LITERACY AND NUMERACY ASSESSMENT

IEA LaNA

LaNA 2023 LINKING STUDY

USER GUIDE FOR THE INTERNATIONAL DATABASE

Clara Wilsher Beyer
Oliver Neuschmidt
Sebastian Meyer

IEA TIMSS & PIRLS BOSTON COLLEGE

Clara Wilsher Beyer
International Association for the Evaluation of Educational Achievement
Hamburg, Germany

Oliver Neuschmidt
International Association for the Evaluation of Educational Achievement
Hamburg, Germany

Sebastian Meyer
International Association for the Evaluation of Educational Achievement
Hamburg, Germany

With contributions from

Eugenio Gonzalez
Bethany Fishbein
TIMSS & PIRLS International Study Center, Boston College
United States

Duygu Savaşcı-Smith
Renato Coppi
International Association for the Evaluation of Educational Achievement
Hamburg, Germany

IEA Secretariat
Keizersgracht 311
1016 EE Amsterdam
The Netherlands
Tel.: +31 (0)20 625 3625
Fax: +31 20 420 7136
secretariat@iea.nl

The International Association for the Evaluation of Educational Achievement (IEA), with headquarters in Amsterdam, is an independent, international cooperative of national research institutions and governmental research agencies. It conducts large-scale comparative studies of educational achievement and other aspects of education, with the aim of gaining in-depth understanding of the effects of policies and practices within and across systems of education.

Clara Wilsher Beyer
Oliver Neuschmidt
Sebastian Meyer

LaNA User Guide for the International Database

TIMSS & PIRLS
BOSTON COLLEGE

Contents

List of tables and figures

Tables

Figures

1. Introduction

Overview

The IEA Literacy and Numeracy Assessment (LaNA) is a new assessment designed to assess reading and mathematics skills in primary education. Results from LaNA help participating countries gauge the percentage of students achieving minimum proficiency in reading and mathematics, aiding in monitoring progress towards Sustainable Development Goal (SDG) 4.1.1b. In addition to assessing mathematics and reading outcomes, LaNA collects information about the home and school learning contexts, providing international comparative perspectives on educational factors related to reading and mathematics achievement.

At the end of 2023, a special administration of LaNA was conducted to link the results with IEA's TIMSS mathematics and the PIRLS reading assessments. This linkage allows for the results from LaNA to be reported on a common metric with TIMSS (Trends in International Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study). The special administration is called the *LaNA 2023 Linking Study*. In the linking study, four additional booklets sourced from TIMSS numeracy and PIRLS literacy materials were included and were rotated within the selected LaNA classes. This design allowed for the establishment of a link between LaNA results and the TIMSS and PIRLS main scales, providing an indication of the percentages of students reaching international benchmarks and minimum proficiency levels in mathematics and reading achievement.

The LaNA 2023 Linking Study international database (IDB) makes the data available to researchers, analysts, and other users to support and promote secondary analysis aimed at improving reading literacy education at the primary level. The database includes student achievement data as well as student and school context data for the six participating countries. This user guide describes the contents of the database and presents example analyses. The results of the LaNA 2023 Linking Study are presented in the LaNA 2023 Linking Study results report (von Davier, 2025), with the rationale for the techniques used and the measures created during data collection and compilation included in the appendices.

The LaNA 2023 Linking Study user guide

The primary purpose of this user guide is to introduce users to the LaNA 2023 Linking Study international database (IDB) and demonstrate the basic functionality of the IEA IDB Analyzer through analysis examples. This user guide includes the following chapters:

- **Chapter 2** provides details about the structure and contents of the LaNA 2023 Linking Study 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 the IEA IDB Analyzer software (IEA, 2024), used for analyzing the LaNA 2023 Linking Study data. A variety of statistical functions are detailed and example analyses are presented.
- **Appendix A** includes the international versions of the LaNA 2023 Linking Study student and school context questionnaires.

2. Contents and Structure of the LaNA 2023 Linking Study Data Files

Overview

This chapter describes the contents of the LaNA 2023 Linking Study international database with a special emphasis on the actual data files used in Chapter 3 to analyze the LaNA 2023 Linking Study results with the IEA IDB Analyzer (International Association for the Evaluation of Educational Achievement [IEA], 2024). The various data files and related materials included in the database are described in accordance with the structure summarized in Table 1. Descriptions of data files include the conventions for naming the various file types and variables as well as codes for missing values.

The LaNA 2023 Linking Study international database and this user guide are available on both the LaNA 2023 Linking Study webpage (https://timssandpirls.bc.edu/LaNA/index.html) and the IEA data repository (https://www.iea.nl/data-tools/repository). Zip files containing all data and documents can be downloaded from either of these locations.

Contents of the LaNA 2023 Linking Study international database

The LaNA 2023 Linking Study data files contain each country's student achievement data files, student and school context data files, and support materials. Table 1 describes the components of the LaNA 2023 Linking Study international database (IDB), with a brief description of the support materials available for download.

Table 1: Contents of the LANA 2023 Linking Study international database

Contents	Description
User guide	User Guide and appendices
Data files	LaNA 2023 linking study student and school data files (available in R, SPSS, and R)
Codebooks	Files describing all variables in the LaNA 2023 Linking Study IDB
Data almanacs	Summary data tables for all LaNA 2023 Linking Study achievement items and context variables
Item information	Files including characteristics of each achievement item in the LaNA 2023 Linking Study IDB

Table 2 lists the countries that participated in the LaNA 2023 Linking Study, the identification codes (ISO codes) used in the data files, and the grade tested. These codes are used to identify the individual country files.

Table 2: Countries in the LaNA 2023 Linking Study

	Identification Codes			
Country	**Alpha**	**Numeric**	**Grade**	**Comment**
Burkina Faso	BFA	854	6	-
Egypt	EGY	818	5	-
Nigeria	NGA	566	6	-
Pakistan	PAK	586	5	-
Palestinian National Authority	PSE	275	5	Westbank only
Senegal	SEN	686	6	-

Restricted use version of the LaNA 2023 Linking Study international database

There are two versions of the LaNA 2023 Linking Study IDB. The public use version is available for immediate access from the LaNA 2023 Linking Study IDB webpage and the IEA data repository.

To minimize the risk of disclosing confidential information, certain variables are not available in the public use version. Table 3 lists the variables removed from the public use version of the LaNA 2023 Linking Study IDB. These variables are available in the restricted use version.

Table 3: Variables in the restricted use version of the LaNA 2023 Linking Study international database

Variable	Description
ITBIRTHY/ITBIRTHM	Students' year and month of birth from the tracking forms
ITDATED /ITDATEM	LaNA 2023 Linking Study testing day and month from the tracking forms
ASBG02A/ASBG02B	Students' year and month of birth from the student questionnaire
ACBG01	Total school enrollment from the school questionnaire
ACBG02A	Enrollment of full-time teachers
ACBG02C	Enrollment of part-time teachers

More details about the restricted use variables are available in the codebook files, described later in this chapter. Users who require any of the removed variables to conduct their analyses should contact IEA through its data repository website (listed above) to obtain permission and access to the restricted use version of the LaNA 2023 Linking Study IDB. Note that the variable reporting students' age remains available in the public use version of the IDB.

LaNA 2023 Linking Study international database files

The LaNA 2023 Linking Study international database files contain data from various instruments administered to students and school principals. This includes student responses to achievement items and context questionnaires, as well as estimated achievement scores and derived context variables. The data files are available in R, SPSS, and SAS formats. File names follow a specific convention, indicating the type of data, target population, and country (see Table 4). For example, "ASGBFANL1.Rdata" is an R data file containing Burkina Faso's sixth-grade student context questionnaire data.

Table 4: LaNA 2023 Linking Study data file names

File name	Description
ASA●●●LN1	Student achievement data files
ASS●●●LN1	Scored student achievement files
ASR●●●LN1	Within-country scoring reliability data files
ASG●●●LN1	Student context data files
ACG●●●LN1	School context data files

Note:
●●● = Three-character country abbreviation based on the ISO alpha coding scheme (see Table 3).

Achievement data files

This section covers the LaNA 2023 Linking Study student achievement data files, scored achievement files, and within-country scoring reliability files, including response codes, missing codes, and derived items.

Student achievement data files

The LaNA 2023 Linking Study student achievement data files (ASA) include student responses to achievement items. The ASA files also contain achievement variables (plausible values) for the two LaNA 2023 Linking Study achievement scales, as well as sampling and weighting variables, which are also available in the student context data files (ACG).

Students were given one of eight assessment booklets. LaNA booklets (1-4) contain two mathematics blocks and two reading passages (one literary and one informational), with only multiple-choice items. LaNA linking booklets (5-8) include easier material from previous TIMSS and PIRLS assessments, featuring two mathematics blocks and one longer text, with both multiple-choice and constructed-response items. The ASA files contain the actual responses to multiple-choice questions and score codes for constructed-response items, based on the LaNA scoring guides.

Item variable names

The achievement item variable names are based on an alphanumeric code, which depends on the subject and instrument (LaNA or linking booklets).

Please refer to the codebook for a comprehensive list of item variable names.

Item response codes

A series of conventions were adopted to code the item responses included in the LaNA 2023 Linking Study ASA 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 codebook file and as part of the variable label in the achievement codebook files (described in a later section).

For the constructed-response items included in the linking items, numerical values ranging from 0 to 3 are used, which correspond to the score points awarded to the students' responses.

Derived items

For item RP51R17, students were asked to provide more than one answer or a multiple-part answer, each one being scored appropriately. The pattern of responses across these item parts determined the score on the item as a whole. The total score for the item is contained in the derived variable RP51R17, which can be identified by the word "DERIVED" in the item label.

Codes for missing values

A subset of values is reserved for specific item response codes related to different categories of missing data. Users should read this section carefully, as the use of these codes may affect subsequent analyses.

Not administered response codes

Special codes were given to items that were not administered (R: NA; SPSS: Sysmis; SAS: .A) to distinguish these cases from data that were missing due to student non-response. The not administered code was used when an item was not administered, either by design—from 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:

- Item not assigned to the student: All students participating in the LaNA 2023 Linking Study received one out of eight test booklets. All variables corresponding to items that were not administered to a student were coded as not administered.
- Student absent from session: When a student was not present for a particular testing session, either part 1 or part 2 of an assessment booklet, all items belonging to that session were coded as not administered.
- Item left out or misprinted: When a particular item (or a whole page) was misprinted or otherwise not available to the student, the corresponding variable was coded as not administered.
- Item mistranslated or deleted: An item identified during translation verification or item review as having a translation error such that the nature of the question was altered, or as having poor psychometric properties, was coded as not administered.

Omitted response codes

Omitted response codes (R: 9; SPSS: 9; SAS: .) were used for items that a student should have answered but did not. An omitted response code was given when an item was left blank or when two or more response options were checked for a multiple-choice item.

Not reached response codes

An 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 completed in the remainder of that part of the booklet. This response code (R: 6; SPSS: 6; SAS: .R) was used only for the test responses, as applicable.

LaNA 2023 Linking Study plausible values variables

LaNA 2023 Linking Study achievement scales were created for overall mathematics and overall reading. The LaNA 2023 Linking Study IDB includes five imputations of achievement, known as "plausible values," for each student based on their responses and context data. These values reflect data uncertainty and are the best measures of student achievement for group reporting, though not suitable for reporting individual results. Plausible values can be analyzed using analysis techniques for multiply imputed data. The achievement plausible value variable names are as follows:

- ASMMAT01-05 indicate the five plausible values used for the mathematics scale.
- ASRREA01-05 indicate the five plausible values used for the reading scale.

For example, ASRREA01 is the first plausible value for the administered grade's reading achievement.

LaNA international benchmarks of reading achievement

To help users of the LaNA 2023 Linking Study data understand what performance on the overall reading and mathematics achievement scale signifies in terms of the reading and mathematics skills and strategies demonstrated by students at the end of primary school, LaNA uses five points on the overall reading and mathematics scale to serve as international benchmarks.

As shown in Table 5, the LaNA international benchmarks are 625, 550, 475, 400, and 325, which correspond to the advanced international benchmark, the high international benchmark, the intermediate international benchmark, the low international benchmark, and the LaNA basic international benchmark, respectively. The basic international benchmark was developed specifically for the LaNA 2023 Linking Study. All other benchmarks, originally developed as part of the TIMSS mathematics and PIRLS reading scale descriptions, can be referenced to the LaNA 2023 Linking Study via the established link between LaNA and the TIMSS and PIRLS main scales.

Table 5: LANA international benchmarks of reading and mathematics achievement

Scale score	International benchmark
625	Advanced international benchmark
550	High international benchmark
475	Intermediate international benchmark
400	Low international benchmark
325	Basic international benchmark

The LaNA 2023 Linking Study international database files contain a set of variables indicating which international benchmark the students have reached based on each of the imputed scores they have received. There are five benchmark indicator variables for each subject—one for each plausible value of the overall reading and mathematics scales. The international benchmark indicator variables are named **ASRLBM01-05** and **ASMLBM01-05** for reading and mathematics, respectively. The codes used for the benchmark indicator variables are described in Table 6.

Table 6: LaNA 2023 Linking Study international benchmark variable codes

Code	Description
1	Did not reach the basic international benchmark
2	Reached the basic international benchmark, but did not reach the low international benchmark
3	Reached the low international benchmark, but did not reach the intermediate benchmark
4	Performed at or above the intermediate international benchmark, but did not reach the high benchmark
5	Reached the high international benchmark, but did not reach the advanced benchmark
6	Reached the advanced international benchmark

Scored student achievement files

Student achievement in the LaNA 2023 Linking Study is represented by sets of five plausible values for overall mathematics and reading. These plausible values are the preferred metrics for analyzing student achievement. However, users may also be interested in analyzing performance on individual items. Such analyses require assigning correctness scores to individual items in the LaNA 2023 Linking Study database, rather than using the actual response options selected by students for multiple-choice items. To facilitate such analyses, the LaNA 2023 Linking Study provides a scored version of the achievement files (ASS files).

- For multiple-choice items, response options A through D (represented as numbers 1 to 4 in the LaNA 2023 Linking Study ASA files) are converted to their respective score levels: “1” for correct and “0” for incorrect, based on the correct response key.
- Constructed-response reading items, which are worth 1, 2, or 3 points, are coded the same in both file types.

- For constructed-response mathematics items, the two-digit diagnostic codes are converted to one-digit score points ranging from 0 to 2.

The ASS files maintain the same structure as the ASA files, but the item variables contain score levels instead of student responses. The missing code scheme in the scored file follows the same conventions as the ASA files. However, note that scored achievement items use single-digit score points and that the two-digit missing codes in SPSS and R (e.g., "99," "96") are simplified to their one-digit equivalents ("9," "6").

Within-country scoring reliability data files

The LaNA 2023 Linking Study within-country scoring reliability data files (ASR) contain data that can be used to evaluate the reliability of the LaNA Linking Study 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. 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 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 variable indicating agreement between the two scorers.

In the ASA files, 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 the calculation of the proficiency estimates or to modify the score provided by the first scorer.

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 name as follows:

- The **original score** variable follows the item variable naming convention described earlier.
- The **second score** variable adds "_R" to the original score variable (e.g., RP51Z15_R).
- The **score agreement** adds "_I" to the original score variable (e.g., RP51Z15_I).

Scoring reliability codes

The values contained in both the original score and second score variables are the scores based on the LANA 2023 Linking Study scoring guides. The score agreement indicator variable may have one of two values, indicating either agreement or disagreement between the score codes assigned by the two scorers, as described in Table 7.

Table 7: LaNA 2023 Linking Study score agreement variable codes

Code	Description
2	Different score levels (the score digit of both scores is different)
1	Identical score level codes (agreement on score level—certain mathematics items only)
0	Identical codes (agreement)

Context data files

This section covers the LaNA 2023 Linking Study context data files, naming conventions, and data coding. The files include responses to student and school questionnaires, derived variables for reporting, and LaNA 2023 Linking Study context questionnaire scales.

Student context data files

The student context data (ASG) files contain students' responses to questions about their home, school experiences, and attitudes toward reading, along with their reading and mathematics achievement variables (plausible values). They also contain various identification, tracking, weighting, and derived variables, which are described later in this chapter.

School context data files

The school context data (ACG) files contain principals' responses to the LaNA 2023 Linking Study school questionnaire. For student-level analyses, merge these files with the ASG files using country and school identification variables. Details on merging will be provided in Chapter 3.

Context variable names

The context variable naming convention for the variables in the ASG and ACG files uses a 7- or 8-character string. The following rules are applied in naming the context variables:

- The first character indicating the primary student level is always "A" for LaNA 2023 Linking Study data.
- The second character indicates the type of respondent. The letter "C" identifies data from the school principals, and the letter "S" is for student 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 index variables constructed from these context questionnaire scale Rasch scores.
- The fourth character is used to indicate the subject or topic to which a context question refers. For the LaNA 2023 Linking Study, only the letter G—general question—is used.
- The fifth through eighth characters of all context variables represent the sequential numbering of the questions as presented in their respective questionnaires.

Please refer to the codebook for a comprehensive list of context variable names.

Context question location convention

The context variable naming convention shows the order of questions in the questionnaires. Each question has a unique location code, combining its sequence number and a three-character string indicating the source with StQ for the student questionnaire and ScQ for the school questionnaire. For example, StQ-09A refers to part A of question 9 in the student questionnaire, stored as ASBG09A in the data files. This convention is used in codebooks, data almanacs, and supplements to this draft user guide.

The variable names are included in the LaNA 2023 Linking Study international context questionnaires provided in Appendix A.

Context variable response codes

The values for context variables depend on the item format and available options. Categorical questions use sequential numerical values corresponding to the response options (e.g., 1 for the first option, 2 for the second). Open-ended questions are coded with the actual number given as a response.

Codes for missing values

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

Special codes indicate items that were not administered (R: NA; SPSS: Sysmis; SAS: .A), distinguishing them from non-responses. These codes are used when an entire questionnaire was not completed, or when questions were left out, misprinted, mistranslated, or not internationally comparable.

Omitted response codes

Omitted response codes (R: 9, 99, 999, ...; SPSS: 9, 99, 999, ...; SAS: .) were used for unanswered questions. No distinction is made between items left blank and those with invalid answers (e.g., multiple responses in a categorical question or unreadable responses). Invalid or inconsistent data from countries were also coded as omitted.

Please note that the length of the omitted and the not administered response codes given to a variable in the R and SPSS data files depends on the number of characters needed to represent the variable.

Context questionnaire scales and derived variables

This section covers the LaNA 2023 Linking Study context questionnaire scales and derived variables.

Context questionnaire scales

In the LaNA 2023 Linking Study context questionnaires, there are instances in which multiple questions about a single construct were combined using Rasch scaling to create a comprehensive score. These scales were categorized into indices. Scales and derived indices will be described in Appendix D of the LaNA 2023 Linking Study Results report (von Davier, 2025).

Context questionnaire derived variables

In the LaNA 2023 Linking Study IDB, there are two derived context variables. They are based on the responses to the following questions in the school context questionnaire:
ACBG05A, ACBG05B, ACBG05C: For the <grade> students in your school:
“How many days per year is your school scheduled to be open for instruction?”
“__ days” (ACBG05A)
“Within the last 12 months, about how many days was your school closed due to unforeseen circumstances (e.g., strikes, bad weather, etc.)?”
“__ days” (ACBG05B)
“What is the total instructional time, excluding breaks, in a typical day?”
“__ minutes” (ACBG05C)

These are detailed in Table 8 below.

Table 8: Context questionnaire derived variables

Derived variable name	Variable label	Procedure
ACDGIIHLN1	Intended instructional hours per year	ACBG05A * (ACBG05C/60) If any source variable is omitted, the derived variable is coded as omitted. If any source variable is not administered, the derived variable is coded as not administered.
ACDGAIHLN1	Actual instructional hours per year	(ACBG05A - ACBG05B) * (ACBG05C/60) If ACBG05A = ACBG05B, the derived variable is coded as omitted. If any source variable is omitted, the derived variable is coded as omitted. If any source variable is not administered, the derived variable is coded as not administered.

Sampling and weighting variables

Several sampling and weighting variables are included in the LaNA 2023 Linking Study data files. Table 9 includes the various sampling and weighting variable descriptions and indicates their location among the different types of data files in the LaNA 2023 Linking Study international database files.

Table 9: Locations of sampling and weighting-related variables in LaNA 2023 Linking Study data files

		Dataset	
Variable	**Description**	**ASA/ASG**	**ACG**
TOTWGT	Total student weight—sums to the estimated national student population	X	-
SENWGT	Student senate weight—sums to 500 in each country	X	-
HOUWGT	Student house weight—sums to the national student sample size	X	-
SCHWGT	Total school weight—the product of WGTFAC1 and WGTADJ1	-	X
STOTWGTU	Sum of TOTWGT at the school level	-	X
JKZONE	The sampling replicate, or primary sampling unit, to which the student's school is assigned	X	-
JKREP	The sampling zone, or stratum, to which the school is assigned	X	-
JKCZONE	The sampling replicate, or primary sampling unit, to which the school is assigned	-	X
JKCREP	The sampling replicate, or primary sampling unit, to which the school is assigned	-	X
WGTFAC1	School weighting factor	X	X
WGTADJ1	School weighting adjustment	X	X
WGTFAC2	Class weighting factor	X	-
WGTADJ2	Class weighting adjustment	X	-
WGTFAC3	Student weighting factor	X	-
WGTADJ3	Student weighting adjustment	X	-

TOTWGT is the preferred student sampling weight for analyzing student-level data. It sums up to the student population size in each country. For analysis combining all countries, the LaNA 2023 Linking Study provides SENWGT, which standardizes the sum of the weights to 500 per country. HOUWGT adjusts the sum of the weights to match the actual sample size in each country.

The weight variable TOTWGT is designed for use in student-level analyses from all student-level and school-level variables. 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 sampling variables beginning with the letters "JK" are used to compute standard errors based on the jackknife repeated replication methodology.

Structure and design variables

Besides the variables used to store responses to the context questionnaires and achievement booklets, the LaNA 2023 Linking Study 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 LaNA 2023 Linking Study data files, several identification variables are included that provide information to identify countries, students, 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 three-digit country identification code based on the ISO classification as shown in Table 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 code that uniquely identifies participating schools within each country for the LaNA 2023 Linking Study. These codes are not actual school identifiers and are not unique across countries. Schools can be uniquely identified across countries using the combination of IDCNTRY and IDSCHOOL.

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.

IDGRADE

IDGRADE identifies the target grade of the participating students. In the LaNA 2023 Linking Study, the administered grades are "5" or "6" (please refer to Table 3 for more information). Please note that the main data collection for the LaNA 2023 Linking Study has been administered at the beginning of the respective school year.

IDBOOK

IDBOOK identifies the specific assessment booklet that was administered to each student. In the LaNA 2023 Linking Study data, the booklets are assigned numerical values from "1" to "8." Booklet numbers 1 to 4 identify the different LaNA 2023 Linking Study booklet types, while numbers 5 to 8 refer to the LaNA 2023 Linking Study link booklet types.

IDPOP

IDPOP identifies the assessed population. For the LaNA 2023 Linking Study, this is always 1.

Table 10 shows the data files in which the various identification variables are located. Cells are shaded to indicate the combinations of variables used to uniquely identify the records contained in the different data file types.

Table 10: Location of identification variables in LANA 2023 Linking Study data files

	Data file		
Variable	ASA	ASG	ACG
IDCNTRY	X	X	X
IDSCHOOL	X	X	X
IDCLASS	X	X	-
IDSTUD	X	X	X
IDGRADE	X	X	X
IDBOOK	X	X	-
IDPOP	-	-	X

Shading indicates combinations of variables that allow for uniquely identifying records across data files. Table 4 describes the file naming conventions.

In the student context and achievement data files, the variables IDCNTRY and IDSTUD provide a unique identification number to identify all students in the database. 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.

Tracking variables

Tracking variables are used to store information about students and schools provided by the survey tracking forms. Tracking variables also include information about the test administration. Many of these variables have the prefix "IT."

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 LaNA 2023 Linking Study international database.

ITDATED

ITDATE indicates the testing date for each student. This variable is available only in the restricted use version of the LaNA 2023 Linking Study international database.

ITDATEM

ITDATE indicates the testing month for each student. This variable is available only in the restricted use version of the LaNA 2023 Linking Study international database.

ASDAGE

ASDAGE shows each student's age at testing, derived from birth month, year, and test date. While the source variables are only included in the restricted use version of the LaNA 2023 Linking Study international database, ASDAGE is public. Since some students don't know their exact birth date, it's recommended to use the age in years (ASBG02C) for analyses.

Codebook

The LaNA 2023 Linking Study national data file codebook provides item information for every variable included in the data files in a single Excel file. Codebooks can be downloaded as Excel files, which list a tab for each appropriate data file type in the LaNA 2023 Linking Study international database. The tabs include the following key information:

- **Variable name**: the item's unique identifier corresponding to the achievement variable name.
- **Label** of the variable, indicating the content of the item.
- **Width** of the variable.
- The **value scheme**, indicating the possible responses and correct response for multiple-choice items (indicated by an asterisk).
- **Missing scheme (SPSS)** and **missing scheme (SAS)**, indicating the missing value schemes for the different data formats. User-defined R missing values are following the SPSS missing scheme, but they may need special handling. Please refer to the "Important note on handling missing data in R" section at the end of Chapter 3 for more information.
- The **variable class**, either multiple-choice or constructed response.
- **Question location**, indicating the booklet and the sequential location of the item therein.

Users should note that the codebooks include all variables that are available in the public use files and in the restricted use files. The comment field of the codebooks indicates whether variables are available for restricted use only.

Data almanacs

Data almanacs provide weighted summary statistics for all variables in the LaNA 2023 Linking Study. There are two basic types of data almanacs: context data almanacs and achievement data almanacs. 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 LaNA 2023 Linking Study assessment. Separate almanacs are provided for mathematics and reading results.

For each item, the achievement data item almanacs display its classification in the cognitive and content domains, the assessment block information, 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 achievement data almanacs also display the average for the countries participating in the LaNA 2023 Linking Study for each item, with each country weighted equally.

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 statistics:

Multiple-choice items

- N: The number of students to whom the item was administered
- DIFF: The percentage of students that responded correctly to the item.
- A, B, C, D, etc.: The percentage of students choosing each one of the response options for a multiple-choice 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.

- 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.

Constructed-response items

- 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 two or three points, the percentages are weighted across each of the possible score point categories. Omitted and not reached responses are excluded from this calculation.
- Scoring guide codes (e.g., 0, 1, 2, 3): 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.
- 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, and school context questionnaires, including the context questionnaire scales and their indices and the derived variables based on context variables.

For the LaNA 2023 Linking Study, separate almanacs are provided for mathematics and reading results.

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 almanacs display averages across the LaNA 2023 Linking Study 2023 participants for each variable, with equal country weighting. There are two display types, one for categorical variables and one for continuous variables.

Categorical variables

- Sample size (number of students and schools included in the sample);
- Number of valid cases;
- Weighted percentages for each valid response option;
- Weighted percentages for omitted and not administered responses; and
- Weighted mean achievement values for each valid response option, including omitted and not administered codes.

Continuous variables

- Sample size;
- Number of valid cases;
- Weighted percentages for omitted and not administered responses;
- Weighted mean, mode, minimum, maximum, and various percentiles (5th to 95th) for valid data; and
- The weighted mean achievement values of students corresponding to each valid response option, as well as the omitted and not administered codes.

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

Achievement item information

The achievement item information file provides detailed characteristics of each achievement item included in the LaNA 2023 Linking Study, including its associated content area and cognitive domain. The information is organized in a single Excel file with separate tabs for the LaNA items and the LaNA linking items in mathematics and reading.

Achievement IRT item parameters

Please refer to Appendix D of the LaNA 2023 Linking Study Results report (von Davier, 2025) for details on the achievement IRT item parameters.

IEA IDB Analyzer software

The LaNA 2023 Linking Study International Database files are released with the IEA IDB Analyzer, configured for use with the LaNA 2023 Linking Study data. The IEA IDB Analyzer (IEA, 2024) includes a comprehensive manual, accessible via the **View the Help Manual** button on the startup screen. This manual provides detailed information about the full functionality and features of the IEA IDB Analyzer. Chapter 3 details the use of the IEA IDB Analyzer software.

3. Analyzing LaNA 2023 Linking Study Data with the IEA IDB Analyzer

Overview

This chapter describes the general use of the IEA IDB Analyzer software (IEA, 2024) for analyzing the LaNA 2023 Linking Study data. Used in conjunction with R (R Core Team, 2024) and RStudio (RStudio Team, 2024), SPSS (IBM Corporation, 2016), or SAS (SAS Institute, 2016), the IEA IDB Analyzer provides a user-friendly interface to easily merge and analyze the various data file types of the LaNA 2023 Linking Study international database. The software seamlessly accounts for the sampling design information and the multiple imputed achievement scores to produce accurate statistical results.

Five example analyses presented in this chapter illustrate some of the capabilities of the IEA IDB Analyzer (Version 5.0) to compute a variety of statistics, including percentages, means, and confidence intervals of students in specified subgroups, regression coefficients, and percentages of students reaching the LaNA international benchmarks of mathematics achievement. The examples use student and school context data.

Users should be able to perform statistical analyses with the IEA IDB Analyzer with a basic knowledge of the LaNA 2023 Linking Study 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 documentation provided with the international database.

About the IEA IDB Analyzer

Developed by IEA Hamburg, the IEA IDB Analyzer is an application that can be used to combine and analyze data from all IEA's large-scale assessments, as well as analyze data from most major large-scale assessment surveys.

The IEA IDB Analyzer consists of two main modules: the merge module and the analysis module. These two modules are integrated into one single application. A third module introduced with version 5.0 (IEA, 2024) converts SPSS system files to R data files. This module is activated only when the user selects R as the statistical software of choice.

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

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

R and RStudio can be downloaded and installed at no cost from r-project.org and posit.co, respectively (R Core Team, 2020; RStudio Team, 2020). Version 5.0 of the IEA IDB Analyzer requires R version 4.2.0 or later. When running a script produced by the IDB Analyzer, RStudio will print in the console a list of necessary packages.

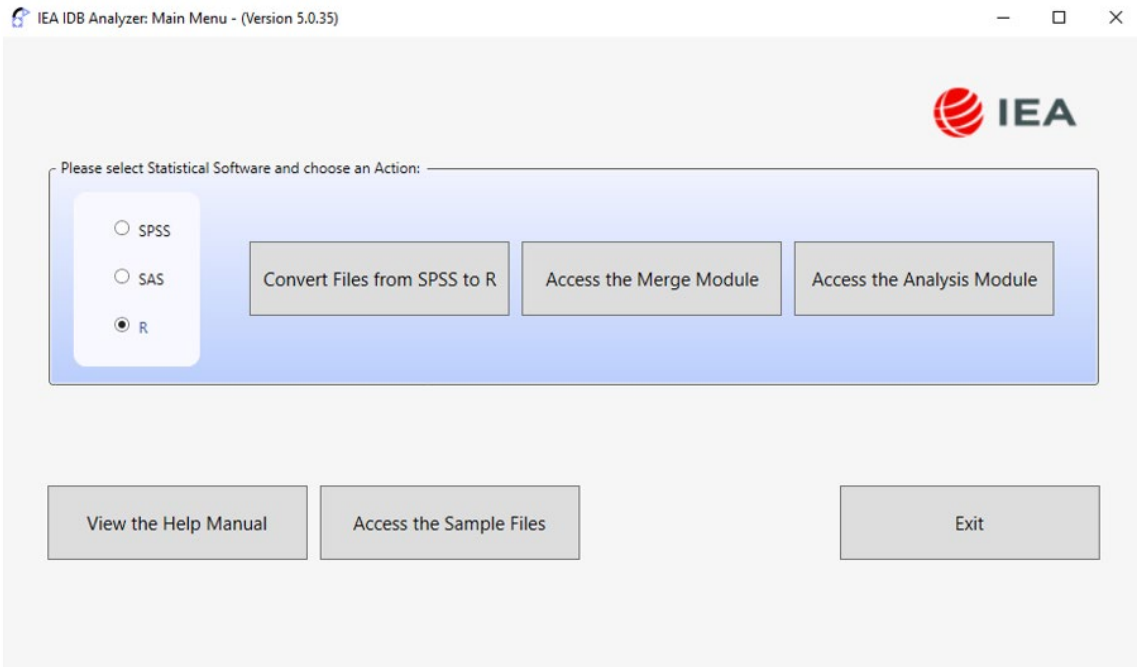
Installing and launching the IEA IDB Analyzer

The latest version of the IEA IDB Analyzer is available for download from the IEA Data and Tools website (https://www.iea.nl/data-tools/tools). When the IEA IDB Analyzer application is launched, the main window will appear, as shown in Figure 1. Users are first directed to choose SPSS, SAS, or R as their statistical software of choice. The examples in this chapter 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 provides 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 **View the Help Manual** button, which users are encouraged to consult for full details on all the functionalities and features of the IEA IDB Analyzer.

Figure 1: IEA IDB Analyzer main window



Merging data files with the IEA IDB analyzer

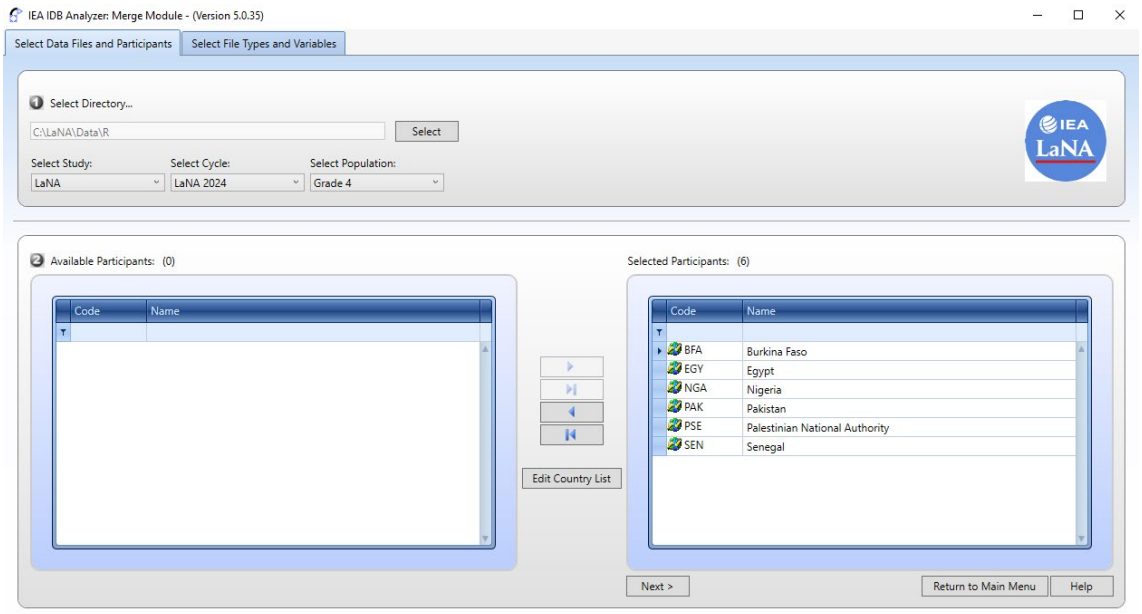
The LaNA 2023 Linking Study data files are disseminated separately by file type or data source and by country. In addition to allowing users to combine datasets from more than one country for cross-country analyses, the merge module allows for the combination of data from different sources (e.g., student and school) into one dataset for subsequent analysis.

Before doing any statistical analysis with the LaNA 2023 Linking Study international database, users should download and copy the contents of the international database 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 to the folder C:\LaNA\Data\R.

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

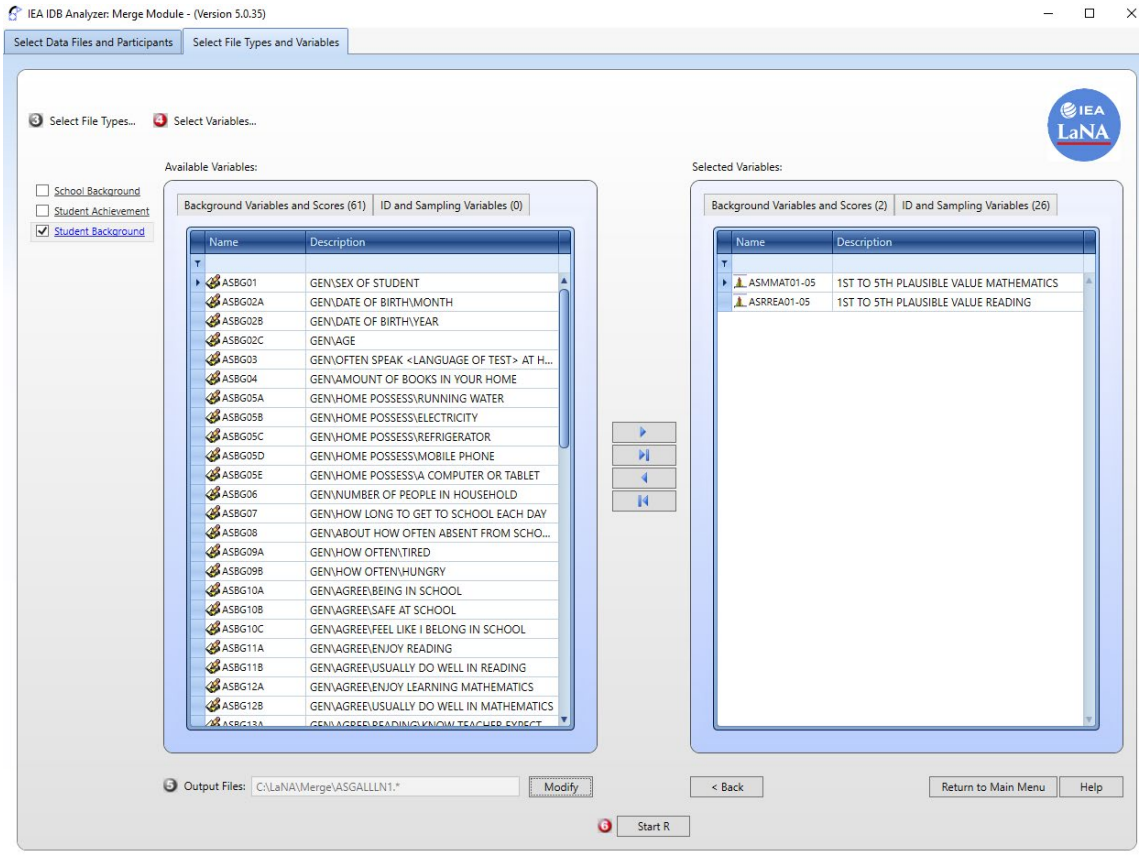
1. Start the IEA IDB Analyzer, select the statistical software you want to use, and then click the **Merge Module** button.
2. Under the **Select Data Files and Participants** tab and in the **Select Directory** field, browse to the folder where all data files are located. All data files must be in the same folder. For example, in Figure 2, all R data files are located in the folder C:\LaNA\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.

Figure 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 Figure 2, all LaNA 2023 Linking Study participants 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 Figure 3, to select the file types and the variables to be included in the merged data file.

Figure 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 Figure 3, the box next to **Student Background** is checked, indicating the LaNA 2023 Linking Study student context data files are selected.[1]
6. Select the variables of interest from the **Available Variables** list in the left panel. The codebook files (described in Chapter 2) as well as Appendix A provide the variable names for storing the data from all questions in the LaNA 2023 Linking Study 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.
8. Click the **Start R** button. The IEA IDB Analyzer will generate and open an R script (*.R) with the necessary code to perform the merge, using the same name and saving it in the same folder with the code necessary to perform the merge. If you follow these steps with an SPSS or SAS data file as the basis, the analyzer will instead create an SPSS syntax file (*.SPS) or a SAS syntax file (*.SAS), respectively.

Once the statistical program of choice 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,

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

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 school context data files

Because the LaNA 2023 Linking Study 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. Therefore, analyzing school context data is better done by linking the students to their schools.

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** 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 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 School 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.
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 will create a script *with the same name and save it in the same folder, containing the code necessary to perform the 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).

Merged data files for the user guide examples

To conduct the analysis examples presented in this chapter, a single merged data file was created following the instructions provided above. The file ASGALLLN1 includes all countries that participated in the LaNA 2023 Linking Study with school and student context data. A list of countries that participated in the LaNA 2023 Linking Study is provided in Table 2.

Conducting analyses with the IEA IDB Analyzer

The IEA IDB Analyzer can perform statistical analyses on any files created using the **Merge Module**. The IEA IDB Analyzer Analysis Module 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 and/or reading achievement, after selecting a **Statistic Type**, users should select the **Use PVs** option from the **Plausible Value Option** drop-down menu.

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

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. When achievement scores are used, the analyses are performed five times (once for each plausible value), and the results are aggregated to produce the correct estimates of achievement and standard errors that incorporate both sampling and imputation errors.

When conducting analyses using contextual variables, users should check whether countries have any missing data on the context variables. High levels of missing data could yield biased results. Similarly, achievement estimates can be unreliable if based on small groups of students.

Statistical procedures in the IEA IDB Analyzer

The following statistical procedures are available in the IEA IDB Analyzer **Analysis Module**.

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, and 5 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 cut point, such as one of the LaNA international benchmarks. This procedure is available only for use with SPSS and SAS in Version 5 of the IEA IDB Analyzer.

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 LaNA 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.

Conducting analyses with LaNA 2023 Linking Study student-level context data

Many types of analyses can be conducted using the student context data files from the LaNA 2023 Linking Study international database. This section guides you through a few analysis examples, including percentages and means, linear regression, and international benchmark analyses.

Example 1 computes average achievement by country, whereas Example 2 computes national average achievement by gender. 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 students reaching each of the LaNA international benchmarks.

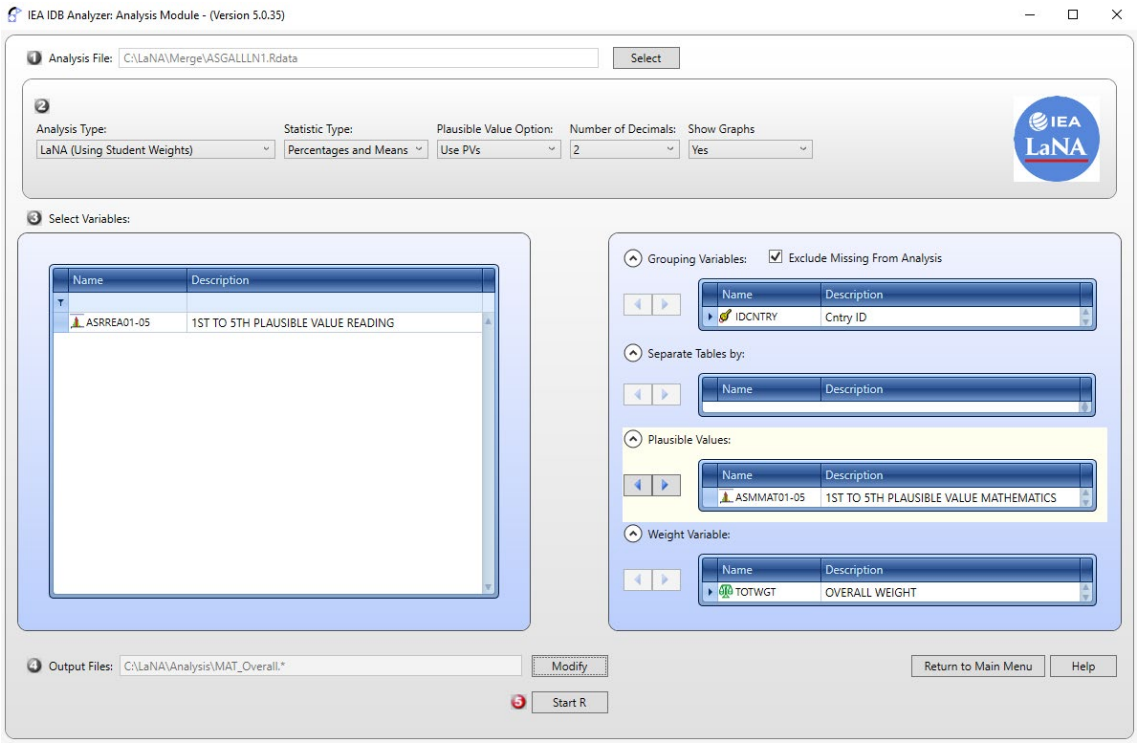
The examples below are based on the R version of the LaNA 2023 Linking Study data files used as input. The process is identical when using SPSS or SAS data files, except that you will see a **Start SPSS** or **Start SAS** button instead of a "Start R" button, and the IEA IDB Analyzer will generate an SPSS syntax file (.SPS) or a SAS syntax file (.SAS) instead of an R script (*.R). After running the corresponding software (R, SPSS, or SAS), output files will be produced that display the same results and statistical parameters, although their layout may vary slightly.

Example 1: Analysis of distribution of mathematics achievement

This first example shows the necessary steps for analyzing the overall distribution of mathematics achievement. Because this analysis is based on plausible values, users should make sure they are included as selected variables when creating the file using the Merge Module and 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 ASGALLLN1, the IEA IDB Analyzer Analysis Module is used to conduct the analysis outlined in the below steps. The completed analysis module setup for this example is shown in Figure 4.

Figure 4: IEA IDB Analyzer Analysis Module setup for example 1 (analysis of distribution of mathematics achievement)



The following steps can be used to analyze the overall distribution of mathematics achievement:

1. Open the A**nalysis Module** of the IEA IDB Analyzer.
2. Select the data file ASGALLLN1 as the **Analysis File** by clicking the **Select** button.
3. Select LaNA (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 S**how 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** box, which excludes cases with missing values on the grouping variables from the analysis. As the IDCNTRY variable does not contain any missing values, checking this box makes no difference 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. 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) of the same name and in the same folder, with the code necessary to perform the analysis. In Figure 4, the syntax file MAT_Overall.R and the output files with the same name will be created and stored in the C:\LaNA\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.

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, an R data file (*.Rdata), and a 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.

Figure 5 displays the results in the R output with the six participating LaNA 2023 Linking Study countries: Burkina Faso, Egypt, Nigeria, Pakistan, Palestinian National Authority, and Senegal. The results are presented in the "Report" section of the HTML output produced by R.

Figure 5: R output for example 1 (analysis of distribution of mathematics 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
Palestinian National Authority	2744	68517	1713.85	0.65	0.03	406.66	5.86	395 - 418	125.26	3.65	0.00	49
Nigeria	4232	3692377	362673.38	34.94	2.31	334.07	9.57	315 - 353	115.03	4.33	0.00	50
Pakistan	8744	3885405	132370.39	36.77	1.55	379.91	5.90	368 - 391	99.08	2.66	0.00	125
Senegal	4636	268786	9282.18	2.54	0.13	352.65	4.77	343 - 362	88.18	2.71	0.00	61
Egypt	5424	2197349	44811.00	20.79	0.81	394.72	7.79	379 - 410	134.52	3.22	0.00	52
Burkina Faso	4916	454969	24605.75	4.31	0.26	394.17	5.43	384 - 405	73.67	4.05	0.00	50

Each country's results are presented on a single line, with countries ordered sequentially according to their numeric ISO code (see Table 3). Results for "Table Average" may be produced (not shown), based on all countries included in the data file. 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 target-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 subsequent column reports the 95% confidence intervals around the mean. 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 on the analysis

variables—the mathematics achievement (percent missing) and the number of variance strata used for computing standard errors, respectively.

We can see from the results that among the six countries that participated in the LaNA 2023 Linking Study, the Palestinian National Authority had the highest average achievement. As shown in the first line of the table, the Palestinian National Authority had valid data for 2,744 students, and these sampled students represented a population of about 68,517 students, indicated by the sum of the weights. The average math achievement in the Palestinian National Authority was 406.66 (standard error of 5.86), and its 95% confidence interval is shown with 395 as the lower and 418 as the upper boundary. The standard deviation was 125.26 (standard error of 3.65).

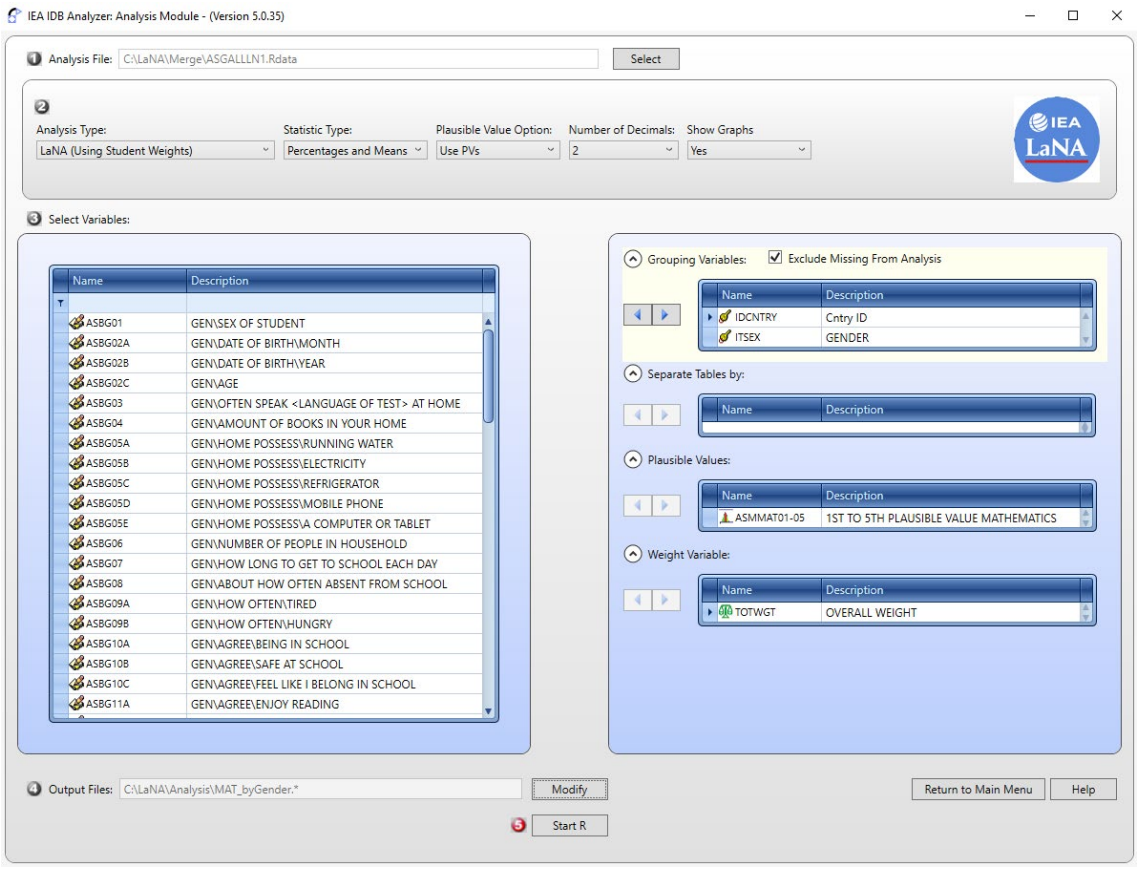
Example 2: Analysis of average mathematics achievement by gender

The second example using student context data creates an analysis that investigates the relationship between students' gender and math achievement. These results are presented below in Figure 6.

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 you to compute the percentages of students in each gender group and their average achievement based on plausible values and their respective standard errors.

The IEA IDB Analyzer Analysis Module is used to conduct this analysis for the six countries participating in the LaNA 2023 Linking Study using the below steps. Figure 6 presents the completed analysis module setup for this example.

Figure 6: IEA IDB Analyzer Analysis Module setup for example 2 (analysis of average achievement by gender)



The following steps can be used to create an analysis that investigates the relationship between student's gender and math achievement:

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the data file ASGALLLN1 as the **Analysis File** by clicking the **Select** button.
3. Select LaNA (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 the 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 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 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) syntax file of the same name and in the same folder, with the code necessary to perform the analysis. In Figure 6, the syntax file MAT_byGender.R and the output files with the same name will be created and stored in the C:\LaNA\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.

The IEA IDB Analyzer produces and saves the results output in three file formats within the folder specified in Step 12—an HTML output file), an R data file (*.Rdata), and a 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 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 and is always created. The additional 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 Figure 7 with the participating LaNA 2023 Linking Study countries Burkina Faso, Egypt, Nigeria, Pakistan, Palestinian National Authority, and Senegal. The results are presented in the "Report" section of the R output.

Countries are ordered sequentially according to their numeric ISO code (see Table 2). 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 (Gender) indicates the category of the grouping variable ITSEX being reported according to the value labels (1: Female, 2: Male). 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 categories 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 subsequent column reports the 95% confidence intervals around the mean. 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.

Figure 7: R output for example 2 (analysis of average achievement by gender)

Report

Analysis for ASMMAT by IDCNTRY ITSEX

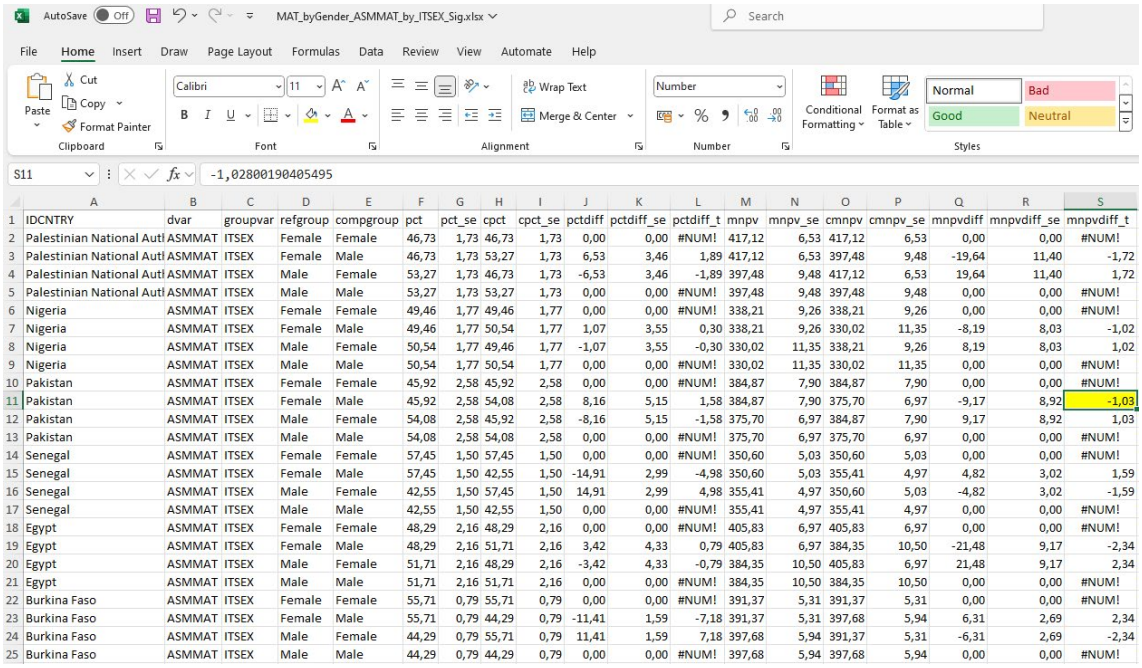
Cntry ID	GENDER	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASMMAT (Mean)	ASMMAT (s.e.)	Confidence Interval (95)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Palestinian National Authority	Female	1382	32021	1452.12	46.73	1.73	417.12	6.53	404 - 430	117.06	3.11	0.00	28
Palestinian National Authority	Male	1362	36496	1474.80	53.27	1.73	397.48	9.48	379 - 416	131.34	5.46	0.00	30
Nigeria	Female	2028	1826345	204245.01	49.46	1.77	338.21	9.26	320 - 356	109.59	4.91	0.00	50
Nigeria	Male	2204	1866031	180331.82	50.54	1.77	330.02	11.35	308 - 352	119.96	4.75	0.00	50
Pakistan	Female	4196	1777983	112397.00	45.92	2.58	384.87	7.90	369 - 400	99.76	3.86	0.00	122
Pakistan	Male	4516	2093990	127442.22	54.08	2.58	375.70	6.97	362 - 389	98.36	3.58	0.00	123
Senegal	Female	2657	154430	7240.33	57.45	1.50	350.60	5.03	341 - 360	88.12	3.07	0.00	61
Senegal	Male	1979	114355	5095.56	42.55	1.50	355.41	4.97	346 - 365	88.19	3.13	0.00	61
Egypt	Female	2629	1061076	52072.51	48.29	2.16	405.83	6.97	392 - 419	126.46	3.42	0.00	52
Egypt	Male	2795	1136272	53043.49	51.71	2.16	384.35	10.50	364 - 405	140.83	3.78	0.00	52
Burkina Faso	Female	2718	253449	14254.63	55.71	0.79	391.37	5.31	381 - 402	72.81	4.05	0.00	50
Burkina Faso	Male	2198	201519	11404.16	44.29	0.79	397.68	5.94	386 - 409	74.59	4.30	0.00	50

The results for Pakistan are interpreted here as an example. Looking at the two lines of results for Pakistan in Figure 7, the population estimates show somewhat more boys than girls: 45.92% of students were girls (standard error of 2.58), and 54.08% were boys (standard error of 2.58). The average mathematics achievement of girls was 384.87 (standard error of 7.90), and it was 375.70 for boys (standard error of 6.97).

While the 95% confidence intervals for boys and girls overlap in Pakistan, the upper boundaries of the confidence interval for girls are eleven points above those for boys. The lower boundaries for the confidence interval associated with girls are 7 points above the lower confidence bound for boys in Pakistan.

The statistical significance of the gender differences can be determined by examining the output file named with the suffix "_sig" (MAT_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 Figure 8.

Figure 8: Excel "Sig" output for example 2 (analysis of average achievement by gender)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
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
2	Palestinian National Autl	ASMMAT	ITSEX	Female	Female	46,73	1,73	46,73	1,73	0,00	0,00	#NUM!	417,12	6,53	417,12	6,53	0,00	0,00	#NUM!
3	Palestinian National Autl	ASMMAT	ITSEX	Female	Male	46,73	1,73	53,27	1,73	6,53	3,46	1,89	417,12	6,53	397,48	9,48	-19,64	11,40	-1,72
4	Palestinian National Autl	ASMMAT	ITSEX	Male	Female	53,27	1,73	46,73	1,73	-6,53	3,46	-1,89	397,48	9,48	417,12	6,53	19,64	11,40	1,72
5	Palestinian National Autl	ASMMAT	ITSEX	Male	Male	53,27	1,73	53,27	1,73	0,00	0,00	#NUM!	397,48	9,48	397,48	9,48	0,00	0,00	#NUM!
6	Nigeria	ASMMAT	ITSEX	Female	Female	49,46	1,77	49,46	1,77	0,00	0,00	#NUM!	338,21	9,26	338,21	9,26	0,00	0,00	#NUM!
7	Nigeria	ASMMAT	ITSEX	Female	Male	49,46	1,77	50,54	1,77	1,07	3,55	0,30	338,21	9,26	330,02	11,35	-8,19	8,03	-1,02
8	Nigeria	ASMMAT	ITSEX	Male	Female	50,54	1,77	49,46	1,77	-1,07	3,55	-0,30	330,02	11,35	338,21	9,26	8,19	8,03	1,02
9	Nigeria	ASMMAT	ITSEX	Male	Male	50,54	1,77	50,54	1,77	0,00	0,00	#NUM!	330,02	11,35	330,02	11,35	0,00	0,00	#NUM!
10	Pakistan	ASMMAT	ITSEX	Female	Female	45,92	2,58	45,92	2,58	0,00	0,00	#NUM!	384,87	7,90	384,87	7,90	0,00	0,00	#NUM!
11	Pakistan	ASMMAT	ITSEX	Female	Male	45,92	2,58	54,08	2,58	8,16	5,15	1,58	384,87	7,90	375,70	6,97	-9,17	8,92	-1,03
12	Pakistan	ASMMAT	ITSEX	Male	Female	54,08	2,58	45,92	2,58	-8,16	5,15	-1,58	375,70	6,97	384,87	7,90	9,17	8,92	1,03
13	Pakistan	ASMMAT	ITSEX	Male	Male	54,08	2,58	54,08	2,58	0,00	0,00	#NUM!	375,70	6,97	375,70	6,97	0,00	0,00	#NUM!
14	Senegal	ASMMAT	ITSEX	Female	Female	57,45	1,50	57,45	1,50	0,00	0,00	#NUM!	350,60	5,03	350,60	5,03	0,00	0,00	#NUM!
15	Senegal	ASMMAT	ITSEX	Female	Male	57,45	1,50	42,55	1,50	-14,91	2,99	-4,98	350,60	5,03	355,41	4,97	4,82	3,02	1,59
16	Senegal	ASMMAT	ITSEX	Male	Female	42,55	1,50	57,45	1,50	14,91	2,99	4,98	355,41	4,97	350,60	5,03	-4,82	3,02	-1,59
17	Senegal	ASMMAT	ITSEX	Male	Male	42,55	1,50	42,55	1,50	0,00	0,00	#NUM!	355,41	4,97	355,41	4,97	0,00	0,00	#NUM!
18	Egypt	ASMMAT	ITSEX	Female	Female	48,29	2,16	48,29	2,16	0,00	0,00	#NUM!	405,83	6,97	405,83	6,97	0,00	0,00	#NUM!
19	Egypt	ASMMAT	ITSEX	Female	Male	48,29	2,16	51,71	2,16	3,42	4,33	0,79	405,83	6,97	384,35	10,50	-21,48	9,17	-2,34
20	Egypt	ASMMAT	ITSEX	Male	Female	51,71	2,16	48,29	2,16	-3,42	4,33	-0,79	384,35	10,50	405,83	6,97	21,48	9,17	2,34
21	Egypt	ASMMAT	ITSEX	Male	Male	51,71	2,16	51,71	2,16	0,00	0,00	#NUM!	384,35	10,50	384,35	10,50	0,00	0,00	#NUM!
22	Burkina Faso	ASMMAT	ITSEX	Female	Female	55,71	0,79	55,71	0,79	0,00	0,00	#NUM!	391,37	5,31	391,37	5,31	0,00	0,00	#NUM!
23	Burkina Faso	ASMMAT	ITSEX	Female	Male	55,71	0,79	44,29	0,79	-11,41	1,59	-7,18	391,37	5,31	397,68	5,94	6,31	2,69	2,34
24	Burkina Faso	ASMMAT	ITSEX	Male	Female	44,29	0,79	55,71	0,79	11,41	1,59	7,18	397,68	5,94	391,37	5,31	-6,31	2,69	-2,34
25	Burkina Faso	ASMMAT	ITSEX	Male	Male	44,29	0,79	44,29	0,79	0,00	0,00	#NUM!	397,68	5,94	397,68	5,94	0,00	0,00	#NUM!

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 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 (female) average and the comparison group (male) 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.

The t-value for the achievement difference between girls and boys in Pakistan is –1.03, 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.

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 target-grade mathematics achievement between girls and boys and determine if it is statistically significant. Thus, the results of this analysis yield the same results, although the output will look slightly different. This example also demonstrates the **Dummy Coding** feature of the IEA IDB Analyzer.

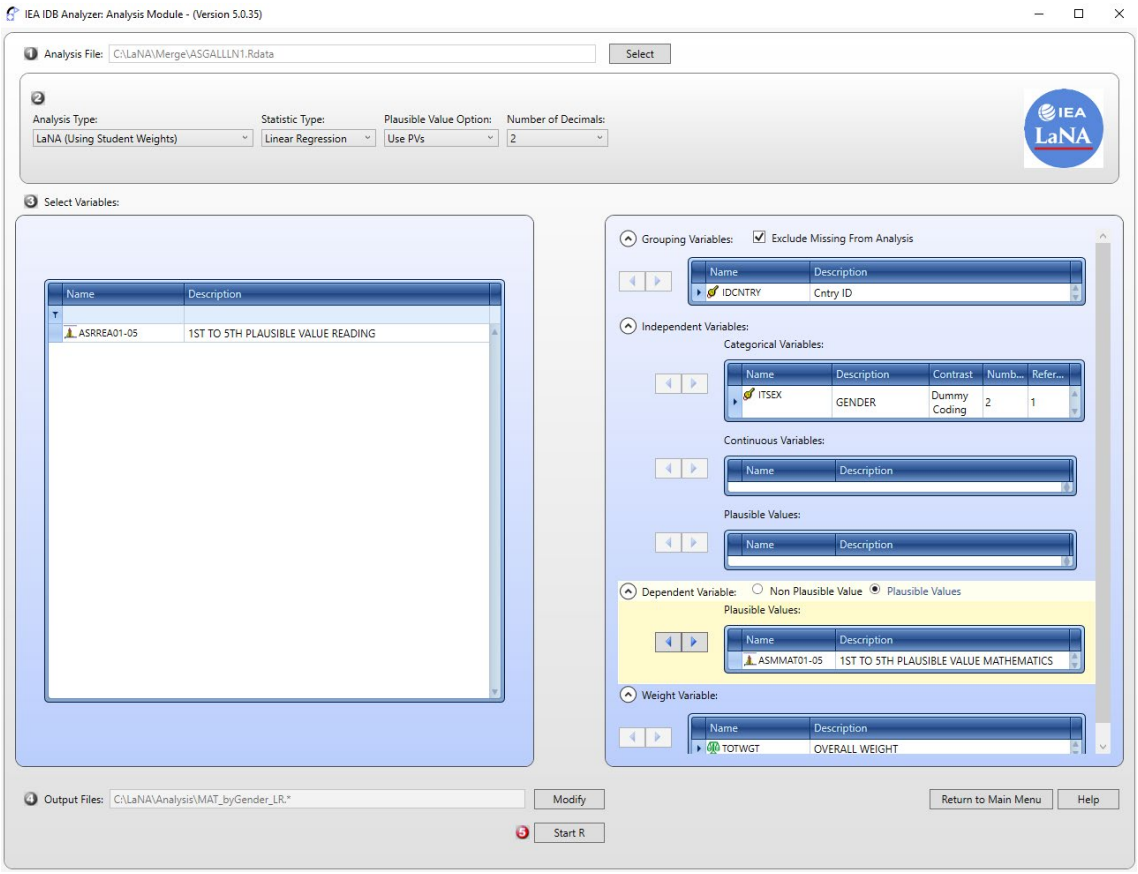
This is a simple example of linear regression, designed to illustrate its connection to the percentages and means used in the previous example. However, regression analysis is far more versatile than mean comparison tests. It allows for examining relationships between variables, handling continuous predictors, controlling for confounding factors, assessing interaction effects, and making predictions.

Additionally, regression can evaluate the relative importance of multiple predictors, address complex data structures (such as hierarchical or longitudinal data), and provide insights into model fit and explanatory power. These capabilities make regression an invaluable tool for understanding and predicting outcomes across diverse research contexts.

The ITSEX variable has a value of one (1) for females and two (2) for males. By using ITSEX as a categorical variable in the IEA IDB Analyzer with **Dummy Coding** and defining category 1 (females) 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 males.

The IEA IDB Analyzer Analysis Module is used to conduct the analysis, with **Linear Regression** defined as the statistic type using the below steps. Figure 9 shows the completed analysis module setup for this example.

Figure 9: IEA IDB Analyzer Analysis Module setup for example 3 (linear regression analysis with student data)



The following steps can be used to create an analysis with linear regression:

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ASGALLLN1 as the **Analysis File** by clicking the **Select** button.
3. Select **LaNA (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 Figure 9 the syntax file MAT_byGender_LR.R and the output files with the same name will be created and stored in the C:\LaNA\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.

When performing a linear regression analysis using the IEA IDB Analyzer, several output files are generated. The primary results include the regression coefficients, which are saved in a file with the suffix "_Coef." Additional output files are created for descriptive statistics by country, labeled with the suffix "_Desc." These files provide information on the intercept (females' average achievement) and the regression coefficients (the change in achievement from girls to males). Another file, with the suffix "_Model" (e.g., MAT_byGender_LR_Model), contains the estimated R-square values for the regression models. All of these results are also included in the HTML output produced by R.

Figure 10 displays the main results for this example analysis—the regression coefficients—in the R output file for the countries participating in the LaNA 2023 Linking Study: Burkina Faso, Egypt, Nigeria, Pakistan, Palestinian National Authority, and Senegal. Countries are ordered numerically according to their numeric ISO code (see Table 2), 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.

Figure 10: R regression coefficient output for example 3 (linear regression analysis with student data)

Regression Coefficients

Cntry ID	Variable	Regression Coefficient	Regression Coefficient (s.e.)	Regression Coefficient (t-value)	Stndrdzd. Coefficient	Stndrdzd. Coefficient (s.e.)	Stndrdzd. Coefficient (t-value)
Palestinian National Authority	(CONSTANT)	417.12	6.53	63.83	NaN	NaN	NaN
Palestinian National Authority	ITSEX_D2	-19.64	11.40	-1.72	-0.08	0.04	-1.75
Nigeria	(CONSTANT)	338.21	9.26	36.53	NaN	NaN	NaN
Nigeria	ITSEX_D2	-8.19	8.03	-1.02	-0.04	0.03	-1.02
Pakistan	(CONSTANT)	384.87	7.90	48.73	NaN	NaN	NaN
Pakistan	ITSEX_D2	-9.17	8.92	-1.03	-0.05	0.04	-1.02
Senegal	(CONSTANT)	350.60	5.03	69.68	NaN	NaN	NaN
Senegal	ITSEX_D2	4.82	3.02	1.59	0.03	0.02	1.60
Egypt	(CONSTANT)	405.83	6.97	58.23	NaN	NaN	NaN
Egypt	ITSEX_D2	-21.48	9.17	-2.34	-0.08	0.03	-2.37
Burkina Faso	(CONSTANT)	391.37	5.31	73.73	NaN	NaN	NaN
Burkina Faso	ITSEX_D2	6.31	2.69	2.34	0.04	0.02	2.38

The countries are identified in the first column (Cntry ID), and the second column (Variable) indicates the intercept (CONSTANT) or the regression coefficient being reported. The third column reports the regression coefficient ('b' in Excel), indicating, for the intercept, the average value of the dependent variable for the reference group (females in this case), and for the regression coefficients, the average difference in the dependent variable from the intercept. The fourth column is the standard error of the regression coefficient (Regression Coefficient (s.e.)). The fifth 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 third, fourth, and fifth columns, whereby the dependent and independent variables are standardized to have a mean of zero (0) and a standard deviation of one (1).

In Figure 10, the first line of results for Pakistan, labeled "(CONSTANT)," for example, indicates the estimated average mathematics achievement of girls in Pakistan: 384.87 with a standard error of 7.90. This estimate concurs with the results obtained in the previous example (Figure 7). From the second line of results labeled "ITSEX_D2," the males in Pakistan had a negative average mathematics achievement difference from girls of −9.17 with an estimated standard error of 8.92. The t-value for the coefficient is −1.03, which is greater than −1.96 (the lower critical value for $\alpha = 0.05$), indicating this achievement difference is not statistically significant. Counting the two regression coefficients together (384.87−9.17) yields the estimated average mathematics achievement of boys in Pakistan, which was 375.70 in Figure 7.

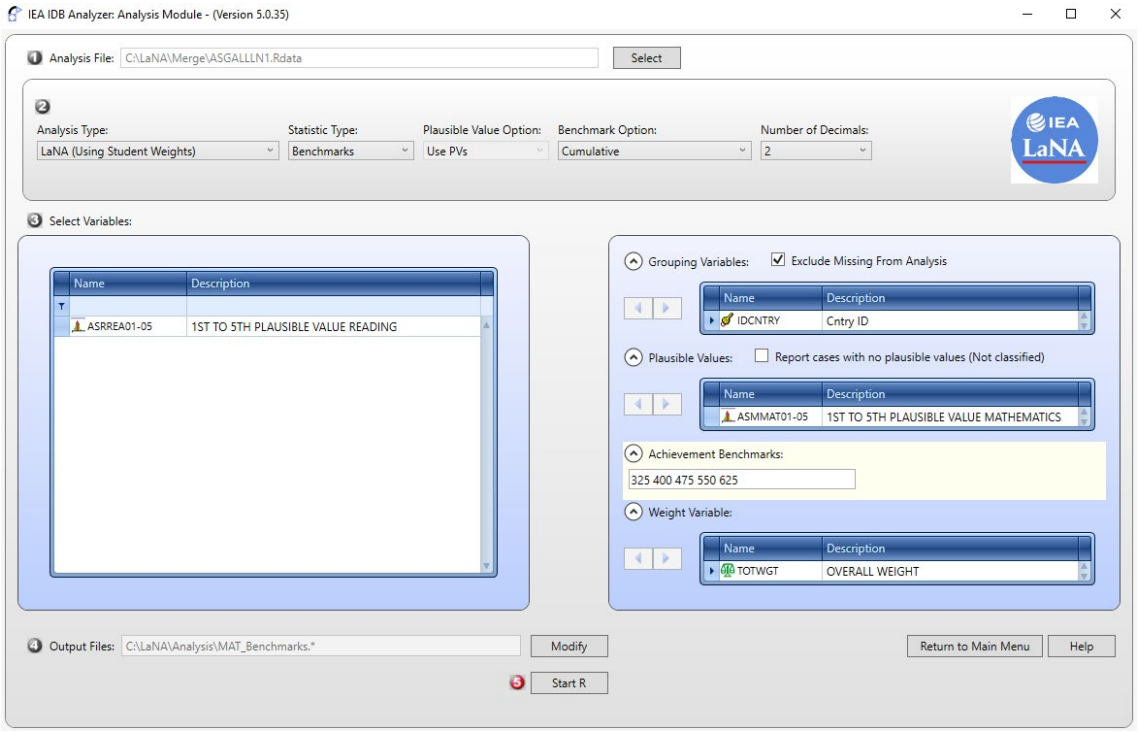
Example 4: Analysis of LaNA international benchmarks

This section describes how to use the IEA IDB Analyzer to perform analyses of student achievement in relation to the LaNA international benchmarks of mathematics achievement. This example computes

the percentages of students reaching each of the five LaNA International Benchmarks of mathematics achievement (advanced, high, intermediate, low, and basic) using the merged ASGALLLN1 data file described earlier under Merging Data Files with the IEA IDB Analyzer. The steps outlined in the following paragraphs are identical to those used for calculating the percentages of students reaching the reading benchmarks.

This example is conducted in the IEA IDB Analyzer Analysis Module using the steps outlined below. The completed analysis module setup is shown in Figure 11.

Figure 11: IEA IDB Analyzer Analysis Module setup for example 4 (analysis of LaNA international benchmarks)



The following steps can be used to create an analysis of student achievement in relation to the LaNA international benchmarks:

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Specify the data file ASGALLLN1 as the **Analysis File** by clicking the S**elect** button.
3. Select **LaNA (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 LaNA 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 LaNA International Benchmarks in ascending order—375, 400, 475, 550, and 625 (basic, low, intermediate, high, and advanced, respectively).

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 Figure 11, the syntax file MAT_Benchmarks.R and the output files with the same name will be created and stored in the folder C:\LaNA\Analysis.
12. Click the **Start R** button 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), an R data file (*.Rdata), and a Microsoft Excel Worksheet (*.xlsx). Graphs are included only in the HTML (or SPSS/SAS) output files. Figure 12 presents the results of Example 4 as shown in the R output, under the "Report" section. Results are shown for the first two countries: the Palestinian National Authority and Nigeria.

Countries are ordered according to their numeric ISO code (see Table 2), and each country's results are displayed on four lines, one for each LaNA 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.)).

Figure 12: R Output for example 4 (analysis of LaNA international benchmarks)

Report

Percentage by Performance Group of ASMMAT

Cntry ID	Performance Group	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)
Palestinian National Authority	1. At or Above 325	2004	50311	1812.34	73.43	1.71
Palestinian National Authority	2. At or Above 400	1465	37107	1667.12	54.16	1.97
Palestinian National Authority	3. At or Above 475	823	21169	1373.25	30.90	1.85
Palestinian National Authority	4. At or Above 550	327	8476	921.08	12.37	1.31
Palestinian National Authority	5. At or Above 625	87	2352	674.15	3.43	0.98
Nigeria	1. At or Above 325	1990	1994305	241048.31	54.01	3.72
Nigeria	2. At or Above 400	1160	1101265	143217.05	29.83	3.14
Nigeria	3. At or Above 475	487	421355	71842.76	11.41	1.97
Nigeria	4. At or Above 550	114	94144	24840.09	2.55	0.71
Nigeria	5. At or Above 625	14	11349	7599.44	0.31	0.21

As shown in the four lines of results listed for Nigeria, an estimated percentage of 54.01% of the students in Nigeria performed at or above the basic international benchmark of 325 scale score points, with a standard error of 3.72; an estimated percentage of 29.83% of students reached the Low International Benchmark, with a standard error of 3.14; an estimated percentage of 11.41% of students reached the intermediate international benchmark, with a standard error of 1.97; an estimated percentage of 2.55% of students reached the high international benchmark, with a standard error of 0.71 and an estimated percentage of 0.31% of students reached the advanced international benchmark, with a standard error of 0.21.

Conducting analyses with LaNA 2023 Linking Study school context data

When analyzing school context data from the LaNA 2023 Linking Study international database, it is preferable that they be analyzed to make inferences about students attending schools with a given characteristic, rather than about schools with 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.

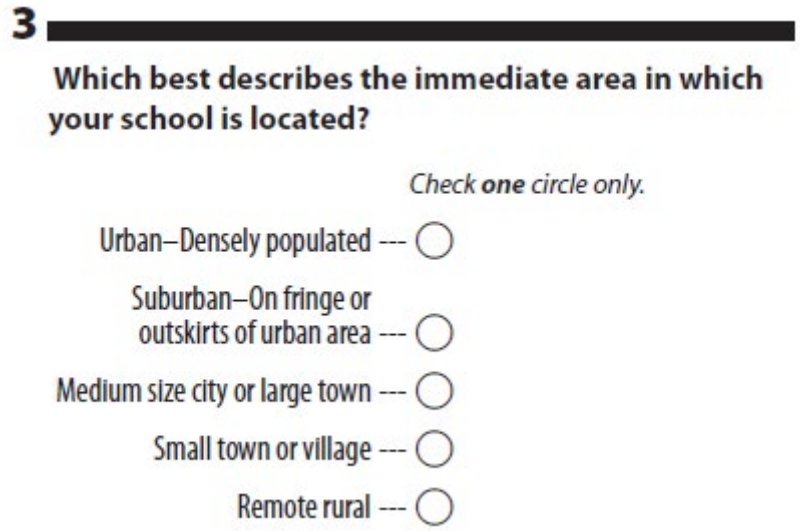
This section presents an analysis conducted using the IEA IDB Analyzer with school context data. Example 5 uses school context data combined with student context data to calculate the percentages and average achievement of target-grade students attending schools with students from different immediate locations.

Example 5: Analysis of average achievement by immediate area of school location

In this example, the **Percentages and Means** statistic type is used along with the **Use PVs** option to estimate the percentages of students with their average mathematics achievement by reporting categories of the immediate area of school location as reported by school principals.

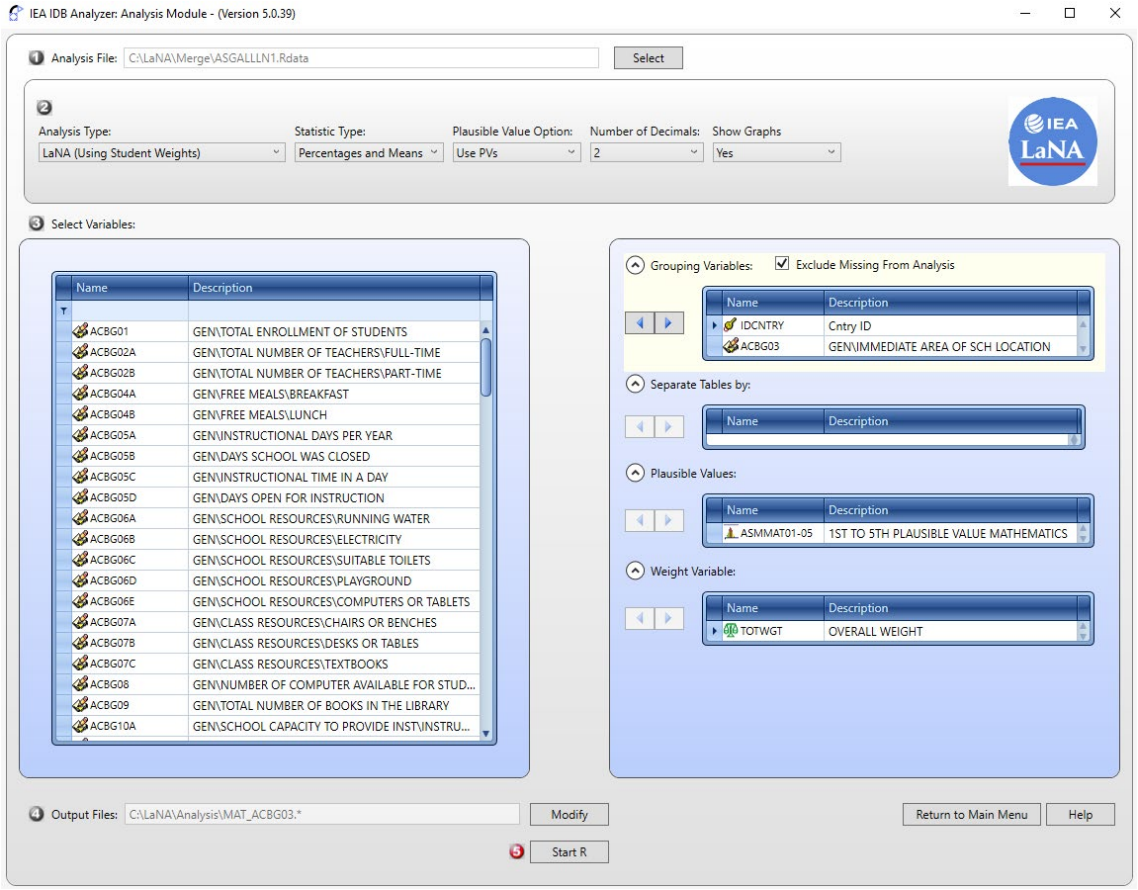
The codebook for the school context data file indicates that the school questionnaire variable ACBG03 contains information on the immediate area of school location in five categories, as shown in Figure 13, which depicts the corresponding question three of the school questionnaire with its five response options.

Figure 13: Question 3 from the LaNA 2023 Linking Study school questionnaire



The completed analysis module setup is shown in Figure 14, using the steps outlined below the figure.

Figure 14: IEA IDB Analyzer Analysis Module setup for example 5 (analysis of average achievement by immediate area of school location)



The following steps can be used to create an analysis of average achievement by immediate area of school location:

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ASGALLLN1 as the **Analysis File** by clicking the **Select** button.
3. Select **LaNA (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 ACBG03.
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 the immediate area of the school location by country, a clustered bar graph of average achievement for each category of school location by country, and a stacked bar graph of average percent of students for each category of school location by country. R also provides separate graphs for each country.
8. Specify the variable ACBG03 as a second grouping variable by first clicking the **Grouping Variables** field to activate it. Then, select ACBG03 from the list of variables in the left panel and move it to the G**rouping 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 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 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 Figure 14, the syntax file MAT_ACBG03.R and the output files with the same name will be created and stored in the folder C:\LaNA\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 Figure 15. The results are presented in the same manner as in Example 2, with countries identified in the first column, and the second column indicates the category of the grouping variable ACBG03 being reported according to the value labels (1: Urban—densely populated; 2: Suburban—on the fringe or outskirts of an urban area; 3: Medium size city or large town; 4: Small town or village; 5: Remote rural). 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 ACBG03 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 subsequent column reports the 95% confidence intervals around the mean. 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.

Figure 15: R output for example 5 (analysis of average achievement by school location)

Report

Analysis for ASMMAT by IDCNTRY ACBG03

Cntry ID	GENAREA OF SCH LOCATION	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
Senegal	Urban–Densely populated	1341	54113	9401.00	20.34	3.44	372.88	12.73	348 to 398	92.78	7.36	0.00	19
Senegal	Suburban–On fringe or outskirts of urban area	1195	58878	9687.63	22.14	3.53	391.20	8.96	374 to 409	79.56	3.10	0.00	19
Senegal	Medium size city or large town	507	33369	10032.67	12.54	3.67	356.12	14.93	327 to 385	82.74	4.49	0.00	13
Senegal	Small town or village	956	64656	10281.06	24.31	3.86	321.92	7.10	308 to 336	78.41	2.54	0.00	25
Senegal	Remote rural	598	54976	9702.40	20.67	3.68	324.59	10.89	303 to 346	85.13	5.83	0.00	20

In this example, each country's results are presented on five lines, one for each value of the ACBG03 variable. There are fewer lines if any category does not have any observations. As shown in the five lines of results for Senegal, 20.34% (standard error of 3.44) of the target-grade students attended schools in "urban-densely populated" areas and their average achievement was 372.88 (standard error of 12.73); 22.14% (standard error of 3.53) attended schools located in "suburban—on fringe or outskirts of urban area" and their average achievement was 391.20 (standard error of 8.96); 12.54% (standard error of 3.67) attended schools in a "medium size city or large town" and their average achievement was 356.12 (standard error 14.93); 24.31% (standard error 3.86) attended schools located in a "small town or village" and their average achievement was 321.92 (standard error 7.10); and 20.67% (standard error of 3.68) attended schools located in "remote/rural" areas and their average achievement was 324.59 (standard error of 10.89).

When examining the 95% confidence intervals, it becomes evident that the first category overlaps with the second and third categories. In contrast, the upper bounds of the fourth and fifth categories fall below the lower bound of the first category. This lack of overlap in confidence intervals suggests that a significance test for the mean difference, with an alpha level of 5%, would likely reject the null hypothesis.

Additional analyses conducted with the IEA IDB Analyzer

Computing correlations

In addition to the analyses described above, the IEA IDB Analyzer can compute correlations among a set of variables. Thus, it can compute correlations between context variables, between achievement scores, and between a combination of both. While these types of analyses are not demonstrated here, the steps for conducting them are like those described previously: select the grouping variables, the analysis variables, the achievement scores (if necessary), and confirm the weight variable.

The output will display, for each group defined by the grouping variables, the correlation coefficients for each possible pair of variables.

Calculating percentiles of a distribution

The **Percentiles** statistic type is an additional tool provided by the IEA IDB Analyzer for analyzing the LaNA 2023 Linking Study data. This procedure will compute the percentiles of a distribution within any specified subgroups, along with appropriate standard errors. The distribution can be either a non-PV-based analysis variable—such as a context questionnaire scale—or a specified set of plausible values.

Performing logistic regression

The IEA IDB Analyzer can perform a logistic regression, with or without plausible values. Logistic regression is used to predict a dichotomous outcome based on one or more predictor variables. Users can specify grouping variables, a dependent variable, and independent variables—with or without interactions—that can be categorical, continuous, or plausible values. Users will find useful information on performing logistic regression in the IEA IDB Analyzer help manual. Note that this analysis is not yet available for R in IEA IDB Analyzer version 5.0.

Important notice on handling missing data in the R, SPSS, and SAS source files

It is strongly recommended to use the IDB Analyzer for all analyses involving LaNA 2023 Linking Study data. However, the source files utilized by the IEA IDB Analyzer can also be opened with corresponding statistical software packages. In such cases, it is important to understand how each software package handles missing data.

SPSS files include numeric user-defined missing values, which are properly recognized and accounted for in any analysis. Similarly, SAS data files use alphanumeric missing codes, which are also treated as missing values during analysis.

However, 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, 999, etc., while "not reached" codes in the achievement files are represented as 6, 96, 996, and so on (see Chapter 2, Section on Missing Data for details).

If you analyze the R data files directly in R or RStudio without addressing 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")

For example, attr(ASGALLLN1$ACBG03, "na_values") returns '9' as the only numeric missing code in variable ACBG03 of the merged school data file ASGALLLN1.Rdata.

Alternatively, you can consult the "Missing Scheme Detailed (SPSS)" column in the LaNA 2023 Linking Study 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.

This is an additional key reason why it is highly recommended to use the IEA IDB Analyzer for your analyses.

Appendix A – International Context Questionnaires

The included questionnaires have the variable name appended. For example, Question 1 of the student questionnaire has the variable name ASBG01.

LaNA 2023 Linking Study international student context questionnaire

1

Are you a girl or a boy?

Fill one circle only.

Girl -- ◯

Boy -- ◯

ASBG01

2

When were you born?

Fill the circles next to the month and year you were born.

a) Month	b) Year	c) I am …
January -- ◯	2009 -- ◯	_____ years old
February -- ◯	2010 -- ◯	
March -- ◯	2011 -- ◯	
April -- ◯	2012 -- ◯	
May -- ◯	2013 -- ◯	
June -- ◯	2014 -- ◯	
July -- ◯	2015 -- ◯	
August -- ◯	Other -- ◯	
September -- ◯		
October -- ◯		
November -- ◯		
December -- ◯		
ASBG02A	ASBG02B	ASBG02C

3

How often do you speak <language of test> at home?

Fill one circle only.

ASBG03

- I always speak <language of test> at home -- ◯
- I almost always speak <language of test> at home -- ◯
- I sometimes speak <language of test> and sometimes speak another language at home -- ◯
- I never speak <language of test> at home -- ◯

4

About how many books are there in your home? (Do not count magazines, newspapers, or your school books.)

Fill one circle only.

ASBG04

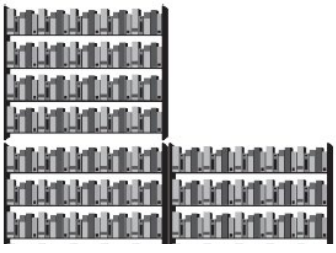
- None or very few (0–10 books) -- ◯ This shows 10 books

- Enough to fill one shelf (11–25 books) -- ◯ This shows 25 books

- Enough to fill one bookcase (26–100 books) -- ◯ This shows 100 books

- Enough to fill two bookcases (101–200 books) -- ◯ This shows 200 books

- Enough to fill three or more bookcases (more than 200) -- ◯ This shows more than 200 books



5

Do you have any of these things at your home?

Fill one circle for each line.

	Yes	No	
a) Running water	○	○	ASBG05A
b) Electricity	○	○	ASBG05B
c) Refrigerator	○	○	ASBG05C
d) Mobile phone	○	○	ASBG05D
e) Computer or tablet	○	○	ASBG05E

6

How many people live in your home? (Do not count yourself.)

Fill one circle only.

- 1–3 ○
- 4–6 ○
- 7–9 ○
- 10 or more ○

ASBG06

7

How long does it take you to get to school each day?

Fill one circle only.

- 30 minutes or less ○
- 30 minutes up to 1 hour ○
- From 1 hour up to 2 hours ○
- 2 hours or more ○

ASBG07

8

About how often are you absent from school?

Fill one circle only.

Once a week -- ◯
Once every two weeks -- ◯
Once a month -- ◯
Once every two months -- ◯
Never or almost never -- ◯

ASBG08

9

How often do you feel this way when you arrive at school?

Fill one circle for each line.

	Every day	Almost every day	Sometimes	Never	
a) I feel tired	◯	◯	◯	◯	ASBG09A
b) I feel hungry	◯	◯	◯	◯	ASBG09B

10

How much do you agree with these statements about school?

Fill one circle for each line.

	Agree a lot	Agree a little	Disagree a little	Disagree a lot	
a) I like being in school	◯	◯	◯	◯	ASBG10A
b) I feel safe when I am at school	◯	◯	◯	◯	ASBG10B
c) I feel like I belong at this school	◯	◯	◯	◯	ASBG10C

11

How much do you agree with these statements about <u>reading</u>?

Fill one circle for each line.

	Agree a lot	Agree a little	Disagree a little	Disagree a lot	
a) I enjoy reading	○	○	○	○	ASBG11A
b) I usually do well in reading	○	○	○	○	ASBG11B

12

How much do you agree with these statements about <u>mathematics</u>?

Fill one circle for each line.

	Agree a lot	Agree a little	Disagree a little	Disagree a lot	
a) I enjoy learning mathematics	○	○	○	○	ASBG12A
b) I usually do well in mathematics	○	○	○	○	ASBG12B

13

How much do you agree with these statements about your <u>reading lessons</u>?

Fill one circle for each line.

	Agree a lot	Agree a little	Disagree a little	Disagree a lot	
a) I know what my teacher expects me to do	○	○	○	○	ASBG13A
b) My teacher is easy to understand	○	○	○	○	ASBG13B
c) I am interested in what my teacher says	○	○	○	○	ASBG13C

14

How much do you agree with these statements about your <u>mathematics lessons</u>?

Fill one circle for each line.

	Agree a lot	Agree a little	Disagree a little	Disagree a lot	
a) I know what my teacher expects me to do	○	○	○	○	ASBG14A
b) My teacher is easy to understand	○	○	○	○	ASBG14B
c) I am interested in what my teacher says	○	○	○	○	ASBG14C

LaNA 2023 Linking Study international school context questionnaire

1

What is the total current enrollment of students at your school?

_____ students
Write in the number.

ACBG01

2

What is the total current number of teachers at your school?

_____ full time teachers
Write in the number.

ACBG02A

_____ part time teachers
Write in the number.

ACBG02B

3

Which best describes the immediate area in which your school is located?

*Check **one** circle only.*

- Urban–Densely populated ---
- Suburban–On fringe or outskirts of urban area ---
- Medium size city or large town ---
- Small town or village ---
- Remote rural ---

ACBG03

4

Does your school provide free meals for students?

*Check **one** circle for each line.*

	Yes, for all students	Yes, for some students	No	
a) Breakfast	○	○	○	ACBG04A
b) Lunch	○	○	○	ACBG04B

5

For the <grade> students in your school:

A. How many days per year is your school scheduled to be open for instruction?

_____ days
Write in the number.

ACBG05A

B. Within the last 12 months, about how many days was your school closed due to unforeseen circumstances (e.g., strikes, bad weather, etc.)?

_____ days
Write in the number.

ACBG05B

C. What is the total instructional time, excluding breaks, in a typical day?

_____ minutes
Write in the number of minutes per day.
Please convert the number of hours into minutes.

ACBG05C

D. In one calendar week, how many days is the school open for instruction?

*Check **one** circle only.*

- 6 days ---
- 5 1/2 days ---
- 5 days ---
- 4 1/2 days ---
- 4 days ---
- Other ---

ACBG05D

6

Does your school have the following resources?

*Check **one** circle for each line.*

		Yes	No
a) Running water	ACBG06A	○	○
b) Electricity	ACBG06B	○	○
c) Suitable toilets or latrines	ACBG06C	○	○
d) Playground	ACBG06D	○	○
e) Computers or tablets for students	ACBG06E	○	○

7

Does your school have the following classroom resources?

*Check **one** circle for each line.*

		Yes	No
a) Chairs or benches	ACBG07A	○	○
b) Desks or tables	ACBG07B	○	○
c) Textbooks for each student	ACBG07C	○	○

8

How many computers (including tablets) does your school have for use by <grade> students?

_____ computers ACBG08

Write in the number.

9

Approximately how many books (print) with different titles does your school library have (exclude magazine and periodicals)?

_____ books ACBG09

Write in the number.

10

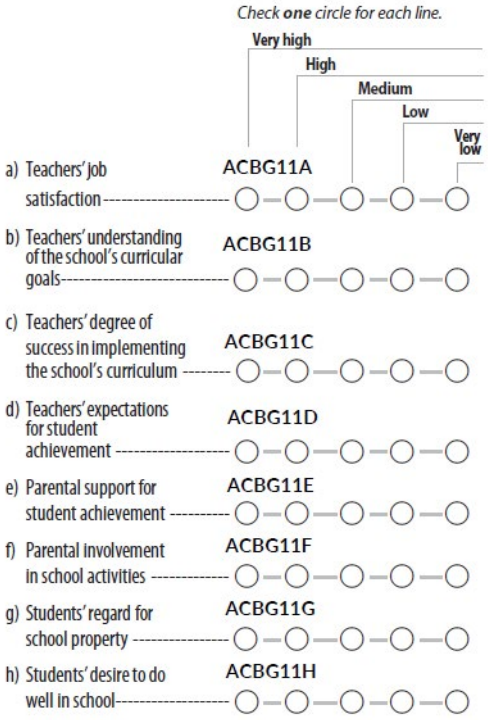
How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following?

*Check **one** circle for each line.*

	Not at all	A little	Some	A lot	
a) Instructional materials (e.g., textbooks)	○	○	○	○	ACBG10A
b) Supplies (e.g., papers, pencils, materials)	○	○	○	○	ACBG10B
c) Qualified teachers	○	○	○	○	ACBG10C

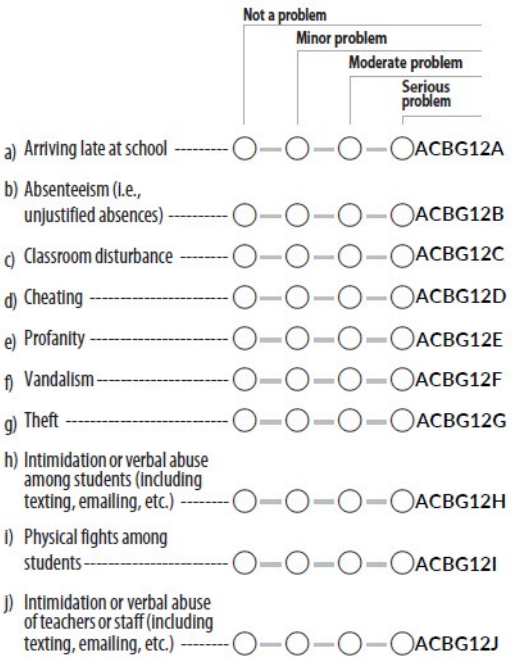
11

How would you characterize each of the following within your school?



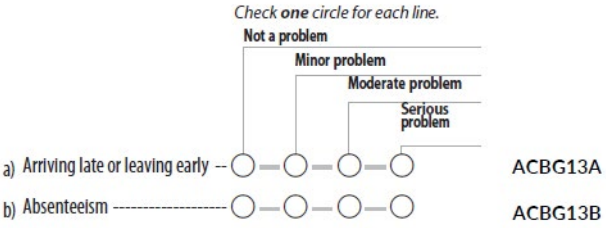
12

To what degree is each of the following a problem among <grade> students in your school?



13

To what degree is each of the following a problem among teachers in your school?



References

IBM Corporation. (2016). IBM SPSS Statistics for Windows (Version 24.0) [Computer software]. Armonk, NY: IBM Corporation.

International Association for the Evaluation of Educational Achievement (IEA). (2024). IEA IDB Analyzer (Version 5.0) [Computer software]. Hamburg, Germany: IEA Hamburg. Retrieved from https://www.iea.nl/data

R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org

RStudio Team (2024). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/

SAS Institute. (2016). SAS for Windows (Version 9.4) [Computer software]. Cary, NC: SAS Institute, Inc.

von Davier, M., Khorramdel, L., Reynolds, K., Aldrich, C. E. A., Bookbinder, A., Kennedy, A., Gonzalez, E. (2025). LaNA 2023 Linking Study Results. Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

BOSTON COLLEGE
timssandpirls.bc.edu

The IEA Literacy and Numeracy Assessment (LaNA) is designed for low- and middle-income countries to generate reliable national data on emerging reading and numeracy skills. Rooted in the well-established TIMSS and PIRLS frameworks, LaNA provides an opportunity for countries to benchmark their students' achievement on the TIMSS mathematics and PIRLS reading scales. By offering internationally comparable data, LaNA supports evidence-based policymaking, evaluates educational interventions, and tracks progress toward SDG Indicator 4.1.1b. Beyond assessment results, LaNA also serves as a capacity-building initiative, equipping national teams with valuable experience in standardized test administration, data collection, and reporting.

The LaNA User Guide provides essential information about the LaNA international database. It outlines the database structure, key variables, and methodologies for weighting and variance estimation in data analysis. Additionally, the guide provides step-by-step instructions for using the IEA's International Database (IDB) Analyzer software, which facilitates the analysis of large-scale assessment data. This guide is accompanied by the international versions of the LaNA background questionnaires.

IEA TIMSS & PIRLS BOSTON COLLEGE