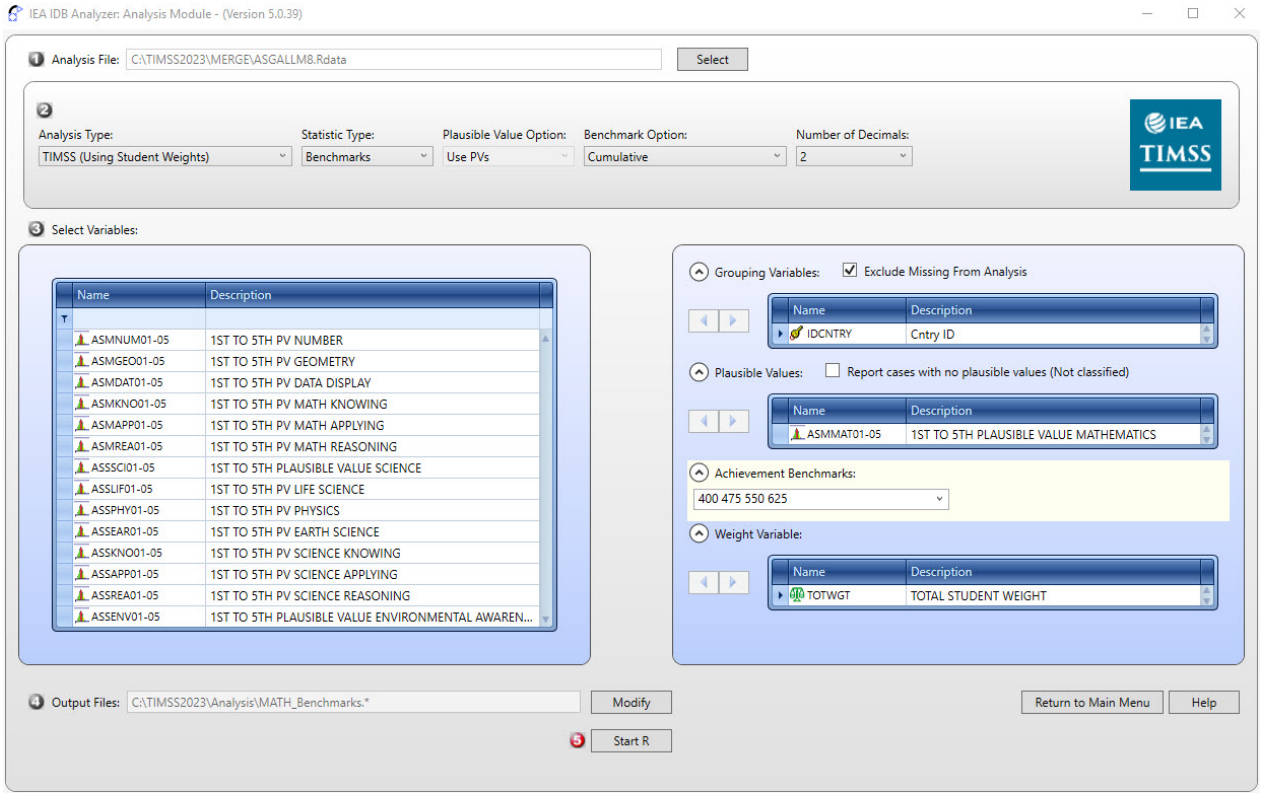
Example 4—Analysis of TIMSS International Benchmarks

This section describes how to use the IEA IDB Analyzer to perform analyses of student achievement in relation to the TIMSS International Benchmarks. This example computes the percentages of students reaching each of the four TIMSS 2023 International Benchmarks of fourth-grade mathematics achievement (Advanced, High, Intermediate, and Low) using the merged ASGALLM8 data file described earlier under the section on Merging Data Files with the IEA IDB Analyzer. These results, presented in Exhibit 1.1.4 of the TIMSS 2023 International Results report, are shown below in Exhibit 1.14.

↓ Exhibit 1.14: International Results Exhibit of Example 4—Analysis of TIMSS International Benchmarks

This example is conducted in the **Analysis Module** of the IEA IDB Analyzer with the following steps. The completed Analysis Module is shown in Exhibit 1.15.

Exhibit 1.15: IEA IDB Analyzer Analysis Module Setup for Example 4—Analysis of TIMSS International Benchmarks



1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Specify the data file ASGALLM8 as the **Analysis File** by clicking the **Select** button.
3. Select **TIMSS (Using Student Weights)** as the **Analysis Type**.
4. Select **Benchmarks** as the **Statistic Type**.
5. Select the **Cumulative** option under the **Benchmark Option** drop-down menu to get cumulated percentages of students reaching the TIMSS International Benchmarks.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The variable IDCNTRY is selected automatically for **Grouping Variables**. No additional grouping variables are needed for this analysis.
8. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASMMAT01–05 from the list of available variables in the left panel, and move it to the right **Plausible Values** field by clicking the **right arrow** (▸) button.
9. In the **Achievement Benchmarks** field, specify the average achievement score for each of the TIMSS International Benchmarks in ascending order: 400, 475, 550, and 625 (Low, Intermediate, High, and Advanced, respectively). These values can be entered manually with each separated by a blank space, or they can be selected by clicking on the drop-down menu available for this field.
10. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis uses student data.
11. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer will create a syntax file with the code necessary to perform the analysis. In Exhibit 1.15, the syntax file MATH_Benchmarks.R and the output files with the same name will be created and stored in the folder C:\TIMSS2023\Analysis.

Click the **Start R** button (or Start SPSS/SAS) to create the syntax file and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** (🏃) button (or select **Submit** in the **Run** menu).

The IDB Analyzer produces and saves the results output in three file formats within the folder specified in Step 11—an HTML output file (or output in SPSS/SAS), R data file (*.Rdata), and Microsoft Excel Worksheet (*.xlsx). Graphs are included only in the HTML (or SPSS/SAS) output files. Exhibit 1.16 presents the results of Example 4 as shown in the

R output, under the "Report" section. Results are shown for four of the example countries: Bulgaria, Canada, Finland, and Morocco.

Exhibit 1.16: R Output for Example 4—Analysis of TIMSS International Benchmarks

Report

Percentage by Performance Group of ASMMAT

Cntry ID	cutvar	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)
Bulgaria	1. At or Above 400	3841	49780	948.98	90.89	1.34
Bulgaria	2. At or Above 475	3303	40664	1020.62	74.24	1.66
Bulgaria	3. At or Above 550	2146	24853	755.82	45.38	1.29
Bulgaria	4. At or Above 625	675	7404	516.36	13.52	0.94
Canada	1. At or Above 400	12039	244668	4775.63	89.76	0.61
Canada	2. At or Above 475	8416	176333	4021.39	64.69	1.00
Canada	3. At or Above 550	3538	78548	2823.47	28.82	0.92
Canada	4. At or Above 625	693	17104	1445.67	6.27	0.52
Finland	1. At or Above 400	5444	57361	1225.93	93.81	0.65
Finland	2. At or Above 475	4467	46705	1147.55	76.38	1.10
Finland	3. At or Above 550	2464	25416	969.51	41.57	1.34
Finland	4. At or Above 625	633	6450	492.38	10.55	0.78
Morocco	1. At or Above 400	3363	340976	16113.09	46.39	1.94
Morocco	2. At or Above 475	1424	159311	12726.56	21.67	1.62
Morocco	3. At or Above 550	400	49731	7317.84	6.77	0.96
Morocco	4. At or Above 625	79	10952	3128.70	1.49	0.42

Countries are ordered according to their numeric ISO code, and each country's results are displayed on four lines, one for each TIMSS International Benchmark. The countries are identified in the first column (Cntry ID) and the second column (cutvar) indicates the benchmark level being reported (this is labeled "Performance Group" in SPSS). The third column reports the number of valid cases (N of Cases), the fourth column reports the sum of weights of the sampled students (Sum of TOTWGT) corresponding to the number of students in the population represented by the sample, and the fifth column is the standard error of the sum of weights (Sum of TOTWGT (s.e.)). The last two columns report the

cumulative percentage of students reaching each benchmark (Percent) and its standard error (Percent (s.e.)).

Conducting Analyses with TIMSS Home Context Data

This section presents an analysis conducted using the IEA IDB Analyzer with home context data from the TIMSS 2023 International Database. Home context data were collected from the parents of fourth-grade students with the TIMSS 2023 Home Questionnaire. Like the previous section, the example below is an actual analysis used to produce exhibits in the *TIMSS 2023 International Results in Mathematics and Science* report.

In general, before conducting analyses using TIMSS contextual variables such as those in the home context data files, users should refer to the relevant codebook for the data file to identify the appropriate variables and understand the coding scheme. From the TIMSS 2023 International Database webpage, the Context Questionnaire Variables downloads present all the context questionnaires administered in TIMSS 2023 and the associated variable names under which the data are saved. The National Adaptations Database downloads should also be checked for any national adaptations made to the questionnaire items that may impact international comparability.

Analyzing home context data from the TIMSS 2023 International Database requires that the home context data files (named beginning with "ASH") be merged with either student achievement files ("ASA") or student context files ("ASG") to retrieve the achievement scores and required sample design variables. This example uses home context data merged with student context data described earlier in the section on Merging Data Files with the IEA IDB Analyzer (ASHALLM8). Example 5 computes the average score on a context questionnaire scale, along with the percentages of students—with their average achievement—for each of the categories of the scale's corresponding index. The analysis replicates Exhibit 3.2.5 of *TIMSS 2023 International Results in Mathematics and Science*, repeated below in Exhibit 1.17.

↓ Exhibit 1.17: International Results Exhibit of Example 5—Analysis of a Context Questionnaire Scale with Home Context Data

At the bottom of Exhibit 1.17, several countries are listed below the International Average row with annotations or without any results reported. Throughout TIMSS' history, some countries have struggled to attain high participation rates from parents, and some countries did not administer the home questionnaire. The exhibits display special annotations to caution readers about notably low response rates. If data were available for less than 40% of students in a country, the country is annotated with a "y," and the data are not reported. If data were available for 40%–50% of students, the country is annotated with an "x," and the data are reported but do not contribute to the International Average.

Example 5—Analysis of a Context Questionnaire Scale with Home Context Data

As described in Chapter 15 of *TIMSS 2023 Technical Report*, TIMSS 2023 reports context questionnaire data by creating context questionnaire scales based on Rasch modeling. The context questionnaire scales are available in the International Database for analysis within the context data file corresponding to the respondent of the questionnaire items. Each context questionnaire scale variable is a Rasch score with an international centerpoint of 10 and standard deviation of 2. From each context questionnaire scale, an index is derived that divides the range of scores on that scale into three categories: the most desirable scores (high values), the least desirable scores (low values), and the remaining scores in between.

Many of these context questionnaire scales and their corresponding indices were reported in *TIMSS 2023 International Results in Mathematics and Science*. Exhibit 1.17 shows one such example, Exhibit 3.2.5 of the report, with results for the TIMSS 2023 *Early Literacy and Numeracy Tasks* scale, based on parents' reports regarding how well their children could do 11 tasks when they began primary school. Results on the Rasch scale are reported for each country as an "Average Scale Score" and its corresponding index is reported as the percentages of students with parents in each category—"very well," "moderately well," and "not well"—along with their average mathematics achievement.

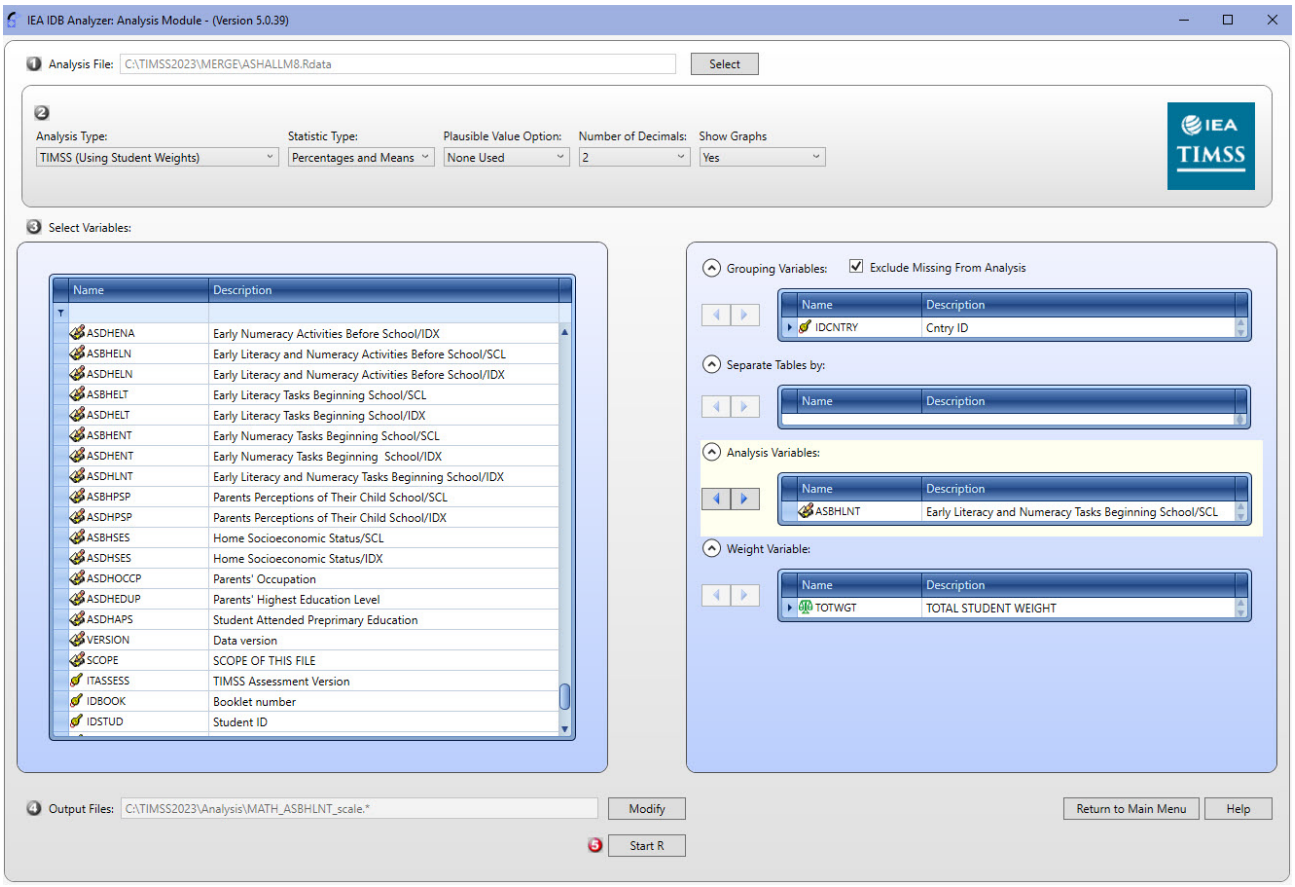
This example replicates both the average scale score column of Exhibit 1.17 and the columns reporting percentages of students and their average achievement in each category. This is accomplished in two analysis steps, both using the merged ASHALLM8 data file.

Step 1: Compute Average Scale Scores

The first step of Example 5 will compute the average scale score of the *Early Literacy and Numeracy Tasks* scale for each country, using the scale variable ASBHLNT. It is conducted in the **Analysis Module** of the IEA IDB Analyzer using the following steps. The completed Analysis Module for this step is shown in Exhibit 1.18.

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Specify the data file ASHALLM8 as the **Analysis File** by clicking the **Select** button.
3. Select **TIMSS (Using Student Weights)** as the **Analysis Type**.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **None Used** as the **Plausible Value Option**, because achievement scores are not used for this part of the analysis.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.

Exhibit 1.18: IEA IDB Analyzer Analysis Module Setup for Example 5—Analysis of a Context Questionnaire Scale with Home Context Data (Step 1)



7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce two bar graphs in the output file: one for average scale score by country, and one for average percent of the total students in each country.
8. The IDB Analyzer automatically selects the variable IDCNTRY for the **Grouping Variables**. No additional grouping variables are needed for this analysis. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (non-achievement) variables. See the IEA IDB Analyzer Help manual for more information.
10. Specify the variable ASBHLNT to be used for the analysis by first clicking the **Analysis Variables** field to activate it. Then, select ASBHLNT from the list of available variables in the left panel, and move it to the right **Analysis Variables** field by clicking the **right arrow** (▸) button.

11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis uses student context data combined with home context data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer will create an R, SPSS, or SAS syntax file of the same name and in the same folder, with the code necessary to perform the analysis. In Exhibit 1.18, the syntax file MATH_ASBHLNT_scale.R and the output files with the same name will be created and stored in the folder C:\TIMSS2023\Analysis.
13. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** ( ) button (or select **Submit** in the **Run** menu).

The IDB Analyzer produces and saves the results output in three file formats within the folder specified in Step 12—an HTML output file (or output in SPSS/SAS), R data file (*.Rdata), and Microsoft Excel Worksheet (*.xlsx). Graphs are included only in the HTML output files.

Exhibit 1.19 shows the results for the first step of Example 5 in the R output file with six example countries: Bulgaria, Canada, Finland, Morocco, New Zealand, and Oman. This step of the analysis produces the same results output files as described in Example 1, with IDCNTRY as the grouping variable and the scale variable ASBHLNT as the outcome. Countries are ordered according to their numeric ISO code.

Exhibit 1.19: R Output for Example 5—Analysis of a Context Questionnaire Scale with Home Context Data (Step 1)

Report

Analysis for ASBHLNT by IDCNTRY

Cntry ID	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASBHLNT (Mean)	ASBHLNT (s.e.)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	4072	54297	610.67	5.25	0.11	9.51	0.08	2.21	0.06	0.87	77
Canada	5783	113125	3815.65	10.93	0.41	10.45	0.04	1.95	0.03	58.50	125
Finland	5244	54967	1126.25	5.31	0.14	10.01	0.04	1.95	0.02	10.10	87
Morocco	7569	721600	20131.90	69.71	0.66	9.37	0.10	2.37	0.07	1.83	125
Oman	7150	69779	1063.71	6.74	0.16	11.36	0.03	1.96	0.02	11.08	118
New Zealand	1872	21381	813.77	2.07	0.09	9.10	0.05	1.66	0.04	63.88	77

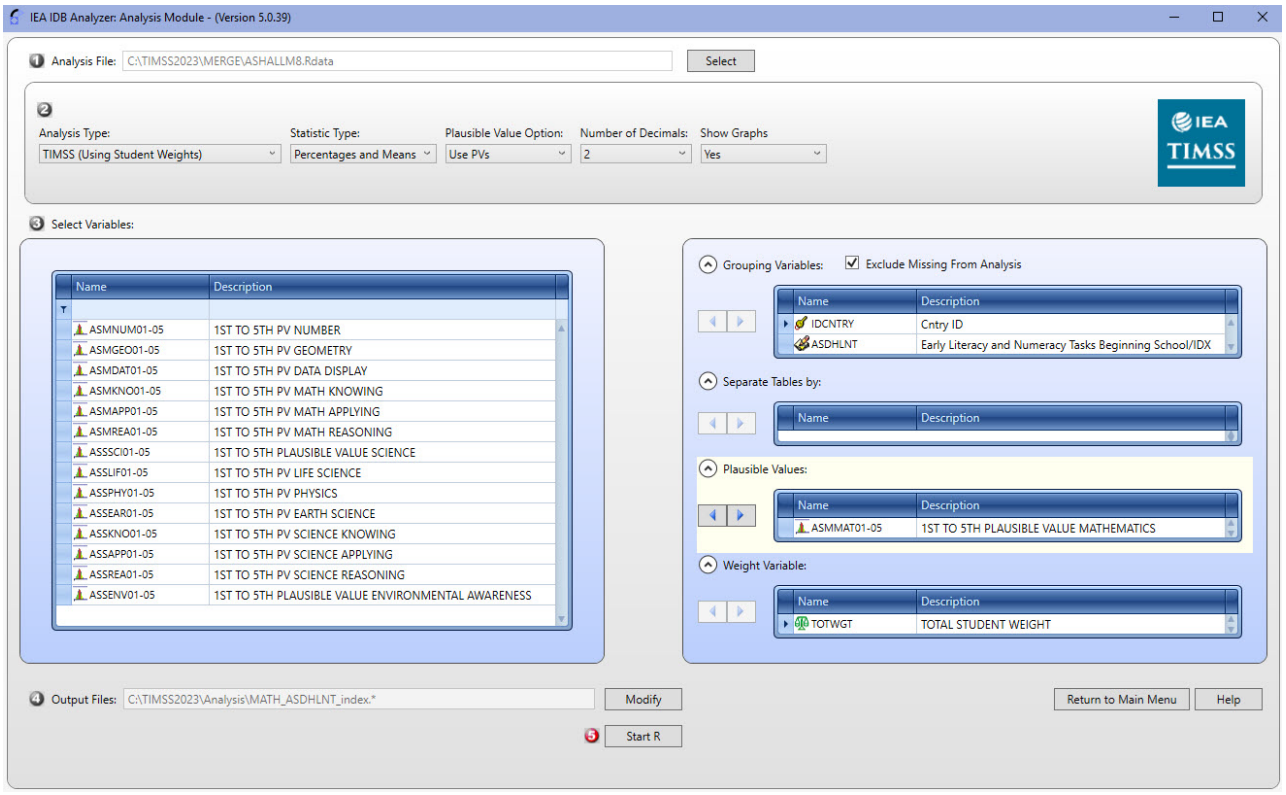
Canada is interpreted as an example. On average, students in Canada scored 10.45 based on their parents' responses to the items on the *Early Literacy and Numeracy Tasks* context questionnaire scale, with a standard error of 0.04. This is located approximately at the international centerpoint of 10. Chapter 15 in *TIMSS 2023 Technical Report* provides more information about the TIMSS context scales and their interpretation.

When analyzing context data, it is important to check the "Percent Missing" column in the output, which shows that 58.50% of students in New Canada do not have data for the scale. In Exhibit 1.17, Canada is listed below the International Average because they had a high level of missing data on the ASBHLNT variable. The "x" annotation indicates that data were available for only 40%–50% of students, so their results should be interpreted with caution. The data are reported but do not contribute to the International Average. Note that the "Percent Missing" column only reports missing data on the outcome variable.

Step 2: Compute Percentages of Students and Average Achievement by Scale Category

In the second step of this example, the percentages of students are computed—with their average mathematics achievement—in each category of the corresponding index variable ASDHLNT. This step is conducted by the **Analysis Module** of the IEA IDB Analyzer using the following steps. Exhibit 1.20 shows the completed Analysis Module for this step.

Exhibit 1.20: IEA IDB Analyzer Analysis Module Setup for Example 5—Analysis of a Context Questionnaire Scale with Home Context Data (Step 2)



1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Specify the data file ASHALLM8 as the **Analysis File** by clicking the **Select** button.
3. Select **TIMSS (Using Student Weights)** as the **Analysis Type**.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**, because average achievement will be computed by the grouping variable ASDHLNT.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce three graphs in the output file: a line graph of average achievement for each scale category by country, a clustered bar graph of average achievement for each scale category by country, and a stacked bar graph of average percent of students for each scale category by country. R also produces separate graphs for each country.
8. Specify the variable ASDHLNT as a second grouping variable by first clicking the **Grouping Variables** field to activate it. Then, select ASDHLNT from the list of variables in the left panel, and move it to the **Grouping Variables** field by clicking the **right arrow** (▸) button. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (not achievement) variables. See the Help manual for more information.
10. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASMMAT01–05 from the list of available variables in the left panel, and move it to the right **Plausible Values** field by clicking the **right arrow** (▸) button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis combines student context data with home context data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. In Exhibit 1.20, the syntax file MATH_ASDHLNT_index.R and the output files with the same name will be created and stored in the C:\TIMSS2023\Analysis folder.
13. Click the **Start R** button to create the R script and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard.

Exhibit 1.21 shows the results for the second step of Example 5 with the six example countries. The results are presented in the "Report" section of the output. This analysis produces the results output files consistent with those described in Example 2, with ASDHLNT as the second grouping variable instead of ITSEX. Each country's results are displayed on three lines, one for each value of the scale index variable.

Exhibit 1.21: R Output for Example 5—Analysis of a Context Questionnaire Scale with Home Context Data (Step 2)

Report

Analysis for ASMMAT by IDCNTRY ASDHLNT

Cntry ID	Early Literacy and Numeracy Tasks Beginning School/IDX	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASMMAT (Mean)	ASMMAT (s.e.)	Confidence Interval (95)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	Very Well	802	9676	524.45	17.82	0.96	576.94	5.06	567 to 587	79.77	3.66	0.00	74
Bulgaria	Moderately Well	2140	26913	752.37	49.57	1.40	546.17	3.03	540 to 552	77.04	2.42	0.00	77
Bulgaria	Not Well	1130	17708	972.78	32.61	1.67	482.24	6.78	469 to 496	92.82	3.93	0.00	77
Canada	Very Well	1502	31881	1580.02	28.18	0.91	562.48	3.64	555 to 570	75.97	2.90	0.00	125
Canada	Moderately Well	3165	59975	2243.05	53.02	0.87	515.30	2.60	510 to 520	77.87	1.75	0.00	125
Canada	Not Well	1116	21269	1011.55	18.80	0.73	486.59	3.34	480 to 493	75.02	3.06	0.00	125
Finland	Very Well	1222	12538	487.74	22.81	0.81	580.63	3.06	575 to 587	69.82	1.86	0.00	87
Finland	Moderately Well	2576	26844	815.74	48.84	0.89	536.77	2.47	532 to 542	68.06	1.30	0.00	87
Finland	Not Well	1446	15585	533.27	28.35	0.86	492.42	2.89	487 to 498	75.36	1.81	0.00	87
Morocco	Very Well	1414	124670	7273.61	17.28	0.96	439.16	5.34	429 to 450	92.61	3.31	0.00	116
Morocco	Moderately Well	3959	354535	14772.53	49.13	1.39	396.58	4.90	387 to 406	98.81	2.96	0.00	125
Morocco	Not Well	2196	242395	14026.87	33.59	1.70	366.46	7.33	352 to 381	103.88	4.58	0.00	125
Oman	Very Well	3350	33259	637.07	47.66	0.72	440.99	4.59	432 to 450	96.65	3.43	0.00	117
Oman	Moderately Well	3150	30573	730.68	43.81	0.68	408.10	4.52	399 to 417	97.17	2.28	0.00	117
Oman	Not Well	650	5947	303.10	8.52	0.41	379.27	5.98	368 to 391	94.06	3.40	0.00	114
New Zealand	Very Well	157	1747	192.49	8.17	0.89	555.07	7.28	541 to 569	88.56	6.20	0.00	57
New Zealand	Moderately Well	915	10394	504.75	48.61	1.52	526.95	3.88	519 to 535	82.07	2.31	0.00	77
New Zealand	Not Well	800	9241	488.73	43.22	1.42	493.40	4.32	485 to 502	87.18	2.41	0.00	77

As shown in the three lines of the results for Canada, 28.18% of fourth-grade students in Canada had students whose parents reported they could do literacy and numeracy

tasks “very well” (standard error of 0.91) and their average mathematics achievement was 562.48 (standard error of 3.64); 53.02% of students who could do the tasks “moderately well” when beginning primary school (standard error of 0.87) and their average mathematics achievement was 515.30 (standard error of 2.60); and 18.80% of students could do the tasks “not well” (standard error of 0.73) and their average mathematics achievement was 486.59 (standard error of 3.34).

In the first step of this example, the “Percent Missing” column indicated that Canada had a high proportion of missing data on the context scale outcome. Because in Step 8, the box was checked for **Excluding Missing From Analysis**, any cases missing data on the grouping variable were excluded and not accounted for in the output. Nevertheless, it is important to consider missing data when interpreting the results of analyses using context data. Users of the data can check for the proportion of missing data on the context variables in the Data Almanacs, described in Chapter 2 of this User Guide.

Conducting Analyses with TIMSS School Context Data

When analyzing school context data from the TIMSS 2023 International Database, it is preferable that they be analyzed to make inferences about students attending schools with a given characteristic, rather than about schools of a given characteristic. Analyzing school context data with student achievement requires that the school context data files be merged with the student context data files to retrieve the achievement scores and required sample design variables.

This section presents an analysis conducted using the IEA IDB Analyzer with school context data collected from principals of schools attended by fourth-grade students with the TIMSS 2023 School Questionnaire. Example 6 uses school context data merged with student context data to compute the percentages and average achievement of fourth-grade students who attend schools composed of students with different levels of socioeconomic background. The results of this analysis are presented in Exhibit 4.1.2 of *TIMSS 2023 International Results in Mathematics and Science,* repeated below in Exhibit 1.22.

↓ Exhibit 1.22: International Results Exhibit of Example 6—Analysis of Average Achievement by School Socioeconomic Composition

Example 6—Analysis of Average Achievement by School Socioeconomic Composition

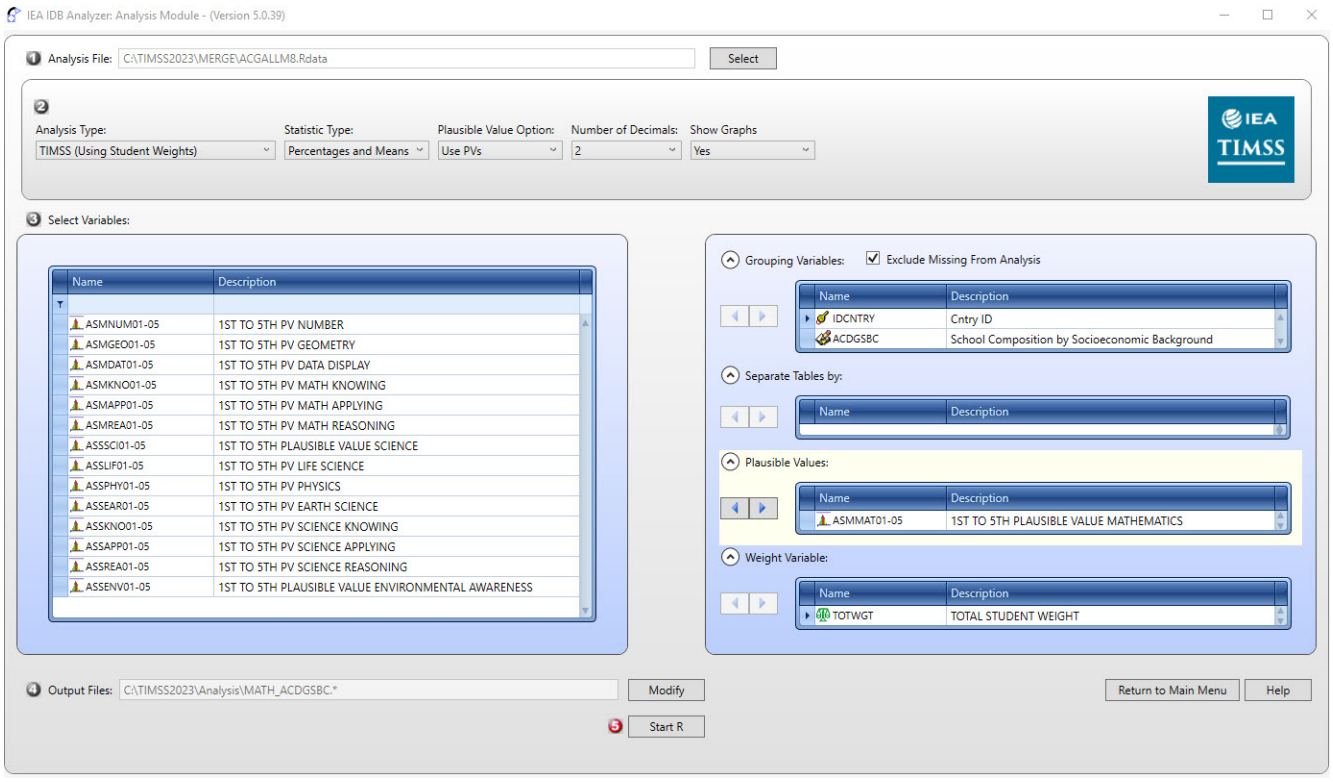
In this example, the **Percentages and Means** statistic type is used along with the **Use PVs** option to estimate the percentages of fourth-grade students with their average mathematics achievement by reporting categories of students’ socioeconomic background as reported by school principals.

Before conducting analyses using school context variables, users should refer to the relevant codebook for the data file to identify the appropriate variables related to the school's composition of students by socioeconomic background and understand the coding scheme. From the TIMSS 2023 International Database webpage, the Context Questionnaire Variables downloads present all the context questionnaires administered in TIMSS 2023 and the associated variable names under which the data are saved. The National Adaptations Database downloads should also be checked for any national adaptations made to the questionnaire items that may impact international comparability.

The codebook for the school context data file indicates that the derived variable ACDGSBC contains information on the socioeconomic composition of schools in three categories represented by the columns of Exhibit 1.22. As described in the Derived Context Variables download, two source variables were used to derive ACDGSBC: ACBG03A for the percentage of students economically disadvantaged and ACBG03B for the percentage of students economically affluent. Schools are characterized as "more affluent," "more disadvantaged," or "neither more affluent nor more disadvantaged."

This example uses the merged data file ACGALLM8 described earlier in the section on Merging Data Files with the IEA IDB Analyzer. This example analysis is conducted in the **Analysis Module** of the IEA IDB Analyzer using the following steps. The completed Analysis Module is shown in Exhibit 1.23.

Exhibit 1.23: IEA IDB Analyzer Analysis Module Setup for Example 6—Analysis of Average Achievement by School Socioeconomic Composition



1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ACGALLM8 as the **Analysis File** by clicking the **Select** button.
3. Select **TIMSS** (**Using Student Weights**) as the **Analysis Type,** because the school context data is analyzed as student attributes.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**, because average achievement will be computed by the grouping variable ACDGSBC.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce three graphs in the output file: a line graph of average achievement for each category of school composition by country, a clustered bar graph of average achievement for each category of school composition by country, and a stacked bar graph of average percent of students for each category of school composition by country. R also provides separate graphs for each country.
8. Specify the variable ACDGSBC as a second grouping variable by first clicking the **Grouping Variables** field to activate it. Then, select ACDGSBC from the list of variables in the left panel, and move it to the **Grouping Variables** field by clicking the **right arrow** (▸) button. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (not achievement) variables. See the Help manual for more information.
10. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASMMAT01–05 from the list of available variables in the left panel, and move it to the right **Plausible Values** field by clicking the **right arrow** (▸) button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because of the **Analysis Type** selected in Step 3 for this analysis which uses school context data linked to student context data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer also will create an R syntax file (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the analysis. In Exhibit 1.23, the syntax file MATH_ACDGSBC.R and the

output files with the same name will be created and stored in the folder C:\TIMSS2023\Analysis.

13. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** ( ) button (or select **Submit** in the **Run** menu).

The results as shown in the R output file are presented in Exhibit 1.24 with four of the example countries: Bulgaria, Canada, Finland, and Morocco. The results are presented in the same manner as in Example 2 and Example 5 (step 2), with countries identified in the first column and the second column the categories of ACDGSBC. Each country's results are presented on three lines, one for each value of the ACDGSBC variable. There are fewer lines if any category does not have any observations.

Exhibit 1.24: R Output for Example 6—Analysis of Average Achievement by School Socioeconomic Composition

Report

Analysis for ASMMAT by IDCNTRY ACDGSBC

Cntry ID	School Composition by Socioeconomic Background	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASMMAT (Mean)	ASMMAT (s.e.)	Confidence Interval (95)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	More Affluent	1603	16781	1890.21	37.29	3.90	567.61	5.52	557 to 578	76.97	2.60	0.00	37
Bulgaria	Neither More Affluent nor More Disadvantaged	1371	17915	1823.35	39.81	3.82	539.28	5.05	529 to 549	79.03	3.20	0.00	38
Bulgaria	More Disadvantaged	469	10301	1241.24	22.89	2.58	458.92	13.87	432 to 486	90.37	4.93	0.00	21
Canada	More Affluent	5580	108635	6433.75	47.66	2.58	521.66	2.88	516 to 527	79.10	1.82	0.00	114
Canada	Neither More Affluent nor More Disadvantaged	4219	83550	5706.65	36.66	2.41	497.05	3.35	490 to 504	78.25	1.76	0.00	107
Canada	More Disadvantaged	1866	35740	5070.24	15.68	2.12	480.32	5.46	470 to 491	83.19	2.41	0.00	75
Finland	More Affluent	1505	14726	2038.99	26.29	3.59	544.57	4.17	536 to 553	78.04	2.40	0.00	39
Finland	Neither More Affluent nor More Disadvantaged	2917	30481	2496.64	54.42	3.94	528.17	2.99	522 to 534	77.07	1.68	0.00	62
Finland	More Disadvantaged	1000	10800	1924.40	19.28	3.41	506.07	6.78	493 to 519	83.49	3.46	0.00	28
Morocco	More Affluent	201	38930	13864.30	7.36	2.57	426.50	32.44	363 to 490	110.66	14.53	0.00	9
Morocco	Neither More Affluent nor More Disadvantaged	1033	121577	19473.53	22.97	3.68	435.26	9.01	418 to 453	95.97	4.80	0.00	35
Morocco	More Disadvantaged	4328	368762	30059.40	69.67	3.69	376.79	6.57	364 to 390	100.49	3.85	0.00	96

Conducting Analyses with TIMSS Teacher Context Data

This section presents an analysis conducted using the IEA IDB Analyzer with teacher context data from the TIMSS 2023 International Database. Analyses with TIMSS teacher context data seek to make inferences about students whose teachers have a given characteristic, attitude, or instructional practice. Because the teachers in TIMSS do not constitute representative samples of teachers, inferences should not be made about teachers themselves.

As an example of an analysis using teacher context data, Example 7 investigates the relationship between students' achievement and their teachers' formal education. Because this example analyzes teacher context data with student achievement, teacher context data should be merged with student data through the student-teacher linkage file to retrieve the achievement scores and required sample design variables (see earlier section on Merging Data Files with the IEA IDB Analyzer).

Example 7—Analysis of Average Achievement by Teachers' Formal Education

This example involves using the **Percentages and Means** statistic type with the **Use PVs** option to estimate the percentages of students with their average achievement by reporting categories of teachers' level of formal education. The results of this analysis are similar to those presented in Exhibit 5.1.2 of *TIMSS 2023 International Results in Mathematics and Science,* repeated below in Exhibit 1.25. However, Exhibit 5.1.2 of combines the original variable into four categories for reporting and does not report average achievement. Using the **Percentages Only** procedure in the IDB Analyzer would exclude average achievement from the analysis.

↓ Exhibit 1.25: International Results Exhibit of Example 7—Analysis of Average Achievement by Teachers' Formal Education

Before conducting analyses using TIMSS contextual variables, users should refer to the relevant codebook for the data file to identify the appropriate variables related to teachers' education and understand the coding scheme. From the TIMSS 2023 International Database webpage, the Context Questionnaire Variables downloads present all the context questionnaires administered in TIMSS 2023 and the associated variable names under which the data are saved. The National Adaptations Database downloads should also be checked for any national adaptations made to the questionnaire items that may impact international comparability.

The codebook for the teacher context data file indicates that the variable ATBG04 contains information on fourth-grade teachers' level of formal education completed in seven categories:

1. Did not complete <Upper secondary education—ISCED Level 3>

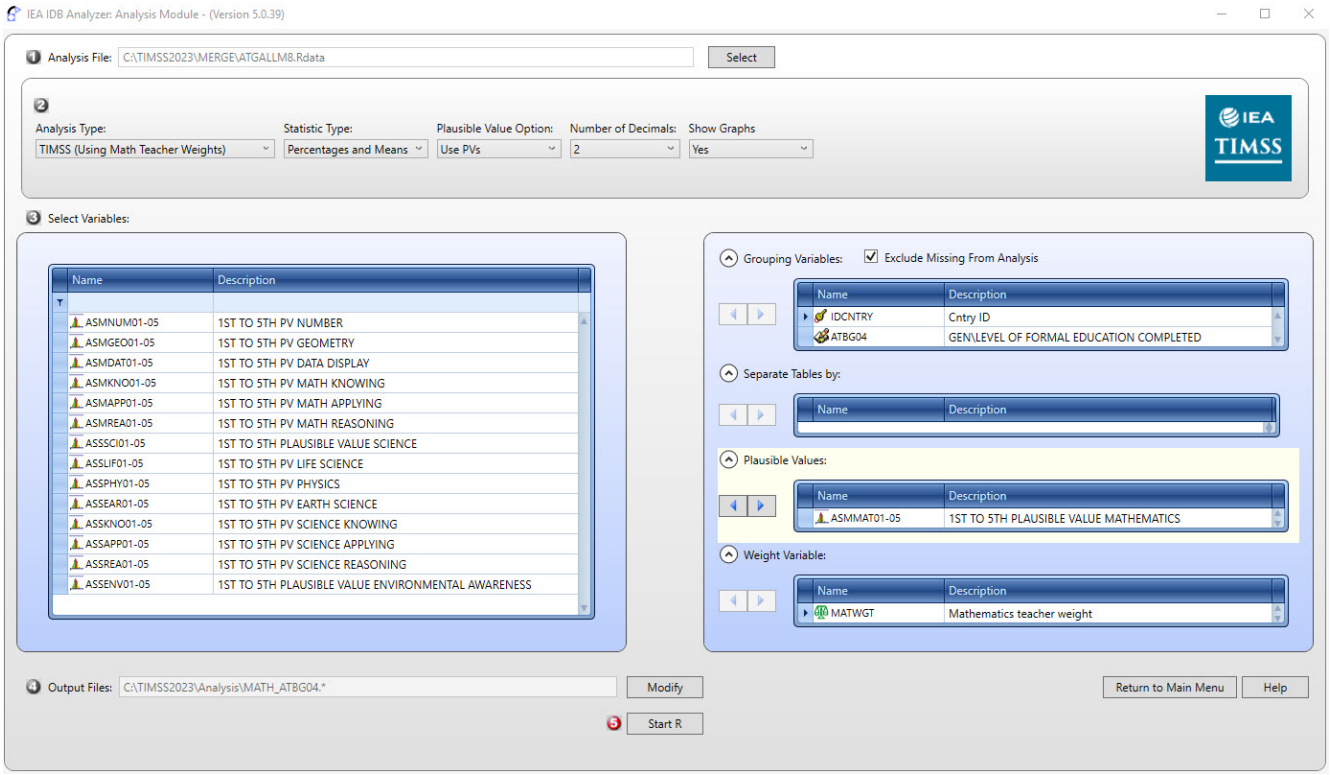
2. <Upper secondary education—ISCED Level 3>
3. <Post-secondary, non-tertiary education—ISCED Level 4>
4. <Short-cycle tertiary education—ISCED Level 5>
5. <Bachelor's or equivalent level—ISCED Level 6>
6. <Master's or equivalent level—ISCED Level 7>
7. <Doctor or equivalent level—ISCED Level 8>

The ATBG04 item was specifically designed to be adapted to each unique education system, indicated by the category labels being contained within brackets in the codebooks, e.g., "<Bachelor's or equivalent level—ISCED Level 6>." Despite the various education system structures and translations across countries, the variable is made internationally comparable by adhering to the International Standard Classification of Education (ISCED; UNESCO, 2012), which provides an internationally accepted classification scheme for describing levels of schooling across countries. For all such variables, National Adaptations Database downloads provide English back translations for the answer categories for each country.

This example uses the merged ATGALLM8 data file described earlier in this chapter in the section on Merging Data Files with the IEA IDB Analyzer. The analysis is conducted in the **Analysis Module** of the IEA IDB Analyzer, shown completed in Exhibit 1.26, using the following steps:

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ATGALLM8 as the **Analysis File** by clicking the **Select** button.
3. Select **TIMSS (Using Math Teacher Weights)** as the **Analysis Type**.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**, because average achievement will be computed by the grouping variable ATBG04.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce three graphs in the output file: a line graph of average achievement for each category of teachers' education level by country, a clustered bar graph of average achievement for each category of teachers' education level by country, and a stacked bar graph of average percent of students for each category of teachers' education level by country. R also produces separate graphs by country.

Exhibit 1.26: IEA IDB Analyzer Analysis Module Setup for Example 7—Analysis of Average Achievement by Teachers' Formal Education



8. Specify the variable ATBG04 as a second grouping variable by first clicking the **Grouping Variables** field to activate it. Then, select ATBG04 from the list of variables in the left panel, and move it to the **Grouping Variables** field by clicking the **right arrow** (▸) button. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (not achievement) variables. See the Help manual for more information.
10. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASMMAT01–05 from the list of available variables in the left panel, and move it to the right **Plausible Values** field by clicking the **right arrow** (▸) button.
11. The **Weight Variable** is selected automatically by the software; MATWGT is selected by default because of the **Analysis Type** selected in Step 3.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer will create an R syntax file (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the

analysis. In Exhibit 1.26, the syntax file MATH_ATBG04.R and the output files with the same name will be created and stored in the C:\TIMSS2023\Analysis folder.

Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** ( ) button (or select **Submit** in the **Run** menu).

This analysis produces the results output files in the same manner as described for Example 2 and Example 5 (step 2), with countries identified in the first column and the second column describing the categories of the analysis variable ATBG04. Exhibit 1.27 shows the results as shown in the R output file with two example countries: Bulgaria and Canada.

Each country's results are displayed on up to seven lines, one for each value of the ATBG04 variable. There are fewer lines if any category does not have any observations. For example, in Exhibit 1.27, Bulgaria has only five lines of results, because no teacher responded in the first two categories. A country with no valid data on this variable would not appear in the results output.

For categories with very few students, standard errors associated with percentages and achievement means may appear unusually high or low because variance estimates and achievement cannot be estimated reliably for small numbers of students. In *TIMSS 2023 International Results in Mathematics and Science*, standard errors and average achievement is not reported for groups of students representing less than 2.5% of a given country.

Exhibit 1.27: R Output for Example 7—Analysis of Average Achievement by Teachers' Formal Education

Analysis for ASMMAT by IDCNTRY ATBG04

Cntry ID	GENOF FORMAL EDUCATION COMPLETED	N of Cases	Sum of MATWGT	Sum of MATWGT (s.e.)	Percent	Percent (s.e.)	ASMMAT (Mean)	ASMMAT (s.e.)	Confidence Interval (95)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	<Post-secondary, non-tertiary education—ISCED Level 4>	9	398	398.45	0.73	0.73	415.51	NaN	NaN to NaN	86.21	NaN	0.00	1
Bulgaria	<Short-cycle tertiary education—ISCED Level 5>	264	3779	903.13	6.97	1.67	509.72	15.65	479 to 540	91.98	7.01	0.00	13
Bulgaria	<Bachelor's or equivalent level—ISCED Level 6>	782	12277	1575.44	22.63	2.90	512.84	11.97	489 to 536	93.74	7.59	0.00	38
Bulgaria	<Master's or equivalent level—ISCED Level 7>	2993	37611	1874.10	69.33	3.19	538.75	4.55	530 to 548	88.15	3.11	0.00	70
Bulgaria	<Doctor or equivalent level—ISCED Level 8>	22	184	184.41	0.34	0.34	597.50	NaN	NaN to NaN	62.59	NaN	0.00	1
Canada	<Upper secondary education—ISCED Level 3> (If you have not completed <postsecondary or tertiary education>, you will b	25	141	140.63	0.06	0.06	470.51	NaN	NaN to NaN	73.42	NaN	0.00	1
Canada	<Post-secondary, non-tertiary education—ISCED Level 4>	18	296	250.59	0.12	0.10	485.39	84.77	319 to 652	72.43	22.90	0.00	2
Canada	<Short-cycle tertiary education—ISCED Level 5>	18	187	132.05	0.08	0.05	391.39	55.82	282 to 501	65.03	28.92	0.00	2
Canada	<Bachelor's or equivalent level—ISCED Level 6>	10534	207360	7032.92	83.95	1.54	503.16	2.29	499 to 508	80.82	1.28	0.00	125
Canada	<Master's or equivalent level—ISCED Level 7>	1886	37040	3616.97	15.00	1.50	511.28	5.66	500 to 522	81.27	2.70	0.00	80
Canada	<Doctor or equivalent level—ISCED Level	89	1985	912.24	0.80	0.37	541.60	24.42	494 to 589	98.78	19.72	0.00	7

References

IBM Corporation. (2024). IBM SPSS Statistics for Windows (version 30.0) [Computer software]. Armonk, NY: IBM Corporation.

International Association for the Evaluation of Educational Achievement (IEA). (2025). IEA IDB Analyzer (Version 5.0) [Computer software]. Hamburg, Germany: IEA Hamburg. Retrieved from https://www.iea.nl/data

Johnson E. G., Rust K. F. (1993). Effective df for variance estimates from a complex sample survey. In *1993 Proceedings: Survey research methods section* (pp. 863–866). American Statistical Association.

Posit Software, PBC formerly RStudio (2025). RStudio: Integrated Development Environment for R (version 2024.09.1+394). Boston, MA, USA: PBC. https://posit.co/products/open-source/rstudio/

R Development Core Team (2025). R: A language and environment for statistical computing (version 4.4.2). Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org

SAS Institute. (2023). SAS for Windows (version 9.4) [Computer software]. Cary, NC: SAS Institute, Inc.

Satterthwaite F. E. (1941). Synthesis of variance. *Psychometrika, 6*(5), 309–316. https://doi.org/10.1007/BF02288586

Satterthwaite F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*(6), 110–114. https://doi.org/10.2307/3002019

Student. (1908). The probable error of a mean. *Biometrika*, *6*(1), 1–25. https://doi.org/10.2307/2331554

UNESCO. (2012). *International Standard Classification of Education (ISCED) 2011*. Montreal, Quebec: UNESCO Institute of Statistics. Retrieved from http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

von Davier, M. (2025). An Improved Satterthwaite (1941, 1946) Effective df Approximation. Journal of Educational and Behavioral Statistics, 0(0). https://doi.org/10.3102/10769986241309329

Welch B. L. (1947). The generalization of “Student’s” problem when several different population variances are involved. *Biometrika, 34*(1/2), 28–35. https://doi.org/10.2307/2332510

CHAPTER 2

Contents and Structure of the TIMSS 2023 International Database

Overview

This chapter describes the contents of the TIMSS 2023 International Database. The various data files and related materials included in the database are described, in accordance with the structure summarized in Exhibit 2.1. Descriptions of data files include the conventions for naming the various file types and variables as well as codes for missing values.

The TIMSS 2023 International Database and all supplemental materials are available on the Boston College, TIMSS & PIRLS International Study Center website: https://timss2023.org/data, and through IEA's Data Repository.

Exhibit 2.1: Contents of the TIMSS 2023 International Database

User Guide	TIMSS 2023 User Guide for the International Database
Data Files	
R Data	TIMSS 2023 student, process, home, school, and teacher R data files
SPSS Data	TIMSS 2023 student, process, home, school, and teacher SPSS data files
SAS Data	TIMSS 2023 student, process, home, school, and teacher SAS data files
Curriculum Data	TIMSS 2023 Curriculum Questionnaire data Excel files
TCMA Data	TIMSS 2023 Test-Curriculum Matching Analysis data Excel files
Supplemental Material	
Codebooks	List describing all variables in the TIMSS 2023 R, SPSS, and SAS data files
Data Almanacs	Summary statistics for all TIMSS 2023 achievement items and context variables
Achievement Item Information	TIMSS 2023 achievement item information
Achievement IRT Parameters	IRT item parameters for TIMSS 2023 achievement items
Context Questionnaire Variables	International versions of the TIMSS 2023 Context Questionnaires
National Adaptations Database	National adaptations to the TIMSS 2023 Context Questionnaires
Derived Context Variables	Variables derived from the student, home, teacher, and school context data
Special Programs	R, SPSS, and SAS programs to score achievement items

The TIMSS 2023 International Database contains data for student assessment results in mathematics and science, process data, and extensive contextual information collected

from participating students as well as their parents, their teachers, and their school principals. Additional data collected from TIMSS National Research Coordinators highlight the national contexts of participating countries, focusing on education policies and mathematics and science curricula.

Exhibit 2.2 lists the TIMSS 2023 countries and benchmarking participants with data in the International Database, along with their alpha and numeric codes used to identify data files and records within the data files. Nearly all participating countries and benchmarking systems administered TIMSS in the digital format. A small number of countries administered a paper version of the assessment containing only assessment material from TIMSS 2019. At the fourth grade, a less difficult paper assessment option was offered with easier mathematics material for countries that requested this option. Results for the different versions of the assessments were linked and reported on the same TIMSS achievement scales. Data file structure is the same for the different options.

↓ Exhibit 2.2: Countries in the TIMSS 2023 International Database

Restricted-Use Version of the TIMSS 2023 International Database

There are two versions of the International Database. The public-use version is available for immediate access from the TIMSS 2023 International Database webpage and the IEA Study Data Repository. A number of variables were removed from the public use version in order to minimize the risk of disclosing confidential information. Exhibit 2.3 lists the variables removed from the public-use version of the TIMSS International Database that are available in the restricted-use version.

Exhibit 2.3: Variables Available in the Restricted-Use Version of the TIMSS 2023 International Database

Variable	Description
ITBIRTHY / ITBIRTHM	Students' year and month of birth from the tracking forms
ITDATE	TIMSS 2023 testing date from the tracking forms
ITMODE_x	Mode of administration for the TIMSS assessments and context questionnaires
ITDEV	Type of device used for the TIMSS assessment
ASBG02A / ASBG02B BSBG02A / BSBG02B	Students' year and month of birth from the student questionnaire
ACBG01 / BCBG01	Total school enrollment from the school questionnaire
ACBG02 / BCBG02	School enrollment in the target grade from the school questionnaire

Users who require access to the restricted-use version should contact IEA through its Data Repository to obtain permission. Note that the variable reporting students' age remains available in the public-use version of the International Database.

TIMSS 2023 Data Files

The TIMSS 2023 International Database includes data collected from all instruments administered to the students and their parents, school principals, and teachers. This includes the scored responses to the achievement items, assessment process data for screen-level time and visits, and the responses to the student, home, school, and teacher context questionnaires. These data files also include the achievement scores estimated for participating students, as well as context variables derived for reporting in *TIMSS 2023 International Results in Mathematics and Science*. National Research Coordinators' responses to the TIMSS 2023 Curriculum Questionnaire and Test-Curriculum Matching Analysis (TCMA) also are part of the International Database and are described later in this chapter.

The next few sections describe the format and contents of the TIMSS 2023 data files. With the exception of the curriculum data files and TCMA data files, the TIMSS 2023 data files are provided in R format (*.Rdata), SPSS format (*.SAV), and SAS format (*.SAS7BDAT).

The file names given to the various data file types in TIMSS 2023 are shown in Exhibit 2.4. These are named with the suffix "M8" corresponding to the eighth cycle of TIMSS. Fourth-grade data files are named beginning with "A" and eighth-grade data files are named beginning with "B." The second and third characters indicate the data file type, and the fourth through sixth characters indicate the country according to ISO 3166 alpha code (↓ Exhibit 2.2). For example, ASGKORM8.Rdata is an R file that contains Korea's TIMSS 2023 fourth-grade student context questionnaire data.

For each file type, a separate data file is provided for each participating country. All data files and the variables they contain are described in the following sections, beginning with the student achievement data files.

Exhibit 2.4: TIMSS 2023 Data File Names

Assessment	File Name	Description
Grade 4	ACG●●●M8	School context data files
	ASA●●●M8	Student achievement data files
	ASP●●●M8	Student process data files
	ASR●●●M8	Within-country scoring reliability data files
	ASG●●●M8	Student context data files
	ASH●●●M8	Home context data files
	AST●●●M8	Student-teacher linkage files
	ATG●●●M8	Teacher context data files
Grade 8	BCG●●●M8	School context data files
	BSA●●●M8	Student achievement data files
	BSP●●●M8	Student process data files
	BSR●●●M8	Within-country scoring reliability data files
	BSG●●●M8	Student context data files
	BST●●●M8	Student-teacher linkage files
	BTM●●●M8	Mathematics teacher context data files
	BTS●●●M8	Science teacher context data files

●●● = Three-character country abbreviation based on the ISO alpha coding scheme (see Exhibit 2.2)

Student Achievement Data Files (ASA/BSA)

The TIMSS 2023 student achievement data files contain the scored responses to the individual achievement items in the TIMSS 2023 assessments. The student achievement data files are used for performing item-level analyses. Achievement variables (plausible values) for all of the TIMSS 2023 achievement scales and sampling and weighting variables are available in the student achievement data files.

As described in the TIMSS 2023 Assessment Design, students who participated in TIMSS digitally were administered one of 14 assessment booklets, each containing sets of mathematics and science items. Students in countries who administered a paper-based option (including less difficult at the fourth grade) were administered one of eight booklets consisting only of trend items from TIMSS 2019.

The TIMSS 2023 student achievement data files include variables for all TIMSS digital items and all TIMSS paper items. The fourth-grade ASA files also include variables for the less difficult mathematics items. However, countries only have responses to variables corresponding to their participation.

In all the booklets administered as part of TIMSS 2023, some of the items were multiple-choice and some were constructed-response. The student achievement data files

contain the actual responses to the multiple-choice questions (e.g., where 1 corresponds to the first response option, 2 to the second option, and so on), and the score codes assigned to the constructed-response items based on the TIMSS 2023 scoring guides.

Achievement Item Variable Naming Convention

The achievement item variable names are based on a nine-character alphanumeric code, which observes the following rules:

- The first character is either "M" for mathematics items, or "S" for science items.
- The second character indicates the assessment mode or type. The letter, "E" indicates digital, "Q" indicates digital for items in TIMSS Problem-Solving and Inquiry (PSI) tasks (see Chapter 1 in *TIMSS 2023 Technical Report*), "P" indicates paper, and "N" indicates less difficult paper mathematics.
- The third character indicates the assessment cycle when the item was first presented in TIMSS. The code "1" was used for items introduced in TIMSS 1995. The items in the TIMSS 2023 assessment have "6" for items introduced in 2015, "7" for items in 2019, or "8" for new items in 2023. The less difficult paper mathematics items have the code "1" for items introduced in 2015 or "2" for items developed in 2019.
- The fourth character is either "1" for fourth grade items or "2" for eighth grade items.
- The fifth through seventh characters represent a unique three-digit number used to identify the items.
    - TIMSS PSI items have a letter identifier for the fifth character corresponding to the particular PSI task, followed by two numeric digits. In fourth-grade mathematics, the letter "A" is used for the *Little Penguins* PSI, "B" for the *Amazon River*, "R" for a set of *Robots* items, and "Z" for *Amazing Bamboo*. In fourth-grade science, the letter "R" is used for *Earth's Motions* and "S" for *Sugar Experiment*. In eighth-grade mathematics, "D" is used for *Dinosaur Speed*, "C" for the *Singing Competition*, and "N" for *Tennis Balls*. In eighth-grade science, "T" is used for *Earth's Motions*, "L" for *Elephants*, and "S" for *Sunken Ship*.
- The eighth character indicates the item part and appears only when required. It is generally a letter from "A" to "H," depending on how many parts there are to a particular item.
- A ninth character is present when an item part itself consists of further parts.

As an example of an item variable name, MQ71A04A is the first part (part A) of a fourth-grade mathematics PSI TIMSS item introduced in 2019 and whose sequential number in the task is 4.

Achievement Item Response Codes

A series of conventions were adopted to code the item responses included in the student achievement data files. The value assigned to each item response depends on the item format.

For multiple-choice items, numerical values 1, 2, 3, 4, etc., are used to correspond to the response options A, B, C, D, etc., respectively. For these items, the correct response key is included in the item information files and as part of the variable label in the achievement codebook files (described in a later section). R, SPSS, and SAS programs are included as part of the TIMSS 2023 International Database to derive correctness scores (score points) for these items based on their item response codes and key (see Chapter 3).

Each constructed-response item has its own scoring guide that uses a one- or two-digit coding scheme. Items developed in TIMSS 2019 and earlier cycles use the two-digit scheme, and items developed in TIMSS 2023 and later use the one-digit scheme.

The two-digit coding scheme uses two-digit codes to characterize the response. The first digit of the code refers to the number of score points given to the response. For correct or partially correct responses, the first digit is 1 for one-point responses or 2 for two-point responses. For an incorrect response, the first digit is 7. The second digit of the score provides diagnostic information for correct and incorrect responses, such as indicating a specific method used to solve the problem or to track a common student misconception or error. An incorrect response not fitting a pre-defined incorrect code is given a 79 for “other incorrect.” If no diagnostic categories are defined, all incorrect responses receive code 79. A special two-digit code, 99, is used for responses left completely blank.

In the one-digit coding, the 0 is used for incorrect responses. For correct or partially correct responses, 1 is used for one-point responses and 2 is used for two-point responses. For some items, code 7 can be assigned to an incorrect response with a notable misconception or to an incorrect response with some elements of a correct answer which is nonetheless wrong. Code 9 is used for responses left completely blank.

Chapter 3 describes the programs that convert the two-digit response codes into correctness scores.

Derived Achievement Item Variables

For some items, students were asked to provide more than one answer or a multiple-part answer, each one being scored separately. The pattern of responses across these item

parts determined the score on the item as a whole. These multi-part items can be worth 1 or 2 score points, according to their scoring guide. A list of all derived item variables in the TIMSS 2023 assessments and the rules for awarding item response codes is provided in Appendix 10G of *TIMSS 2023 Technical Report.*

The schemes for naming and assigning response codes for derived items follow the same conventions for constructed-response items described in the previous sections on Achievement Item Variable Naming Convention and Achievement Item Response Codes. The total score for the item is contained in a derived variable, identified by the word "DERIVED" in the item label.

For example, derived item variable SP71009 contains the combined score for its six parts labeled SP71009A through SP71009F. This derived item response was assigned 2 score points (complete comprehension) if all six parts were answered correctly, 1 score point (partial comprehension) if four or five parts were answered correctly, and 0 points otherwise.

Codes for Missing Values on Achievement Items

A subset of values was reserved for specific item response codes related to different categories of missing data. It is recommended that users read this section with particular care, as the way in which these missing codes are used may have implications for subsequent analyses.

Not Administered Response Codes (R: NA; SPSS: SYSMIS; SAS: .A)

Special codes were given to items that were *Not Administered* to distinguish these cases from data that were missing due to student non-response. In general, the Not Administered code was used when an item was not administered, either by design—based on the rotation of items across the assessment booklets—or unintentionally when an item was misprinted or otherwise unavailable for a student to respond, including for technical reasons. The Not Administered code was used in the following cases:

- Assessment item not assigned to the student
- Student absent from the session when the item was administered
- Item misprinted, mistranslated, or deleted due to poor psychometric functioning

Omitted or Invalid Response Codes (R: 9, 99; SPSS: 9, 99; SAS: .)

Omitted or invalid response codes were used for items that a student should have answered but did not. The omitted response code was also used when two or more response options were checked for a paper multiple-choice item.

Not Reached Response Codes (R: 6, 96; SPSS: 6, 96; SAS: .R)
An assessment item was considered *Not Reached* when, within part 1 or part 2 of a booklet, the item itself and the item immediately preceding it were not answered, and no other items were answered in the remainder of that part of the booklet.

Missing Values in R
When using R data files (*.Rdata) without the IDB Analyzer for analysis, it is crucial to understand how missing data is represented and handled. In R, there is only one true native missing code: *NA*. However, the R data files were originally created from SPSS files, where multiple user-defined missing codes can be represented as specific numeric values. For example, in the SPSS files, missing codes for "omitted" responses are often represented as 9, 99, etc., while "not reached" codes in the achievement files are represented as 6, 96, and so on.

If you analyze the R data files directly in R without accounting for these user-defined missing codes, your estimates will likely be incorrect. This is because R interprets these values as valid numeric data rather than missing values. To ensure accurate analysis in R you must recode all numerical missing codes to R's native missing code, NA, before proceeding with your analysis. To determine which numeric values have been defined as missing values, you can check directly in your R data file by executing the following command line:

attr(<your R data file>$<variable of interest>, "na_values")

Alternatively, you can consult the "Missing Scheme Detailed (SPSS)" column in the codebook. This resource provides the necessary details for correctly identifying and handling missing data. This should also be taken into account when you add variables to your R data files.

Modifications to Achievement Item Data

Some modifications were applied to the achievement item data based on quality checks and psychometric analyses conducted for the TIMSS 2023 psychometric scaling. For example, a 2-point item may have been collapsed into a 1-point item, or the data for a particular country may have been deleted. The analyses conducted and the criteria for making these changes are described in Chapter 10 of *TIMSS 2023 Technical Report*. Appendix 10F includes a list of all item modifications reflected in the TIMSS 2023 International Database files.

TIMSS Achievement Variables

TIMSS achievement scales were produced for overall mathematics and science as well as for the content and cognitive subdomains defined in the *TIMSS 2023 Assessment Frameworks*.

The TIMSS 2023 achievement scales are listed in Exhibit 2.5.

Exhibit 2.5: TIMSS 2023 Achievement Scales

Assessment	Achievement Scales				
Grade 4	Overall	MAT	Mathematics	SCI	Science
	Content Domains	NUM	Number	LIF	Life Science
		GEO	Measurement and Geometry	PHY	Physical Science
		DAT	Data	EAR	Earth Science
	Cognitive Domains	KNO	Knowing	KNO	Knowing
		APP	Applying	APP	Applying
		REA	Reasoning	REA	Reasoning
	Other			ENV	Environmental Knowledge
Grade 8	Overall	MAT	Mathematics	SCI	Science
	Content Domains	NUM	Number	BIO	Biology
		ALG	Algebra	CHE	Chemistry
		GEO	Geometry and Measurement	PHY	Physics
		DAT	Data and Probability	EAR	Earth Science
	Cognitive Domains	KNO	Knowing	KNO	Knowing
		APP	Applying	APP	Applying
		REA	Reasoning	REA	Reasoning
	Other			ENV	Environmental Knowledge

For each achievement scale, the TIMSS 2023 International Database provides five imputations of achievement. The five imputations are known as "plausible values," and the variability between them for any given individual reflects the uncertainty in the estimation. A detailed description of the TIMSS 2023 scaling approach and how these achievement scales were created is available in *TIMSS 2023 Technical Report*.

The plausible values for any given scale are the best available measures of student achievement on that scale in the TIMSS 2023 International Database and should be used as the outcome measure in any study of student achievement. It is important to note that these plausible values are not suitable measures of individual student achievement. Plausible values can be analyzed readily using the IEA IDB Analyzer.

The plausible value variable names are based on an eight-character alphanumeric code, which adheres to the following rules:

- The first character is either "A" for a fourth-grade score, or "B" for an eighth-grade score.
- The second character is always "S" to indicate it is a student score variable.

- The third character is either “M” for a mathematics score, or “S” for a science score, whether it is an overall score, a content domain score, or a cognitive domain score.
- The fourth through sixth characters are a three-character code describing the achievement scale, as shown in Exhibit 2.5.
- The seventh character is a 0.
- The eighth character indicates the plausible value: 1, 2, 3, 4, or 5.

For example, ASMAPP01 is the first plausible value on the fourth-grade mathematics applying cognitive domain achievement scale.

TIMSS International Benchmarks of Achievement

To help users of the TIMSS data understand what performance on the overall mathematics and science achievement scales signifies in terms of the mathematics and science students know and can do, TIMSS identified four points on the overall mathematics and science scales to serve as International Benchmarks. As shown in Exhibit 2.6, the TIMSS International Benchmark scores are 625, 550, 475, and 400, which correspond to the Advanced International Benchmark, the High International Benchmark, the Intermediate International Benchmark, and the Low International Benchmark, respectively.

Exhibit 2.6: TIMSS International Benchmarks of Mathematics and Science Achievement

Scale Score	International Benchmark
625	Advanced International Benchmark
550	High International Benchmark
475	Intermediate International Benchmark
400	Low International Benchmark

TIMSS uses a technique known as *scale anchoring* to summarize and describe student achievement at these four points on the scale (see Chapter 14 of *TIMSS 2023 Technical Report*). The *TIMSS 2023 International Results in Mathematics and Science* report presents the results of this scale anchoring and reports the percentage of students in each country reaching each of the TIMSS International Benchmarks.

The TIMSS 2023 International Database contains a set of variables indicating which International Benchmark the students have reached based on each of the five plausible values. The International Benchmark variables follow the achievement score variable naming convention where the fourth through sixth positions have the letters “IBM.” Thus, **ASMIBM01–05** are the five benchmark variables for fourth-grade overall mathematics, **ASSIBM01–05** the five benchmark variables for fourth-grade overall science, **BSMIBM01–**

05 for eighth-grade overall mathematics, and **BSSIBM01–05** for eighth-grade overall science. The codes defined for the benchmark variables are described in Exhibit 2.7.

Exhibit 2.7: TIMSS International Benchmark Variable Codes

Code	Description
1	Performed below the Low International Benchmark
2	Performed at or above the Low International Benchmark, but below the Intermediate Benchmark
3	Performed at or above the Intermediate International Benchmark, but below the High Benchmark
4	Performed at or above the High International Benchmark, but below the Advanced Benchmark
5	Performed at or above the Advanced International Benchmark

Nonresponse Indicator Variables

The TIMSS 2023 student achievement data files include several variables indicating whether students had omitted or not-reached item responses for different subgroups of items based on item type and TIMSS mathematics or science subdomain. These nonresponse indicator variables were utilized in generating plausible values of student achievement. The nonresponse indicator variables use the numerical value 1 if the student answered all items in the subset, or 0 if the student had at least one missing item response in the subset. The nonresponse indicator variable names are based on an eight-character scheme as follows:

- The first character is either “M” for mathematics items, or “S” for science items.
- The second character indicates the assessment mode or type of the group of items. The letter, “E” indicates the digital TIMSS assessment. The paper assessment options did not use these variables.
- The third character indicates the assessment cycle when the variable was first presented in TIMSS. The code “8” indicates that this is a new variable for 2023.
- The fourth character is either “1” for fourth-grade items or “2” for eighth-grade items.
- The fifth through seventh characters indicate the TIMSS subdomain for the group of items according to the *TIMSS 2023 Assessment Frameworks*. The three characters correspond to the subdomains according to the achievement scales listed in Exhibit 2.5. Additionally, a nonresponse indicator ending with “NEN” was added which covers all non-environmental items.

For more information about how these variables were used in achievement scaling, see Chapter 12 of *TIMSS 2023 Technical Report*.

Student Process Data Files (ASP/BSP)

The TIMSS 2023 student process data files contain variables associated with students' screen navigation in the digital TIMSS 2023 assessment. Data for derived process variables are available only for countries that participated in the digital assessment.

The TIMSS 2023 process data files include three types of process variables associated with the individual screens in the assessment:

- Total time on screen (seconds)
- Time on first screen visit (seconds)
- Number of screen visits (frequency)

Codebooks published with the database list all derived variables included in the data files and describe codes for missing values in the process data. The process variables for each screen are named according to the item variable naming convention (described in the previous Achievement Item Variable Naming Convention section for the student achievement data files), followed by the suffix "_S" for total time, "_R" for time on first visit, and "_F" for the number of visits.

Typically, a screen displays a single item and thus will have the same ID as the item shown. However, there are occasional screens that display more than one item sharing a common stem but requiring separate responses. In these instances, the variable names consist of the first seven common characters of the corresponding item names. For example, ME72106A, ME72106B, and ME72106C are three fourth-grade mathematics items with separate responses but share the same stem and are shown on the same screen. The process data variable names associated with that screen are ME72106_S for total time spent on screen, ME72106_R for time spent on first screen visit, and ME72106_F for the frequency of screen visits. PSI screens may not contain any items and use a different naming convention. The Achievement Item Information downloads (see later section) list the screen ID associated with each item ID. The screen ID is always the basis of the process variable name.

The student process data variables were derived from the raw student response data files generated through the TIMSS 2023 digital assessment. RM Results was responsible for the development of the TIMSS 2023 assessment software, including the production of raw data files. The raw event log data files contain timestamped records of students' interactions with the digital assessment. For the generation of derived process variables, specific events corresponding to the screen navigation were used. The data in the student process data files were cleaned for inconsistencies, outliers, and any known technical issues in the production of the raw data files.

In addition to the derived process variables, the student process data files include identification and tracking data for each student (see later sections on Identification

Variables and Tracking Variables), and additional tracking and indicator variables about student participation and cleaning of the process variables. Variables ITPARTPT1 and ITPARTPT2 indicate whether the student has valid item responses (non-missing) in the respective part of the assessment. Process data is not available for a part if no valid item responses are available. Variables FLAGPT1 and FLAGPT2 indicate whether the student had process data removed in a specific part of the assessment. More information about the codes used in these variables can be found in the codebooks corresponding to the data files.

Within-Country Scoring Reliability Data Files (ASR/BSR)

The TIMSS within-country scoring reliability data files contain data that can be used to investigate the reliability of the item scoring for human-scored constructed-response items. The scoring reliability data files contain one record for each student whose responses to constructed-response items, in whole or in part, were double scored during the within-country scoring reliability exercise (see Chapter 7 and Chapter 10 in *TIMSS 2023 Technical Report*). For each constructed-response item requiring human scoring, the following three variables are included in the scoring reliability data files:

- **Original Score:** the score assigned by the first scorer and also present in the student achievement files
- **Second Score:** the score assigned by the second scorer and present only in the scoring reliability files
- **Score Agreement:** a dichotomous variable indicating agreement between the two scorers

In the student achievement data files (ASA/BSA), the variable ILRELIAB indicates whether the students' responses were included for scoring reliability (1) or not included (0).

It should be noted that the Second Score data were used only to evaluate within-country scoring reliability and were not used in computing the achievement scores included in the International Database and presented in *TIMSS 2023 International Results in Mathematics and Science*.

Scoring Reliability Variable Naming Convention

The variable names for the Original Score, Second Score, and Score Agreement variables are based on the same naming convention as for the achievement item variables discussed earlier. The Second Score and Score Agreement variables have one more character added to the Original Score variable as follows:

- The **Original Score** variable follows the item variable naming convention described earlier. The second character has letter "E" for a digital item, the

letter “Q” for a digital PSI item, the letter “P” for a paper item, or the letter “N” for a less difficult mathematics item.

- The **Second Score** variable has the letter “R” added to the Original Score variable after the second character (e.g., MER81033).
- The **Score Agreement** variable has the letter “I” added to the Original Score variable after the second character (e.g., MEI81033).

Scoring Reliability Codes

The values contained in both the Original Score and Second Score variables are the scores assigned using the TIMSS scoring guides. The Score Agreement variable may have one of three values, depending on the degree of agreement between the score codes assigned by the two scorers, as described in Exhibit 2.8. The earlier section on Achievement Item Response Codes describes how codes in the scoring guides translate to score points (levels).

Exhibit 2.8: TIMSS 2023 Score Agreement Variable Codes

Code	Description
0	Identical codes and score levels
1	Identical score levels, but different diagnostic codes
2	Different score levels

Context Data Files

This section describes the TIMSS 2023 context data files and the conventions for naming the various files and variables and coding the data. There are six types of context data files: the first four context data files correspond to the four types of context questionnaires administered in TIMSS 2023 (student, home, school, and teacher); the fifth data file serves to link the student and teacher context data; and the sixth data file corresponds to the TIMSS 2023 Curriculum Questionnaire administered to the National Research Coordinators of each participating country. The Curriculum Data are provided separately from the other data files in the International Database. Comprising mostly descriptive, narrative information about the national education systems, these data are provided in Excel format.

The four context data files for student, home, school, and teacher questionnaires contain the responses to the questions asked in their respective context questionnaires, plus some additional derived context variables used for reporting, as well as variables for the context questionnaire scales. Home context data is only available at the fourth grade. At the eighth grade, teacher data is in separate files for mathematics teachers and science teachers.

Student Context Data Files (ASG/BSG)

All participating students were administered a context questionnaire with questions related to their home context, school experiences, and attitudes toward mathematics and science. Regardless of whether they were participating in the digital or paper assessment, students were administered a questionnaire at the end of their testing session. The student context data files contain students' responses to all of these questions. They also contain students' mathematics and science achievement variables (plausible values) to facilitate analyses of relationships between student characteristics and achievement.

Two versions of the student questionnaire were administered at the eighth grade. One version is for educational systems where science is taught as an integrated subject (General/Integrated Science version). The other version is for educational systems where the sciences—biology, physics, chemistry, and earth science—are taught as separate subjects (Separate Science version). For eighth-grade students who were administered the General/Integrated Science version, questions that were given only in the Separate Science version were coded as "Not Administered." For students who were assigned the Separate Science version, questions that were asked only in the General/Integrated Science version were coded as "Not Administered." At the fourth grade, there was a single version of the student questionnaire, tailored for general science.

The student context data files also contain a number of identification, tracking, sampling, and weighting variables, and derived variables that were used for producing exhibits in *TIMSS 2023 International Results in Mathematics and Science*. These variables are described later in the section on Structure and Design Variables..

Home Context Data Files (ASH)

For countries that participated at the fourth grade, TIMSS 2023 included a home questionnaire, which was completed by the students' parents or guardians. It asked questions about preparations for primary schooling, including attendance in preschool and numeracy and literacy activities in the home before the child began school, such as reading books, singing songs, writing letters or words, or adding and subtracting. Parents answered questions about home resources, their highest level of education, and their employment status. Analyzing data from the home questionnaire requires that the home context data files be merged with the student context data files using the country and student identification variables (see later section on Structure and Design Variables). Details of the merging procedure with the IEA IDB Analyzer are described in Chapter 1 of this User Guide.

School Context Data Files (ACG/BCG)

The school context data files contain principals' responses to the questions in the TIMSS 2023 School Questionnaire. Although school-level analyses where the schools are the units of analysis can be performed, it is preferable to analyze school-level variables as

attributes of students. To perform student-level analyses with school data, the school context data files must be merged with the student context data files using the country and school identification variables.

Teacher Context Data Files (ATG/BTM/BTS)

The mathematics and science teachers of the students that were sampled in TIMSS 2023 were administered at least one questionnaire with questions pertaining to their teaching context and attitudes and their teaching practices in the classes of the sampled students. Each teacher was asked to respond to a questionnaire for each class taught that contained sampled students. The teacher context data files contain one record for each of the classes taught either by a mathematics or a science teacher. If a teacher taught more than one class, they were expected to complete multiple questionnaires, but to respond only once to general context questions and multiple, separate class- or subject-specific questions for each class they taught. In some cases, although the teacher responded to more than one questionnaire, responses to only one were obtained. In these cases, there were as many records entered in the teacher context data file as sampled classes were taught by the teacher, and the context information in part A from the completed questionnaire was entered into these teacher records.

There were two types of teacher questionnaires administered at the eighth grade: one for the mathematics teachers, and one for the science teachers. As described in Exhibit 2.4, the responses of teachers to the mathematics questionnaire are found in the BTM files and the responses of teachers to the science questionnaire are found in the BTS files. At the fourth grade, the situation was more straightforward, with a single teacher questionnaire requesting information on both mathematics and science, and all teachers' responses are found in the ATG files.

In the teacher context data files at both grades, each teacher has a unique identification number (IDTEACH) and a link number (IDLINK) that is specific to the sampled class taught by the teacher and to which the information in the data record corresponds. The IDTEACH and IDLINK combination uniquely identifies, within a country, a teacher teaching a specific class. The IDTEALIN variable is a concatenation of IDTEACH and IDLINK. Students linked to teachers identified by the same IDTEACH but different IDLINK are taught by the same teacher but in different classes. The teacher context data files cannot be merged directly with the student data files and they do not contain sampling and weighting information, nor achievement scores. The student-teacher linkage data files, described next, serve that purpose.

It is important to note that the teachers in the teacher context data files do not constitute a representative sample of teachers in a country, but rather are the teachers who taught a representative sample of students. The teacher data, therefore, should be

thought of as attributes of the students to which they are linked, and should be analyzed only in conjunction with the student-teacher linkage data files.

Student-Teacher Linkage Data Files (AST/BST)

The TIMSS student-teacher linkage data files contain information required to link the student and teacher data files. The student-teacher linkage data files contain one entry per student-teacher linkage combination in the data. For instance, if a student has three teachers, the file has three entries corresponding to that student. The sole purpose of the student-teacher linkage data files is to link students with their teacher-level data to perform appropriate student-level analyses where teacher characteristics are considered as attributes of the students. The student-teacher linkage data files also include sampling and weighting information and achievement scores to facilitate the analyses of teacher data.

Curriculum Data

The TIMSS 2023 Curriculum Data contain the responses provided by the National Research Coordinators of the participating countries to the TIMSS 2023 Curriculum Questionnaire. The data are available to download in Excel files for each grade from the TIMSS 2023 International Database webpage. Separate tabs contain responses for general (non-subject specific) questions, mathematics questions, and science questions, respectively.

Context Variable Naming Convention

The context variable naming convention for the variables in the student, home, school, and teacher context data files uses a seven- or eight-character string. The following rules are applied in naming the context variables:

- The first character is always “A” for fourth-grade data, or “B” for eighth-grade data.
- The second character indicates the type of respondent. The letter “C” identifies data from the school principals, the letter “T” for teacher data, and the letter “S” for student and parent data.
- The third character is used to indicate the source of the data. The letter “B” is used for all context variables reporting responses to the context questionnaires. The letter “D” is used for variables derived from responses in the context questionnaires. In addition, the letter “B” is used for the Rasch scores to context questionnaire scales derived from questionnaire data, and the letter “D” is used for the categorical (index) variables constructed from these context questionnaire scale Rasch scores (see Chapter 15 in *TIMSS 2023 Technical Report*).
- The fourth character is used to indicate the subject or topic to which a context question refers. The following letters are used:

G—General questions (not subject specific)

H—Home questionnaire questions

M—Questions related to mathematics

S—Questions related to science

B—Questions related to biology

C—Questions related to chemistry

E—Questions related to earth science

P—Questions related to physics

- The fifth through eighth characters of all context variables represent the sequential numbering of the questions as presented in their respective questionnaires.

The curriculum data files follow their own variable naming convention whereby the first three characters of a variable name are as follows:

GEN—General questions (not subject-specific)

MA4—Questions related to fourth-grade mathematics

SC4—Questions related to fourth-grade science

MA8—Questions related to eighth-grade mathematics

SC8—Questions related to eighth-grade science.

The remaining characters in the curriculum data variable naming convention refer to the question location, as summarized in Exhibit 2.9.

Exhibit 2.9: TIMSS 2023 Context Variable and Question Location Naming Convention

Questionnaire	Question Location Name	Context Variable Name	Description
Student Questionnaire	SQG-●●●	ASG●●●	Grade 4 general questions
		BSG●●●	Grade 8 general questions
	SQMS-●●●	ASBM●●●	Grade 4 mathematics questions
		ASBS●●●	Grade 4 science questions
	SQM-●●●	BSBM●●●	Grade 8 mathematics questions
	SQIS-●●●	BSBS●●●	Grade 8 integrated science questions
	SQSS-●●● (Grade 8 Separate science version)	BSBS●●●	Grade 8 general science questions
		BSBB●●●	Grade 8 biology questions
		BSBC●●●	Grade 8 chemistry questions
		BSBE●●●	Grade 8 earth science questions
		BSBP●●●	Grade 8 physics questions
Home Questionnaire	HQ-●●●	ASH●●●	Grade 4 all questions
School Questionnaire	ScQ-●●●	ACG●●●	Grade 4 all questions
		BCG●●●	Grade 8 all questions
Teacher Questionnaire	TQG-●●●	ATG●●●	Grade 4 general questions
		BTG●●●	Grade 8 general questions
	TQM-●●●	ATM●●●	Grade 4 mathematics questions
		BTM●●●	Grade 8 mathematics questions
	TQS-●●●	ATS●●●	Grade 4 science questions
		BTS●●●	Grade 8 science questions
Curriculum Questionnaire	CQG-●●●	GEN●●●	General questions
	CQM4-●●●	MA4●●●	Grade 4 mathematics questions
	CQS4-●●●	SC4●●●	Grade 4 science questions
	CQM8-●●●	MA8●●●	Grade 8 mathematics questions
	CQS8-●●●	SC8●●●	Grade 8 science questions

●●● = Sequential numbering of the question location in the questionnaire

Context Question Location Convention

The context variable naming convention indicates explicitly the ordering of questions in the context questionnaires. Each question also was assigned a unique location code. This unique code includes the sequence number of the question within the questionnaire—the same sequence number found in the question's variable name—appended to a three-character string corresponding to the questionnaire source as shown in Exhibit 2.9. For example, if the question location is given as SQG-08A, it refers to part A of general question 8 in the student context questionnaire. The data for this question is stored in the student context data files (ASG) under variable name ASBG08A. This convention is followed in the codebooks, the data almanacs, and in the description of the variables

included in the Context Questionnaire Variables download, all available on the TIMSS 2023 International Database webpage.

Context Variable Response Codes

The values assigned to each of the context variables depend on the item format and the number of options available. For categorical questions, sequential numerical values are used to correspond to the response options available. The numbers correspond to the sequence of appearance of the response options. For example, the first response option is represented with a 1, the second response option with a 2, etc. Open-ended questions such as "How many students are in this class?" are coded with the actual number given as a response.

Codes for Missing Values on Context Items

A subset of values was reserved for specific item response codes related to different categories of missing data. It is recommended that users read this section with particular care, as the way in which these missing codes are used may have implications for subsequent analyses.

Not Administered Response Codes (R: NA; SPSS: SYSMIS; SAS: .A)
Special codes were given to items that were *Not Administered* to distinguish these cases from data that were missing due to non-response. In general, the Not Administered code was used when an entire questionnaire was not completed. The Not Administered code also was used in the following cases:

- Question was removed from a country's national questionnaire
- Question left out, misprinted, or mistranslated the therefore deemed not comparable across countries

The National Adaptations of the Context Questionnaires downloads available from the TIMSS 2023 International Database website reports all instances of questions removed or questions modified such that the data were not internationally comparable.

Omitted or Invalid Response Codes (R: 9, 99, 999, ...; SPSS: 9, 99, 999, ...; SAS: .)
Omitted or invalid response codes were used for questions that were presented and should have been answered but were not. The length of the Omitted response code given to a variable in the R and SPSS data files depends on the number of characters needed to represent the variable. In all cases, the space necessary to represent the variable is filled with 9s. No distinction is made between items left blank and items with invalid.

Not Applicable Response Codes (R: 6, 96, 996, ...; SPSS: 6, 96, 996, ...; SAS: .B)
Not Applicable response codes were used for the context questionnaire items for which responses were dependent upon a filter question. For example, in the eighth-grade Student Questionnaire—Separate Science Version, if a student answers "No" to being

asked if they studied biology in school this year, all questions corresponding to biology were coded as "Not Applicable."

Context Questionnaire Scales and Derived Variables

In the TIMSS 2023 Context Questionnaires, there are instances where several questions are asked about various aspects of a single construct. In these cases, responses to the individual items were combined to create a score which provided a more comprehensive interpretation of the construct of interest than the individual variables could on their own. These context questionnaire scales also were categorized, usually into three groups, to create an index. The context questionnaire scales and their indices are included in the TIMSS 2023 International Database context data files and described in Chapter 15 of *TIMSS 2023 Technical Report*.

Additional variables were derived from responses to multiple questions to provide more pertinent information for analysis and reporting. Parents' education is an example where responses from both parents were combined into a single variable to report a single educational level. A description of the derived variables included in the TIMSS 2023 International Database are provided in the Derived Context Variables download on the TIMSS 2023 International Database website.

Sampling and Weighting Variables

Several sampling and weighting variables are included in the TIMSS 2023 data files; they are listed and described in Exhibit 2.10. Exhibit 2.11 indicates the location of the various sampling and weighting variables among the different types of data files in the TIMSS 2023 International Database. It is important to note that the teacher context data files, home context data files, and scoring reliability data files do not contain any sampling and weighting variables.

The variable TOTWGT is the overall student sampling weight that should be used for most student-level analyses (except when teacher data are involved—see below). The variable TOTWGT, within each country, sums up to the estimated population of students in the country.

Exhibit 2.10: TIMSS 2023 Sampling and Weighting Variables

Variable	Description
TOTWGT	Total student weight—sums to the national student population
SENWGT	Student senate weight—sums to 500 in each country
HOUWGT	Student house weight—sums to the national student sample size
SCHWGT	Total school weight—the product of WGTFAC1 and WGTADJ1
STOTWGTU	Sum of TOTWGT at the school level
TCHWGT	Overall teacher weight
MATWGT	Mathematics teacher weight
SCIWGT	Science teacher weight
JKZONE	The sampling zone, or stratum, to which the student's school is assigned
JKREP	The sampling replicate, or primary sampling unit, to which the student's school is assigned
JKCZONE	The sampling zone, or stratum, to which the school is assigned
JKCREP	The sampling replicate, or primary sampling unit, to which the school is assigned
WGTFAC1	School weighting factor
WGTADJ1	School weighting adjustment
WGTFAC2	Class weighting factor
WGTADJ2	Class weighting adjustment
WGTFAC3	Student weighting factor
WGTADJ3	Student weighting adjustment

The variable SENWGT is a transformed version of the student TOTWGT that adds to 500 within each participating country. This weight is used during TIMSS operational calibration procedures to control the proportional contribution of each country to the estimates. By using the SENWGT, each country contributes the same to the analysis, regardless of sample or population size, assuming no missing data on the analysis variables.

HOUWGT is another transformed version of TOTWGT that adds to each country's actual sample size. This weight should be used with some legacy analysis programs that use actual sample size, instead of the sum of the weights, for variance estimation. While such programs are rare, users can check for this by running the same analysis using TOTWGT and using HOUWGT and comparing the results. Except for estimates of population size (e.g., number of students), all other estimates should not vary.

Exhibit 2.11: Locations of Sampling and Weighting Variables in TIMSS 2023 Data Files

Variable	Data File				
	ASA/BSA	ASP/BSP	ASG/BSG	AST/BST	ACG/BCG
TOTWGT	●	●	●		
SENWGT	●	●	●		
HOUWGT	●	●	●		
SCHWGT					●
STOTWGTU					●
TCHWGT				●	
JKZONE	●	●	●	●	
JKREP	●	●	●	●	
JKCZONE					●
JKCREP					●
WGTFAC1	●	●	●		●
WGTADJ1	●	●	●		●
WGTFAC2	●	●	●		
WGTADJ2	●	●	●		
WGTFAC3	●	●	●		
WGTADJ3	●	●	●		

Exhibit 2.4 describes the file name conventions.

The weight variables TOTWGT, SENWGT, and HOUWGT are designed for use in student-level analyses from all student-level and school-level files, including the home context data files. The weight variable SCHWGT is designed for use in school-level analyses where the schools are the units of analysis. The weight variable STOTWGTU is a school-level weight which is the sum of TOTWGT for all students within a school.

The weight variables TCHWGT, MATWGT, and SCIWGT are specifically designed for using teacher context data in student-level analyses and are based on TOTWGT. Whereas TCHWGT is used for analyses using all teachers, MATWGT and SCIWGT are used for analyses of mathematics and science teachers, respectively. These teacher weights are located in the student-teacher linkage files (AST and BST), not in the actual teacher context data files (ATG, BTM, and BTS). Analyses with teacher data will be weighted properly by merging the teacher files with the student-teacher linkage files.

All weighting variables beginning with the letters “WGT” provide insight into the multi-stage sampling and weighting methodology applied to the data. These weights are described in detail in Chapter 3 of *TIMSS 2023 Technical Report*.

The sampling variables beginning with the letters "JK" are used to compute standard errors based on the jackknife repeated replication methodology see Chapter 13 of *TIMSS 2023 Technical Report*.

Structure and Design Variables

Besides the variables used to store responses to the context questionnaires and achievement booklets, the data files also contain variables meant to store information that identify and describe the respondents and design information required to properly analyze the data.

Identification Variables

In all data files, several identification variables are included that provide information to identify countries, students, teachers, or schools. These variables also are used to link, or merge, cases between the different data file types. The identification variables have the prefix "ID" and are described below.

IDCNTRY

IDCNTRY is a six-digit country identification code based on the ISO classification as shown in Exhibit 2.2. This variable should always be used as the first linking variable whenever files are linked within and across countries.

IDSCHOOL

IDSCHOOL is a four-digit identification code that uniquely identifies the participating schools within each country. The school codes are generated and assigned specifically for TIMSS 2023 and are not meant to represent actual school identifiers in the participating countries. They are not unique across countries. Schools across countries can be identified uniquely only with the IDCNTRY and IDSCHOOL combination of linking variables.

IDCLASS

IDCLASS is a six-digit identification code that uniquely identifies the sampled classrooms within a country. The variable IDCLASS has a hierarchical structure and is formed by concatenating the IDSCHOOL variable and a two-digit sequential number identifying the sampled classrooms within a school. Classrooms can be identified uniquely in the database by the combination of IDCNTRY and IDCLASS as linking variables.

IDSTUD

IDSTUD is an eight-digit identification code that uniquely identifies each sampled student in a country. The variable IDSTUD also has a hierarchical structure and is formed by concatenating the IDCLASS variable and a two-digit sequential number identifying all students within each classroom. Students can be identified uniquely in the database by the combination of IDCNTRY and IDSTUD as linking variables.

IDTEACH
IDTEACH is a six-digit identification code that uniquely identifies a teacher within a school. It has a hierarchical structure and is formed by the concatenation of IDSCHOOL and a two-digit sequential number within each school.

IDLINK
IDLINK uniquely identifies the class for which a teacher answered a questionnaire. The combination of linking variables IDCNTRY, IDTEACH, and IDLINK uniquely identifies all teacher-class combinations in the database.

IDTEALIN
IDTEALIN is a concatenation of IDTEACH and IDLINK. It can be used with IDCNTRY, instead of IDTEACH and IDLINK, to uniquely identify all teacher-class combinations in the database.

IDGRADE
IDGRADE identifies the target grade of the participating students. In TIMSS 2023, the usual value is “4” at the fourth grade and “8” at the eighth grade. Some countries and benchmarking participants can have the value “5” at the fourth grade and “9” at the eighth grade.

IDBOOK
IDBOOK identifies the specific assessment booklet that was administered to each student. In the TIMSS 2023 data, the digital booklets are given a numerical value from “1” through “14.” For countries that administered a paper option, booklets have a value from “1” through “8.”

Exhibit 2.12 shows in which data files the various identification variables are located. Cells are shaded to indicate the combinations of variables used to identify uniquely the records contained in the different data file types.

Exhibit 2.12: Location of Identification Variables in TIMSS 2023 Data Files

Variable	Data File						
	ASA/BSA	ASP/BSP	ASG/BSG	AST/BST	ATG/BTM/BTS	ACG/BCG	ASH
IDCNTRY	●	●	●	●	●	●	●
IDSCHOOL	●	●	●	●	●	●	●
IDCLASS	●	●	●	●			●
IDSTUD	●	●	●	●			●
IDTEACH				●	●		
IDLINK				●	●		
IDTEALIN				●	●		
IDGRADE	●	●	●	●	●	●	●
IDBOOK	●	●	●	●			

Shading indicates combinations of variables that allow for uniquely identifying records across data files.
Exhibit 2.4 describes the file name conventions.

In the student context, home context, student process, and achievement data files, the variables IDCNTRY and IDSTUD provide a unique identification number to identify all students in the database. Since teachers may teach more than one class, the combination of the IDCNTRY, IDTEACH, and IDLINK (or IDTEALIN) variables in the teacher context data files is needed to identify uniquely all teachers and the classes they teach. Teacher context variables are linked to the appropriate students using the student-teacher linkage data files. The variable IDSCHOOL, contained in all files, is a unique identification number for each school within a country. Combined with IDCNTRY, it can be used to link school context data to corresponding students or teachers.

Tracking Variables

Tracking variables are used to store information about students, teachers, and schools provided by the survey tracking forms, containing lists of students, teachers, and schools used for sampling and administrative purposes. Tracking variables also include information about the test administration. Many of these variables have the prefix “IT.” All tracking variables are included in the student context data files (ASG/BSG). ITLANG and ITMODE are included in the home, school, and teacher context files.

ITSEX

ITSEX indicates the gender of each student as stated in the Student Tracking Forms.

ITBIRTHM and ITBIRTHY

ITBIRTHM and ITBIRTHY indicate the month and year of birth of each student as stated in the Student Tracking Forms. Both variables are available only in the restricted-use version of the TIMSS 2023 International Database.

ITDATE
ITDATE indicates the testing date for each student. This variable is available only in the restricted-use version of the TIMSS 2023 International Database.

ASDAGE/BSDAGE
ASDAGE/BSDAGE indicates the age of each student at the time of testing. It is derived from ITBIRTHM, ITBIRTHY, and ITDATE. While the source variables are only included in the restricted-use version of the TIMSS 2023 International Database, ASDAGE/BSDAGE is available in the public use version.

ITLANG_x
ITLANG_x indicates the language of the TIMSS assessment and context questionnaires, where "x" can take the values "SA," "HQ," "SQ," "CQ," and "TQ" to denote the various survey instruments. The valid codes for ITLANG are specified in the codebook files.

ITMODE_x
ITMODE_x indicates the mode of administration (paper or digital) for the TIMSS assessments and context questionnaires, where "x" can take the values "SA," "HQ," "SQ," "CQ," and "TQ" to denote the various survey instruments. This variable is available only in the restricted-use version of the TIMSS 2023 International Database.

ITDEV
ITDEV indicates the type of digital device used for the TIMSS assessment. This variable is available only in the restricted-use version of the TIMSS 2023 International Database.

ITADMINI
ITADMINI indicates the position of the test administrator who conducted the students' test session: national center staff, teacher, or other.

ITCOURSE
ITCOURSE indicates the subject taught by the teacher in the teacher and student-teacher linkage files. The valid codes for ITCOURSE are specified in the codebook files.

Test-Curriculum Matching Analysis (TCMA) Data

The Test-Curriculum Matching Analysis (TCMA) was conducted to investigate the degree that the TIMSS 2023 mathematics and science assessments matched mathematics and science curricula of the participating countries. To that end, participating countries were asked to indicate which items in the TIMSS 2023 assessments assessed topics were covered by their national curricula. Based on psychometric models tailored to each country's set of curriculum-relevant items, each country was able to see the performance of all countries only on the items that were part of its curriculum, and also the performance of its students on the items judged appropriate for the curricula in other countries. The

analytical method used and the results of the TCMA are presented in Appendix C of *TIMSS 2023 International Results in Mathematics and Science*.

The International Database contains TCMA data downloads for the fourth-grade and eighth-grade TCMA files in Excel format. Only countries that participated in the digital assessment are included.

Supplemental Material

In addition to the data files, a number of supplemental documents are available for download along with the TIMSS 2023 International Database to aid users in analyzing the data. All materials are available to download from the TIMSS 2023 International Database webpage.

Codebooks

All information related to the structure of the TIMSS 2023 data files, as well as the source, format, descriptive labels, and response option codes for all variables, is contained in codebook files. Codebooks can be downloaded as Excel files from the TIMSS 2023 International Database webpage.

In the codebook Excel files, there is a tab for each appropriate data file type in the TIMSS 2023 International Database (see Exhibit 2.4 for the naming convention). These tabs describe the contents and structure of the individual TIMSS 2023 data files. Important codebook fields include LABEL, which contains extended textual information for all variables, QUESTION LOCATION, which provides the location of questions and achievement items within their respective survey instruments, and VALUE SCHEME DETAILED, which lists the acceptable responses allowed for each variable.

Data Almanacs

Data almanacs provide weighted summary statistics for variables in the TIMSS 2023 data files. There are two basic types of data almanacs: achievement data almanacs for the achievement items and context data almanacs for the context variables. All data almanac files are provided in PDF format and Excel format.

Achievement Data Almanacs

The achievement data almanacs provide weighted summary statistics for each participating country on each individual achievement item included in the TIMSS 2023 assessments. The achievement data almanac files available in the International Database are listed in Exhibit 2.13. For each grade, the International Database contain almanacs separate by subject for items in the primary digital assessment (“T23”), the paper assessment (“T23P”), and the fourth-grade TIMSS less difficult paper assessment (“T23LD”).

Exhibit 2.13: TIMSS 2023 Achievement Data Almanacs

Assessment	Grade 4	Grade 8
TIMSS	T23_G4_MAT_ItemAlmanac T23_G4_MAT_TrendItemAlmanac T23_G4_SCI_ItemAlmanac T23_G4_SCI_TrendItemAlmanac	T23_G8_MAT_ItemAlmanac T23_G8_MAT_TrendItemAlmanac T23_G8_SCI_ItemAlmanac T23_G8_SCI_TrendItemAlmanac
TIMSS Paper	T23P_G4_MAT_ItemAlmanac T23P_G4_MAT_TrendItemAlmanac T23P_G4_SCI_ItemAlmanac T23P_G4_SCI_TrendItemAlmanac	T23P_G8_MAT_ItemAlmanac T23P_G8_MAT_TrendItemAlmanac T23P_G8_SCI_ItemAlmanac T23P_G8_SCI_TrendItemAlmanac
TIMSS Less Difficult Paper	T23LD_G4_MAT_ItemAlmanac T23LD_G4_MAT_TrendItemAlmanac T23LD_G4_SCI_ItemAlmanac T23LD_G4_SCI_TrendItemAlmanac	

The achievement data **Item Almanacs** display for each item its classification in the content and cognitive domains, the item block to which it belongs, a brief description of the item, its variable name, whether it is a multiple-choice or constructed-response item, its point value, and the correct response key if it is a multiple-choice item.

The **Trend Item Almanacs** provide summary statistics for achievement items used in both the 2019 and 2023 assessments.

The achievement data almanacs also display the international averages for each item, with each country weighted equally. The benchmark participants, listed below the international averages, are not included in the calculation of international averages.

There are two types of displays in the achievement data almanacs, depending on whether an item is a multiple-choice item or a constructed-response item. The statistics in these almanacs include the following:

- N: The number of students to whom the item was administered
- DIFF: The percentage of students that responded correctly to the item. In the case of constructed-response items worth 0,1, or 2 points, the percentages are weighted across each of the possible score point categories. Omitted and Not Reached responses are excluded from this calculation.
- A, B, C, D, etc.: The percentage of students choosing each one of the response options for a multiple-choice item
- Scoring Guide Codes (e.g., 0, 1, 2, 20, 21, 10, 11, 70, 71): The percentage of student responses assigned each of the codes in the scoring guide for a constructed-response item
- OMITTED: The percentage of students that omitted a response to the item
- NOT REACHED: The percentage of students that did not reach the item

- 2pt, 1pt, 0pt: The percentage of students that scored 2 points,1 points, or 0 points on the item
- GIRL PCT RIGHT and BOY PCT RIGHT: The percentage of girls and boys that either got a multiple-choice item right or obtained the maximum score on a constructed-response item. Omitted and Not Reached responses are excluded from this calculation.

Context Data Almanacs

Context data almanac files contain weighted summary statistics for each participating country on each variable in the student, home, school, and teacher context questionnaires, including the context questionnaire scales and their indices and the derived variables based on context variables. Among the statistics reported is mean achievement by response category. The context data almanacs also display for each variable the question as it was asked, its location in the corresponding questionnaire, and its variable name in the data files. The context data almanac files available in the International Database are listed in Exhibit 2.14. There are separate sets of context almanacs to report mean mathematics and mean science achievement by response category.

Exhibit 2.14: TIMSS 2023 Context Data Almanacs

Assessment	Mathematics	Science
Grade 4	T23_G4_MAT_StudentAlmanac T23_G4_MAT_HomeAlmanac T23_G4_MAT_TeacherAlmanac T23_G4_MAT_SchoolAlmanac	T23_G4_SCI_StudentAlmanac T23_G4_SCI_HomeAlmanac T23_G4_SCI_TeacherAlmanac T23_G4_SCI_SchoolAlmanac
Grade 8	T23_G4_MAT_StudentAlmanac T23_G8_MAT_TeacherAlmanac T23_G4_MAT_SchoolAlmanac	T23_G4_SCI_StudentAlmanac T23_G4_SCI_TeacherAlmanac T23_G4_SCI_SchoolAlmanac

The context data almanacs also display the international averages for each variable, with each country weighted equally. The benchmark participants, listed below the international averages, are not included in the calculation of international averages.

The context almanacs include all countries together at each grade since the same context questionnaires were administered regardless of mode of administration.

There are two types of displays in the context data almanacs, depending on whether the data are categorical (i.e., have a small number of discrete values) or continuous. The almanac display for categorical variables includes the following:

- The sample size (number of students, parents, schools, or teachers included in the sample)
- The number of valid cases (number of students, parents, schools, or teachers for whom valid data were obtained)

- The weighted percentages of students corresponding to each valid response option (percentages based only on the students with valid data, as well as Not Applicable codes when used)
- The weighted percentages of students for whom none of the valid response options were selected, coded as Not Administered or Omitted (percentages based on all students)
- The weighted mean achievement values of students corresponding to each valid response option, as well as the Not Administered and Omitted codes
- In cases where a variable can be coded as Not Applicable because of an earlier filter question, the weighted percentage of students for whom the variable is coded as Not Applicable also is displayed, along with the corresponding weighted mean achievement

The almanac display for continuous variables includes the following:

- The sample size (number of students, parents, schools, or teachers included in the sample)
- The number of valid cases (number of students, parents, schools, or teachers for whom valid data were obtained)
- The weighted percentages of students for whom the variable is coded as Not Administered or Omitted (percentages based on all students)
- The weighted mean, minimum, maximum, and the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles across students (based on the students with valid data)
- In cases where a variable can be coded as Not Applicable because of an earlier filter question, the weighted percentage of students for whom the variable is coded as Not Applicable also is displayed

It is important to note that all statistics reported in the context data almanacs, with the exception of the sample sizes and the number of valid cases, always are based on student-level calculations—for example, the percentage of students whose teachers or schools gave a particular response to a question, because teacher data and school data usually are analyzed as student attributes.

Achievement Item Information

Achievement Item information files include characteristics of each achievement item in the TIMSS 2023 assessment. For each grade, item information in a single Excel file, with separate tabs for the main digital assessment (“T23 G4” or “T23 G8”), the paper assessment, (“T23 G4 Paper” or “T23 G8 Paper”), and the fourth-grade less difficult paper assessment (“T23 G4 LD Paper”). The file includes the following information for all items:

- **Item ID**, the item's unique identifier corresponding to the achievement variable name
- For the primary digital assessment, the **Screen ID** used as the basis for storing derived process variables in ASP/BSP files
- A **Subject** column showing if the item was a mathematics (M) item or science (S) item
- An indicator for the **Grade** of the assessment (4 or 8)
- An indicator for the assessment **Cycle** when the item and its text was first presented (see earlier section on Achievement Item Variable Naming Convention)
- The item's **Block** and **Block Seq**, its sequential location within the block
- The item's **Secure Status**, indicating whether the item is available for "restricted use" after the 2023 assessment, or "secured" for use in future assessments
- The **Content Domain**, **Topic Area**, **Topic**, **Cognitive Domain**, and **Cognitive Area** assessed by the item, according to the *TIMSS 2023 Assessment Frameworks*
- For science items, an indicator of whether the item was classified as **Environmental Knowledge**
- The item's **Maximum Points** value
- The **Item Type**, either multiple-choice (MC) or constructed-response (CR)
- The number of **Response Options** for multiple-choice items
- The correct response **Key** for multiple-choice items
- **Scaling Status**, indicating whether the item was included in the IRT scaling
- A **Label** for the item

Achievement IRT Item Parameters

The International Database includes Excel files with the IRT item parameters estimated for all TIMSS 2023 digital items from the TIMSS 2023 psychometric analysis. The Achievement IRT Parameters downloads on the TIMSS 2023 International Database webpage provide one Excel file for each grade, with separate tabs for mathematics items, science items, and nonresponse indicators. These item parameters are described and presented in Chapter 12 of *TIMSS 2023 Technical Report*, which also reports the linear scale transformation constants that were used to set the TIMSS 2023 achievement scores on the TIMSS trend scale. Item parameters and scale transformations for the paper

assessment options are the same as those used in TIMSS 2019 and can be downloaded from the TIMSS 2019 International Database.

Context Questionnaire Variables

The International Database provides information about the items in the context questionnaires administered in TIMSS 2023, including the curriculum questionnaire. For each grade, ZIP files can be downloaded containing international versions of each questionnaire, annotated with the variable names under which responses are recorded in the data files. In addition, an Excel file provided for each questionnaire contains a list with detailed information for each question, including the question location, the corresponding variable name, and the question text, as well as whether the question is considered to be "trend"—whether a comparable question was asked in TIMSS 2019.

More information about the TIMSS 2023 Context Questionnaires is provided in Chapter 2 of *TIMSS 2023 Technical Report*. The TIMSS questionnaires were designed to provide an opportunity for individual countries to make modifications to some questions or response options. This allowed countries to include the appropriate wording or options most consistent with their own national systems. In the international version of the questionnaires, such questions contain instructions to the National Research Coordinators (NRCs) to substitute the appropriate wording for their country and/or modify or delete any inappropriate questions or options. These instructions were indicated in the questionnaires by text inserted within carets (e.g., <country-specific>). The NRCs were to substitute, if necessary, an appropriate national adaptation that would retain the same basic interpretation as the text within carets. These adaptations are documented in a National Adaptations Database available from the TIMSS 2023 International Database webpage, described below.

National Adaptations to the Context Questionnaires

The National Adaptations Database available through the International Database describes the adaptations countries made to the international version of the TIMSS 2023 Context Questionnaires. This information provides users with a guide to evaluate the availability of internationally comparable data for use in secondary analyses involving the context variables. National adaptations to context questionnaires include questions that countries were required to adapt, questions that were not administered, and questions that countries modified to suit their national context.

For each grade, ZIP files are available containing Excel files for each questionnaire with a list of adaptations. For each question that was adapted, a national entry is included if the version of the question administered was different from the international version in meaning or in context. The following information is provided:

- Question number

- Question stem and response options
- Variable name(s)
- National adaptation, listed by country.

Each national adaptation entry is assigned either code "D" or code "X," representing the following:

D The question is still considered comparable to the international version, and the corresponding data are included in the international database.

X The corresponding data are not included in the international database. The question was not administered, not applicable, or deleted for any of several reasons (e.g., deemed not internationally comparable, removed per country request, or a data problem).

Derived Context Variables

The International Database includes data for variables derived from other source variables in the database, including individual questionnaire items. Derived context variables are used for reporting or conducting other secondary analysis. Descriptions of how these variables were derived can be downloaded from the TIMSS 2023 International Database webpage.

The student context data files (ASG/BSG) contain additional variables derived from achievement item responses used to evaluate the reliability of each country's achievement data as reported in Appendix E of *TIMSS 2023 International Results in Mathematics and Science*. Exhibit 2.15 lists these additional variables. Information about their value scheme is provided in the codebooks.

Exhibit 2.15: TIMSS 2023 Achievement Reliability Derived Variables

Grade 4	Grade 8	Description
ASDMLOWP	BSDMLOWP	Mathematics Achievement Too Low for Estimation
ASDSLOWP	BSDSLOWP	Science Achievement Too Low for Estimation
ASDMZERP	BSDMZERP	Mathematics Zero Points Scored
ASDSZERP	BSDSZERP	Science Zero Points Scored
ASDMCORP	BSDMCORP	Mathematics Percent Correct Points Scored
ASDSCORP	BSDSCORP	Science Percent Correct Points Scored
ASDMMISP	BSDMMISP	Mathematics Percent Missing Item Responses
ASDSMISP	BSDSMISP	Science Percent Missing Item Responses

Special Programs

The International Database includes R, SPSS, and SAS scripts or syntax files which will recode the responses to individual items from the achievement data files to their appropriate score levels. These programs are described in Chapter 3 of this User Guide.

CHAPTER 3

Special Programs

Overview

This chapter describes special programs available in the TIMSS 2023 International Database. The programs recode the responses to the individual achievement item variables in the student achievement data files to their appropriate score levels and can apply user-customized values for item nonresponse. Programs are provided by grade to recode all assessment item variables. ZIP files are provided by type of software used containing one program for fourth-grade data (ASASCRM8) and one for eighth-grade data (BSASCRM8). The programs are available in R, SPSS, and SAS.

Scoring the TIMSS 2023 Items

The ASA/BSA files contain variables for the coded responses to the test items, but analyzing performance on individual scored items may be of interest to some users. Carrying out such analyses require that the individual items in the data be assigned their correctness score levels, rather than the actual response options selected by students for multiple-choice items or the diagnostic codes assigned to students' responses on constructed-response items. Users may also wish to apply certain values to item-level nonresponse codes for Omitted and Not Reached item responses (see section on Codes for Missing Values for achievement data files in Chapter 2). The International Database provides R, SPSS, and SAS programs to perform this task.

For multiple-choice items and some item parts of digital item types, the numbers 1 through 6 are used to represent response options A through F in the achievement data files. These responses must be converted to their appropriate score level ("1" for correct and "0" for incorrect) based on each multiple-choice item's correct response key. For constructed-response items, worth a total of 1 or 2 score points, one- or two-digit diagnostic codes are used in the data files for some items to represent the students' response. These codes must be recoded to represent the correct score-point value of the responses.

For all items, special codes are used to represent *Not Administered*, *Omitted*, or *Not Reached* responses. These special missing codes also may be recoded in order to carry out specific item-level analyses. By default, the Not Administered, Omitted, and Not

Reached response codes are coded by the programs to be missing. These default settings can be modified within the score programs, depending on the requirements of the item-level analyses. For example, users may wish to treat Omitted responses as incorrect, instead of missing.

The International Database includes two R scripts, two SPSS programs, and two SAS programs to recode the responses to individual items from the achievement data files (ASA/BSA) to their appropriate score levels. The ASASCRM8 programs score the digital and paper items for the fourth-grade data files. The BSASCRM8 programs score the digital and paper items for the eighth-grade data files.

If the analysis includes multiple countries, the score programs should use merged data files as input. Consequently, users first must create a merged data file of all required student achievement data files (ASA/BSA) using the Merge Module of the IEA IDB Analyzer, as described in Chapter 1 of this User Guide. The score programs will then create a merged data file with scored achievement items that can be used by the Analysis Module of the IEA IDB Analyzer.

Exhibit 3.1 shows a condensed version of the ASASCRM8.R program to score the individual fourth-grade items. All R, SPSS, and SAS programs have a similar structure (see Exhibits 3.2 and 3.3). To score each individual item, the program code in the R score program must be adapted by completing the following steps:

1. Open ASASCRM8.R or BSASCRM8.R with R or RStudio
2. At the end of the program, specify the folder where the merged R data file of student achievement data files is located in the argument **indir**
3. Specify the folder where the merged R data file of scored achievement items will be located in the argument **outdir**
4. Specify the name of the merged student achievement R data file in the argument **infile**
5. Submit the edited code for processing by R

In this example, the merged R data file of student achievement data files is called ASAALLM8, located in the folder C:\TIMSS2023\Data, and contains the achievement item variables. The resulting data file of scored achievement items will be called ASAALLM8_SCR and saved in the location specified in the **outdir** argument. In this case, the same folder was used for both **indir** and **outdir**. Note that in R, forward slashes (/) should be used instead of backslashes in the file path, so the folder path in Exhibit 3.1 is given as: C:/TIMSS2023/Data/. If using SPSS or SAS, backslashes should be used.

Exhibit 3.1: The ASASCRM8.R Program (Condensed)

```
scoreit <- function(data, item, type, right, NR, OM, other){

. . .

}

Doit <- function(indir=getwd(), outdir=getwd(), infile=""){

 . . .

    #  Score multiple-choice items with A key
    Aright <- c(<List of multiple-choice items where A is correct>)
    Data <- scoreit(data, item=Aright, type="MC", right=1, NR=6, OM=9, other=7)

    #  Score multiple-choice items with B key
    Bright <- c(<List of multiple-choice items where B is correct>)
    Data <- scoreit(data, item=Bright, type="MC", right=2, NR=6, OM=9, other=7)

    #  Score multiple-choice items with C key
    Cright <- c(<List of multiple-choice items where C is correct>)
    Data <- scoreit(data, item=Cright, type="MC", right=3, NR=6, OM=9, other=7)

    #  Score multiple-choice items with D key
    Dright <- c(<List of multiple-choice items where D is correct>)
    Data <- scoreit(data, item=Dright, type="MC", right=4, NR=6, OM=9, other=7)

    #  Score multiple-choice items with E key
    Eright <- c(<List of multiple-choice items where E is correct>)
    Data <- scoreit(data, item=Eright, type="MC", right=5, NR=6, OM=9, other=7)

    #  Score multiple-choice items with F key
    Fright <- c(<List of multiple-choice items where F is correct>)
    Data <- scoreit(data, item=Fright, type="MC", right=6, NR=69, OM=99, other=90)

    #  Score constructed-response items
    Constr <- c(<List of constructed-response items>)
    Data <- scoreit(data, item=constr, type="CR", right="", NR=c(6,96), OM=c(9,99), other=c(7,90))

 . . .

}

doit(indir =  "C:/TIMSS2023/Data/",
     outdir = "C:/TIMSS2023/Data/",
     infile = "ASAALLM8")
```

The data file of scored achievement items will have the same data structure as the student achievement data files, but with the score levels stored in the item variables instead of the student responses. The analysis module of the IEA IDB Analyzer will be able to conduct analyses with this data file.

By default, the score programs treat Omitted and Not Reached responses as missing. If Omitted and/or Not Reached responses are to be treated as incorrect rather than missing, users should replace the following lines of R code (which appear twice in the programs, once for multiple-choice items and once for constructed-response items):

```
#  Code for multiple-choice items
                                                . == NR     ~ NaN, # Not Reached
                                                . == OM     ~ NaN, # Omitted

#  Code for constructed-response items
                                                . %in% NR    ~ NaN, # Not Reached
                                                . %in% OM    ~ NaN, # Omitted
```

with these statements:

```
#  Code for multiple-choice items
                                                . == NR     ~ 0, # Not Reached
                                                . == OM     ~ 0, # Omitted

#  Code for constructed-response items
                                                . %in% NR    ~ 0, # Not Reached
                                                . %in% OM    ~ 0, # Omitted
```

Executing the equivalent SPSS programs (Exhibit 3.2) requires the same steps as the R programs.

Exhibit 3.2: The ASASCRM8.SPS Program (Condensed)

```
DEFINE SCOREIT <List of macro parameters> .
. . .

!ENDDEFINE .

DEFINE DOIT (INDIR  = !CHAREND('/') /
             OUTDIR = !CHAREND('/') /
              INFILE = !CHAREND('/') ) .
. . .

SCOREIT TYPE = MC / ITEM = <List of multiple-choice items where A is correct> /
      RIGHT = 1 / NR = 6 / NA = SYSMIS / OM =  9 / OTHER =  7 .

SCOREIT TYPE = MC / ITEM = <List of multiple-choice items where B is correct> /
      RIGHT = 2 / NR = 6 / NA = SYSMIS / OM =  9 / OTHER =  7 .

SCOREIT TYPE = MC / ITEM = <List of multiple-choice items where C is correct> /
      RIGHT = 3 / NR = 6 / NA = SYSMIS / OM =  9 / OTHER =  7 .

SCOREIT TYPE = MC / ITEM = <List of multiple-choice items where D is correct> /
      RIGHT = 4 / NR = 6 / NA = SYSMIS / OM =  9 / OTHER =  7 .

SCOREIT TYPE = MC / ITEM = <List of multiple-choice items where E is correct> /
      RIGHT = 5 / NR = 6 / NA = SYSMIS / OM =  9 / OTHER =  7 .

SCOREIT TYPE = MC / ITEM = <List of multiple-choice items where F is correct> /
      RIGHT = 6 / NR = 96 / NA = SYSMIS / OM = 99 / OTHER = 90 .

SCOREIT TYPE = CR / ITEM = <List of constructed-response items> /
      RIGHT = 0 / NR = 6 96 / NA = SYSMIS / OM = 9 99 / OTHER = 7 90 .

 . . .

!ENDDEFINE .

DOIT INDIR  = C:\TIMSS2023\Data /
       OUTDIR = C:\TIMSS2023\Data /
       INFILE = ASAALLM8 .
```

If Omitted and Not Reached responses are to be treated as incorrect rather than missing, users should replace the following SPSS statements:

```
* Code for multiple-choice items .

        (!NR = SYSMIS) /* Not Reached */
        (!OM = SYSMIS) /* Omitted */

* Code for constructed-response items .

      !DO !N !IN (!NR) (!N = SYSMIS) !DOEND  /* Not Reached      */
      !DO !M !IN (!OM) (!M = SYSMIS) !DOEND  /* Omitted      */
```

with these statements:

```
* Code for multiple-choice items .

        (!NR = 0) /* Not Reached */
        (!OM = 0) /* Omitted */

* Code for constructed-response items .

      !DO !N !IN (!NR) (!N = 0) !DOEND  /* Not Reached      */
      !DO !M !IN (!OM) (!M = 0) !DOEND  /* Omitted      */
```

Executing the equivalent SAS programs (Exhibit 3.3) requires the same steps as the R and SPSS programs.

Exhibit 3.3: The ASASCRM8.SAS Program (Condensed)

```
%MACRO SCOREIT (ITEM, TYPE, RIGHT, NR, NA, OM, OTHER) ;

   .  .  .

%MEND SCOREIT ;

%MACRO DOIT (INDIR  = ,
             OUTDIR = ,
             INFILE = ) ;

   .  .  .

  ARRAY ARIGHT <List of multiple-choice items where A is correct> ;
  DO OVER ARIGHT ; %SCOREIT (ARIGHT, "MC", 1, .R, .A, ., .I) ; END ;

  ARRAY BRIGHT <List of multiple-choice items where B is correct> ;
  DO OVER BRIGHT ; %SCOREIT (BRIGHT, "MC", 2, .R, .A, ., .I) ; END ;

  ARRAY CRIGHT <List of multiple-choice items where C is correct> ;
  DO OVER CRIGHT ; %SCOREIT (CRIGHT, "MC", 3, .R, .A, ., .I) ; END ;

  ARRAY DRIGHT <List of multiple-choice items where D is correct> ;
  DO OVER DRIGHT ; %SCOREIT (DRIGHT, "MC", 4, .R, .A, ., .I) ; END ;

  ARRAY ERIGHT <List of multiple-choice items where E is correct> ;
  DO OVER ERIGHT ; %SCOREIT (ERIGHT, "MC", 5, .R, .A, ., .I) ; END ;

  ARRAY FRIGHT <List of multiple-choice items where F is correct> ;
  DO OVER FRIGHT ; %SCOREIT (FRIGHT, "MC", 6, .R, .A, ., .I) ; END ;

  ARRAY CONSTR <List of constructed-response items> ;
  DO OVER CONSTR ; %SCOREIT (CONSTR, "CR",  , .R, .A, ., .I) ; END ;

   .  .  .

%MEND DOIT ;

%DOIT (INDIR  = C:\TIMSS2023\Data ,
       OUTDIR = C:\TIMSS2023\Data ,
       INFILE = ASAALLM8 ) ;
```

If Omitted and Not Reached responses are to be treated as incorrect rather than missing, users should replace the following SAS statements (which appear twice in the programs, once for multiple-choice items and once for constructed-response items):

```
    IF &ITEM = &NR THEN SCORE = . ; * Not Reached ;
    IF &ITEM = &OM THEN SCORE = . ; * Omitted ;
```

with these statements:

```
    IF &ITEM = &NR THEN SCORE = 0 ; * Not Reached ;
    IF &ITEM = &OM THEN SCORE = 0 ; * Omitted ;
```

BOSTON
COLLEGE

timssandpirls.bc.edu

© IEA, 2025

IEA

TIMSS & PIRLS
BOSTON COLLEGE