

PIRLS 2021

USER GUIDE FOR THE INTERNATIONAL DATABASE 2nd Edition

Bethany Fishbein
Liqun Yin
Pierre Foy



Copyright © 2024

International Association for the Evaluation of Educational Achievement (IEA)

PIRLS 2021 User Guide for the International Database, 2nd Edition

Bethany Fishbein, Liqun Yin, and Pierre Foy

Suggested citation:

Fishbein, B., Yin, L., & Foy, P. (2024). *PIRLS 2021 User Guide for the International Database* (2nd ed.). Boston College, TIMSS & PIRLS International Study Center. <https://pirls2021.org/data>

Publishers:

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

ISBN: 978-1-889938-69-1

Library of Congress Catalog Card Number: 2023909690

For more information about PIRLS, contact:

TIMSS & PIRLS International Study Center

Lynch School of Education and Human Development

Boston College

Chestnut Hill, MA 02467

United States

tel: +1-617-552-1600

e-mail: timssandpirls@bc.edu

timssandpirls.bc.edu

Boston College is an equal opportunity, affirmative action employer.

The authors are grateful to these individuals for their contributions to this User Guide and its supplements:

Ummugul Bezirhan, Viktoria Böhm, Jessie Bristol, Mark Cockle, Susan Flicop, and Katherine A. Reynolds



TIMSS & PIRLS
International Study Center
Lynch School of Education
BOSTON COLLEGE

Contents

Introduction	1
Overview	1
The PIRLS 2021 User Guide	3
The PIRLS 2021 International Database	3
Exhibit 1: Summary of Contents of the PIRLS 2021 International Database	4
Public Use and Restricted Use Versions of the PIRLS 2021 International Database	4
References	4
Chapter 1: Analyzing PIRLS 2021 Data with the IEA IDB Analyzer	6
Overview	6
About the IEA IDB Analyzer	6
Installing and Launching the IEA IDB Analyzer	7
Exhibit 1.1: IEA IDB Analyzer Main Window	8
Merging Data Files with the IEA IDB Analyzer	8
Exhibit 1.2: IEA IDB Analyzer Merge Module – Select Data Files and Participants	9
Exhibit 1.3: IEA IDB Analyzer Merge Module – Select File Types and Variables	10
Merging Student and Home Context Data Files	11
Merging Student and School Context Data Files	12
Merging Student and Teacher Context Data Files	12
Exhibit 1.4: IEA IDB Analyzer Merge Module – Select File Types and Variables for Merging Student and Teacher Context Data	13
Merged Data Files for the User Guide Examples	14
Conducting Analyses with the IEA IDB Analyzer	15
Statistical Procedures in the IEA IDB Analyzer	15
Definitions of Analysis Variables in the IEA IDB Analyzer	16
Conducting Analyses with PIRLS Student Achievement and Context Data	18
Example 1 – Analysis of Average Reading Achievement	18
Exhibit 1.5: Exhibit of Example 1 – Analysis of Average Reading Achievement (Exhibit 1.3 of <i>PIRLS 2021 International Results in Reading</i>)	19
Exhibit 1.6: IEA IDB Analyzer Analysis Module Setup for Example 1 – Analysis of Average Reading Achievement	20

Exhibit 1.7: R Output for Example 1 – Analysis of Average Achievement.....	22
Example 2 – Analysis of Average Reading Achievement by Gender	23
Exhibit 1.8: Exhibit of Example 2 – Analysis of Average Achievement by Gender (Exhibit 1.5 of <i>PIRLS 2021 International Results in Reading</i>).....	24
Exhibit 1.9: IEA IDB Analyzer Analysis Module Setup for Example 2 – Analysis of Average Achievement by Gender.....	25
Exhibit 1.10: R Output for Example 2 – Analysis of Average Achievement by Gender	28
Exhibit 1.11: Excel “Sig” Output for Example 2 – Analysis of Average Achievement by Gender	29
Example 3 – Linear Regression Analysis with Student Context Data	29
Exhibit 1.12: IEA IDB Analyzer Analysis Module Setup for Example 3 – Linear Regression Analysis with Student Data.....	30
Exhibit 1.13: R Regression Coefficient Output for Example 3 – Linear Regression Analysis with Student Data	32
Example 4 – Analysis of PIRLS International Benchmarks.....	33
Exhibit 1.14: Exhibit of Example 4 – Analysis of PIRLS International Benchmarks (Exhibit 4.2 of <i>PIRLS 2021 International Results in Reading</i>).....	34
Exhibit 1.15: IEA IDB Analyzer Analysis Module Setup for Example 4 – Analysis of PIRLS International Benchmarks.....	35
Exhibit 1.16: R Output for Example 4 – Analysis of PIRLS International Benchmarks.....	37
Conducting Analyses with PIRLS Home Context Data	38
Exhibit 1.17: Exhibit of Example 5 – Analysis of a Context Questionnaire Scale with Home Context Data (Exhibit 5.3 of <i>PIRLS 2021 International Results in Reading</i>).....	39
Example 5 – Analysis of a Context Questionnaire Scale with Home Context Data	40
Step 1: Compute Average Scale Score.....	40
Exhibit 1.18: IEA IDB Analyzer Analysis Module Setup for Example 5 – Analysis of a Context Questionnaire Scale with Home Context Data (Step 1).....	41
Exhibit 1.19: R Output for Example 5 – Analysis of a Context Questionnaire Scale with Home Context Data (Step 1)	43
Step 2: Compute Percentages of Students and Average Achievement by Scale Category	43
Exhibit 1.20: IEA IDB Analyzer Analysis Module Setup for Example 5 – Analysis of a Context Questionnaire Scale with Home Context Data (Step 2).....	44
Exhibit 1.21: R Output for Example 5 – Analysis of a Context Questionnaire Scale with Home Context Data (Step 2)	46
Conducting Analyses with PIRLS School Context Data	47
Example 6 – Analysis of Average Achievement by School Socioeconomic Composition.....	47
Exhibit 1.22: Exhibit of Example 6 – Analysis of Average Achievement by School Socioeconomic Composition (Exhibit 6.1 of <i>PIRLS 2021 International Results in Reading</i>)	48

Exhibit 1.23: IEA IDB Analyzer Analysis Module Setup for Example 6 – Analysis of Average Achievement by School Socioeconomic Composition	49
Exhibit 1.24: R Output for Example 6 – Analysis of Average Achievement by School Socioeconomic Composition	51
Conducting Analyses with PIRLS Teacher Context Data	52
Example 7 – Analysis of Average Achievement by Teachers’ Formal Education	53
Exhibit 1.25: IEA IDB Analyzer Analysis Module Setup for Example 7 – Analysis of Average Achievement by Teachers’ Formal Education	54
Exhibit 1.26: R Output for Example 7 – Analysis of Average Achievement by Teachers’ Formal Education	56
Additional Analyses Conducted with the IEA IDB Analyzer	57
Computing Correlations	57
Calculating Percentiles of a Distribution	58
Performing Logistic Regression	58
References	58

Chapter 2: Contents and Structure of the PIRLS 2021 International Database 60

Overview	60
Exhibit 2.1: Contents of the PIRLS 2021 International Database	60
Exhibit 2.2: Countries in the PIRLS 2021 International Database	61
Supplements to the PIRLS 2021 User Guide	63
Restricted Use Version of the PIRLS 2021 International Database	64
Exhibit 2.3: Variables Available in the Restricted Use Version of the PIRLS 2021 International Database	64
Achievement Items Documentation	65
Item Information	65
IRT Item Parameters and Scale Transformation Constants	66
Item Percent Correct Statistics	66
PIRLS 2021 Data Files	66
Exhibit 2.4: PIRLS 2021 Data File Names	67
Student Achievement Data Files (ASA)	68
Item Variable Naming Convention	68
Item Response Codes	69
Derived Items	69
Codes for Missing Values	70

Modifications to Achievement Item Data	71
PIRLS Achievement Variables	71
Exhibit 2.5: PIRLS 2021 Achievement Scales	72
PIRLS International Benchmarks of Reading Achievement.....	73
Exhibit 2.6: PIRLS 2021 International Benchmarks of Reading Achievement	73
Exhibit 2.7: PIRLS 2021 International Benchmark Variable Codes	74
Non-Response Indicator Variables	74
Student Process Data Files (ASP)	75
Within-Country Scoring Reliability Data Files (ASR)	76
Scoring Reliability Variable Naming Convention	77
Scoring Reliability Codes.....	77
Exhibit 2.8: PIRLS 2021 Score Agreement Variable Codes	77
Context Data Files	77
Student Context Data Files (ASG).....	78
Home Context Data Files (ASH)	78
School Context Data Files (ACG)	79
Teacher Context Data Files (ATG)	79
Student-Teacher Linkage Data Files (AST)	80
Curriculum Data	80
Context Variable Naming Convention.....	80
Exhibit 2.9: PIRLS 2021 Context Variable and Question Location Naming Convention	82
Context Question Location Convention.....	82
Context Variable Response Codes	82
Codes for Missing Values	82
Context Questionnaire Scales and Derived Variables.....	84
Sampling and Weighting Variables.....	84
Exhibit 2.10: PIRLS 2021 Sampling and Weighting Variables.....	85
Exhibit 2.11: Locations of Sampling and Weighting Variables in PIRLS 2021 Data Files.....	86
Structure and Design Variables	87
Identification Variables	87
Exhibit 2.12: Location of Identification Variables in PIRLS 2021 Data Files	89
Tracking Variables	89
Codebooks.....	91
Data Almanacs	91
Achievement Data Almanacs	91

Exhibit 2.13: PIRLS 2021 Achievement Data Almanacs	92
Enjoyment Almanacs	93
Context Data Almanacs.....	93
Exhibit 2.14: PIRLS 2021 Context Data Almanacs.....	94
Special R, SPSS, and SAS Programs.....	95
References.....	96
Chapter 3: Special R, SPSS, and SAS Programs	98
Overview	98
Scoring the PIRLS 2021 Items.....	98
Exhibit 3.1: The ASASCRR5.R Program (Condensed).....	100
Exhibit 3.2: The ASASCRR5.SPS Program (Condensed)	101
Exhibit 3.3: The ASASCRR5.SAS Program (Condensed)	102
References.....	103
Appendix: Organizations and Individuals Responsible for PIRLS 2021	104

Introduction

Overview

PIRLS (the Progress in International Reading Literacy Study) is an international comparative study of students' reading achievement. Conducted on a regular five-year cycle since 2001, PIRLS assesses the reading achievement of students in their fourth year of schooling and has accumulated over 20 years of trend measurements—in 2001, 2006, 2011, 2016, and 2021. PIRLS collects a rich array of information about the national, home, school, and classroom contexts in which students learn to read as well as read to learn. These context data provide international comparative perspectives about the educational factors related to reading achievement.

In 2021, PIRLS transitioned to computer-based assessment. Approximately half the PIRLS 2021 participants opted to administer the digital version, called “digitalPIRLS,” and the other half continued to administer PIRLS with paper booklets (paperPIRLS) as in previous cycles. The digitalPIRLS assessment included electronic versions of the 18 text and item sets that were also in paperPIRLS, plus five online informational ePIRLS tasks. The digitalPIRLS 2021 achievement data were scaled to allow for relative comparisons of performance between digitalPIRLS and paperPIRLS assessments, while also maintaining comparability of trend measurements over time (Bezirhan et al., 2023; Yin et al., 2023a, 2023b). To provide a bridge between digitalPIRLS and paperPIRLS, digitalPIRLS countries also administered the paperPIRLS trend texts and items to a separate sample of students. The PIRLS 2021 Bridge data formed an intermediate link (or “bridge”) between digitalPIRLS countries' computer-based data in 2021 and their paper-based data in 2016, as well as the data from paperPIRLS 2021 countries.

PIRLS 2021 coincided with the height of the COVID-19 pandemic, which subsequently introduced many complexities in analyzing and interpreting the data. Disruptions to school operations during the COVID-19 pandemic necessitated delaying data collection in a number of participating countries, resulting in the PIRLS 2021 data collection occurring over a two-year period instead of only a few months as is typical. Most countries managed to administer the PIRLS 2021 assessment to the targeted fourth grade students at the end of the school year in 2020 or 2021, according to the original plan. However, six countries and one benchmarking participant assessed the next fourth grade cohort of students one year later. In 14 countries and three benchmarking entities, assessment of the fourth grade cohort was delayed until the beginning of the next school year in the fifth grade. These countries had

samples of markedly older students compared to their PIRLS 2016 counterparts, and require caution in making appropriate comparisons (see [Impacts of Modifying the Assessment Schedule on Students' Achievement](#); Mullis et al., 2023).

The PIRLS 2021 International Database makes the data collected and analyzed by the PIRLS 2021 project available to researchers, analysts, and other users to support and promote secondary analysis aimed at improving reading literacy education at the fourth grade level. The database includes student achievement data as well as student, home, school, teacher, and national context data for 57 countries and eight benchmarking participants. Across participating countries and benchmarking entities, including the bridge data, the database includes records for 395,295 students, 384,001 parents, 13,509 school principals, 19,741 teachers, and the National Research Coordinators of each participating country. All participating countries gave IEA Hamburg permission to release their national data contained in the PIRLS 2021 International Database.

For trend countries that participated in previous assessments, PIRLS 2021 provides trends for up to five cycles—2001, 2006, 2011, 2016, and 2021. Countries that participated in digitalPIRLS and administered the paper bridge have data points collected in the two modes of administration. In countries new to PIRLS, the 2021 results can help policy makers and practitioners assess their comparative standing and gauge the rigor and effectiveness of their reading programs. Results of the assessments conducted in 2021 are reported in [PIRLS 2021 International Results in Reading](#) (Mullis et al., 2023).

PIRLS 2021 was an ambitious and demanding study, involving complex procedures for drawing student samples, assessing students' achievement, analyzing the data, and reporting the results. To work effectively with PIRLS data, it is necessary to understand the characteristics of the study, which are described fully in [Methods and Procedures: PIRLS 2021 Technical Report](#) (von Davier et al., 2023). It is intended that this User Guide be used in conjunction with the technical documentation. While the User Guide describes the organization and contents of the database, the chapters of the technical report provide the rationale for the techniques used and for the measures created in the process of data collection and compilation. In particular, differences in national samples due to disruptions to data collection because of the COVID-19 pandemic should be considered when analyzing and interpreting the PIRLS 2021 data. Throughout this User Guide, references are provided to specific technical report chapters and to other PIRLS 2021 documentation and materials where relevant.

The PIRLS 2021 User Guide

This User Guide describes the content and format of the data in the PIRLS 2021 International Database and presents example analyses with the data. Following this introduction, the User Guide includes the following chapters:

- **Chapter 1** introduces the IEA International Database (IDB) Analyzer Software (IEA, 2023) and presents examples of analyses with the PIRLS 2021 data using this software in conjunction with R/RStudio (R Core Team, 2020; RStudio Team, 2020), SPSS (IBM Corporation, 2016), and SAS (SAS Institute, 2016).
- **Chapter 2** serves as a reference for details about the structure and contents of the PIRLS 2021 International Database, including detailed descriptions of the various data files, conventions for naming data files and variables, and descriptions of all the supporting documentation provided with the International Database.
- **Chapter 3** describes special R, SPSS, and SAS programs needed to make full use of the PIRLS 2021 International Database, including programs to score the achievement items according to the assigned item response codes.

The User Guide is accompanied by the following supplements:

- **Supplement 1:** International Versions of the PIRLS 2021 Context Questionnaires
- **Supplement 2:** National Adaptations to the PIRLS 2021 Context Questionnaires
- **Supplement 3:** Variables Derived from the PIRLS 2021 Student, Home, School, and Teacher Context Data

The User Guide and its supplements are available on the PIRLS 2021 International Database and User Guide webpage: <https://pirls2021.org/data>. The primary purpose of this User Guide is to introduce users to the PIRLS 2021 International Database and demonstrate the basic functionality of the IEA IDB Analyzer through simple examples of results published in *PIRLS 2021 International Results in Reading*. The IEA IDB Analyzer comes with its own manual, available through the Help Module, which describes the full functionality and features of the IEA IDB Analyzer. This User Guide also provides references to other PIRLS 2021 publications and documentation to facilitate proper interpretation of data analysis results.

The PIRLS 2021 International Database

The PIRLS 2021 International Database is available on the Boston College, TIMSS & PIRLS International Study Center's website: <https://pirls2021.org/data>. The database also is available

through IEA’s Data Repository: <https://iea.nl/data>. The International Database contains the PIRLS 2021 student achievement data files, student, home, school, and teacher context data files, curriculum data files, and support materials. Exhibit 1 describes the general structure of the International Database, with a brief description of the support materials available for download.

Exhibit 1: Summary of Contents of the PIRLS 2021 International Database

User Guide	User Guide and Supplements
Achievement Items Documentation	PIRLS 2021 item information, IRT item parameters, and percent correct statistics
International Database	
Data (R, SPSS, SAS)	PIRLS 2021 student, home, school, and teacher data files
Curriculum Data	PIRLS 2021 Curriculum Questionnaire data Excel files
Codebooks	Codebook files describing all variables in the PIRLS 2021 International Database
Data Almanacs	Summary statistics for all PIRLS 2021 achievement items and context variables
Special Programs	R, SPSS, and SAS programs to score achievement items

Public Use and Restricted Use Versions of the PIRLS 2021 International Database

The PIRLS 2021 International Database is available in two versions: a public use version and a restricted use version. In the public use version, some variables are excluded to minimize the risk of disclosing confidential information. The list of variables removed from the public use version is given in Chapter 2 of this User Guide. The public use version is available for immediate access from the [PIRLS 2021 International Database webpage](#), and users should be able to replicate all published PIRLS 2021 results with this version of the PIRLS 2021 International Database. Users who require any of the removed variables to conduct their analyses should contact IEA through the [IEA Study Data Repository](#) to obtain permission and access to the restricted use version of the PIRLS 2021 International Database.

References

Bezirhan, U., Foy, P., & von Davier, M. (2023). PIRLS 2021 achievement scaling methodology: Item Response Theory, population models, and linking across modes. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 10.1-10.19). Boston College, TIMSS & PIRLS International Study Center.
<https://doi.org/10.6017/lse.tpisc.tr2103.kb3131>

- 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). (2023). IEA IDB Analyzer (Version 5.0) [Computer software]. Hamburg, Germany: IEA Hamburg. Retrieved from <https://www.iea.nl/data>
- Mullis, I. V. S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., & Wry, E. (2023). *PIRLS 2021 International Results in Reading*. Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb5342>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>
- RStudio Team (2020). 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., Mullis, I. V. S., Fishbein, B., & Foy, P. (Eds.). (2023). *Methods and Procedures: PIRLS 2021 Technical Report*. Boston College, TIMSS & PIRLS International Study Center. <https://pirls2021.org/methods>
- Yin, L., Bezirhan, U., Fishbein, B., & Foy, P. (2023a). Implementing the PIRLS 2021 achievement scaling methodology. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 11.1-11.82). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb3067>
- Yin, L., Fishbein, B., Bezirhan, U., Foy, P., & von Davier, M. (2023b). Examining country-level differences between digitalPIRLS data and bridge data. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 12.1-12.33). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb9281>

CHAPTER 1

Analyzing PIRLS 2021 Data with the IEA IDB Analyzer

Overview

This chapter describes the general use of the IEA's IDB Analyzer software (IEA, 2023) for analyzing the PIRLS 2021 data. Used in conjunction with R (R Core Team, 2020), 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 PIRLS 2021 International Database. The software seamlessly accounts for the sampling information and the multiple imputed achievement scores to produce accurate statistical results.

Seven example analyses presented in this chapter illustrate some of the capabilities of the IEA IDB Analyzer (Version 5.0) to compute a variety of statistics, including means and percentages of students in specified subgroups, mean student achievement in specified subgroups, regression coefficients, and percentages of students reaching the PIRLS 2021 International Benchmarks of Reading Achievement. The examples use student, home, school, and teacher context data files to replicate some of the PIRLS 2021 results included in [*PIRLS 2021 International Results in Reading*](#) (Mullis et al., 2023).

Users should be able to perform statistical analyses with the IEA IDB Analyzer with a basic knowledge of the PIRLS 2021 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 API (application programming interface) for R, SPSS, and SAS—all well-known statistical analysis software. The IEA IDB Analyzer enables users to combine data files from IEA's large-scale assessments and conduct analyses using R, SPSS, or SAS without actually writing programming code. The IEA IDB

Analyzer generates R, SPSS, or SAS syntax that accounts for information from the sampling design in the computation of statistics and their standard errors. In addition, the generated syntax makes appropriate use of plausible values for calculating estimates of achievement scores and their standard errors, combining both sampling variance and imputation variance. Chapter 13 of *Methods and Procedures: PIRLS 2021 Technical Report* provides details about estimating standard errors in the PIRLS 2021 results (Foy & Almaskut, 2023).

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

R and RStudio can be downloaded and installed at no cost from [r-project.org](https://www.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

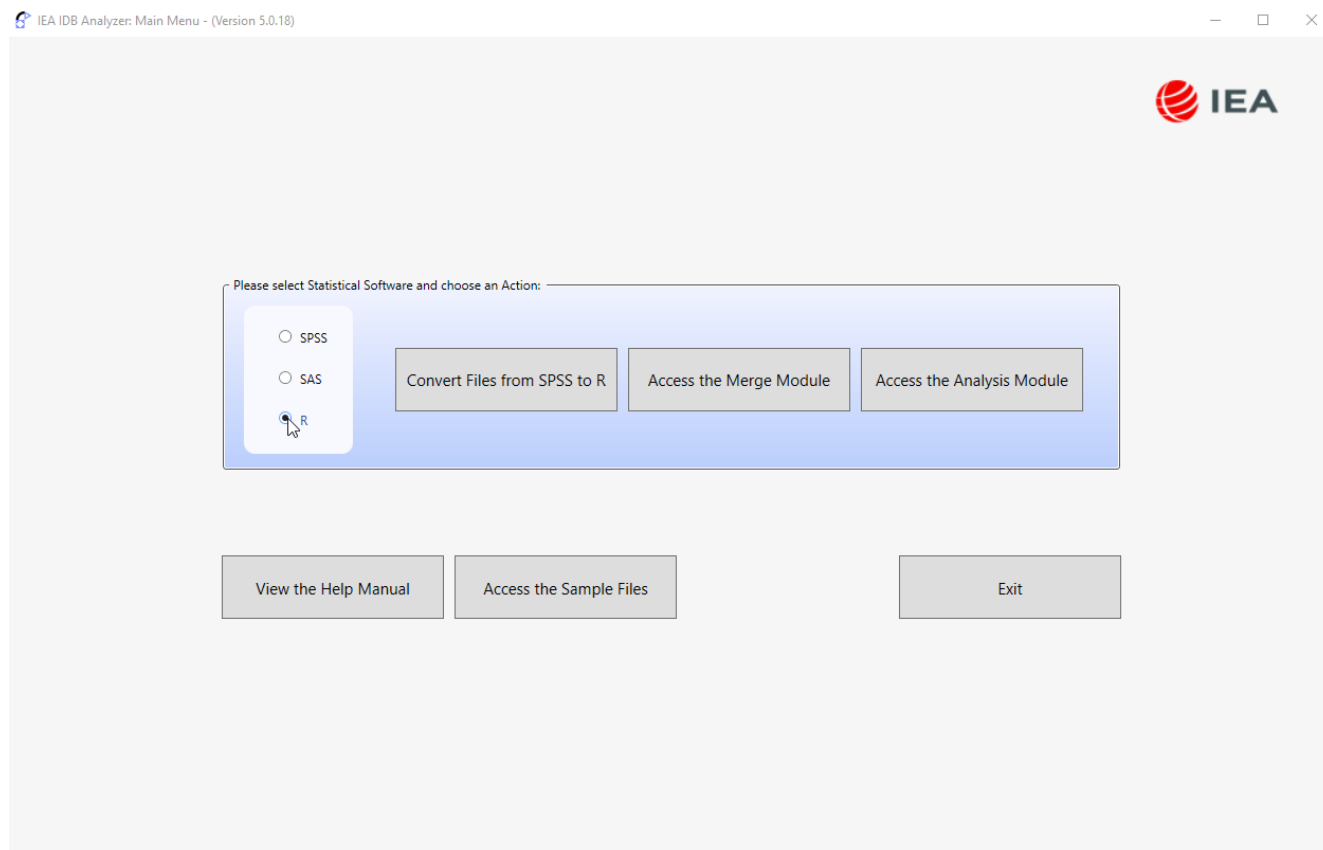
The latest version of the IEA IDB Analyzer—version 5.0—is available for download from the [IEA Data and Tools website](https://iea-data-and-tools.org). When the IEA IDB Analyzer application is launched, the main window will appear, as shown in Exhibit 1.1. Users are first directed to choose SPSS, SAS, or R as their statistical software of choice. The examples in this chapter primarily use R and RStudio.¹ 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 directs users to **Convert Files from SPSS to R**, **Access the Merge Module**, **Access the Analysis Module**, **View the Help Manual**, **Access the Sample Files**, or simply **Exit** the application.

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

¹ The *PIRLS 2016 User Guide for the International Database* (Foy, 2017) presents parallel examples using SAS. The *TIMSS 2019 User Guide for the International Database* (Fishbein et al., 2021) presents parallel examples with TIMSS data using SPSS.

Exhibit 1.1: IEA IDB Analyzer Main Window



Merging Data Files with the IEA IDB Analyzer

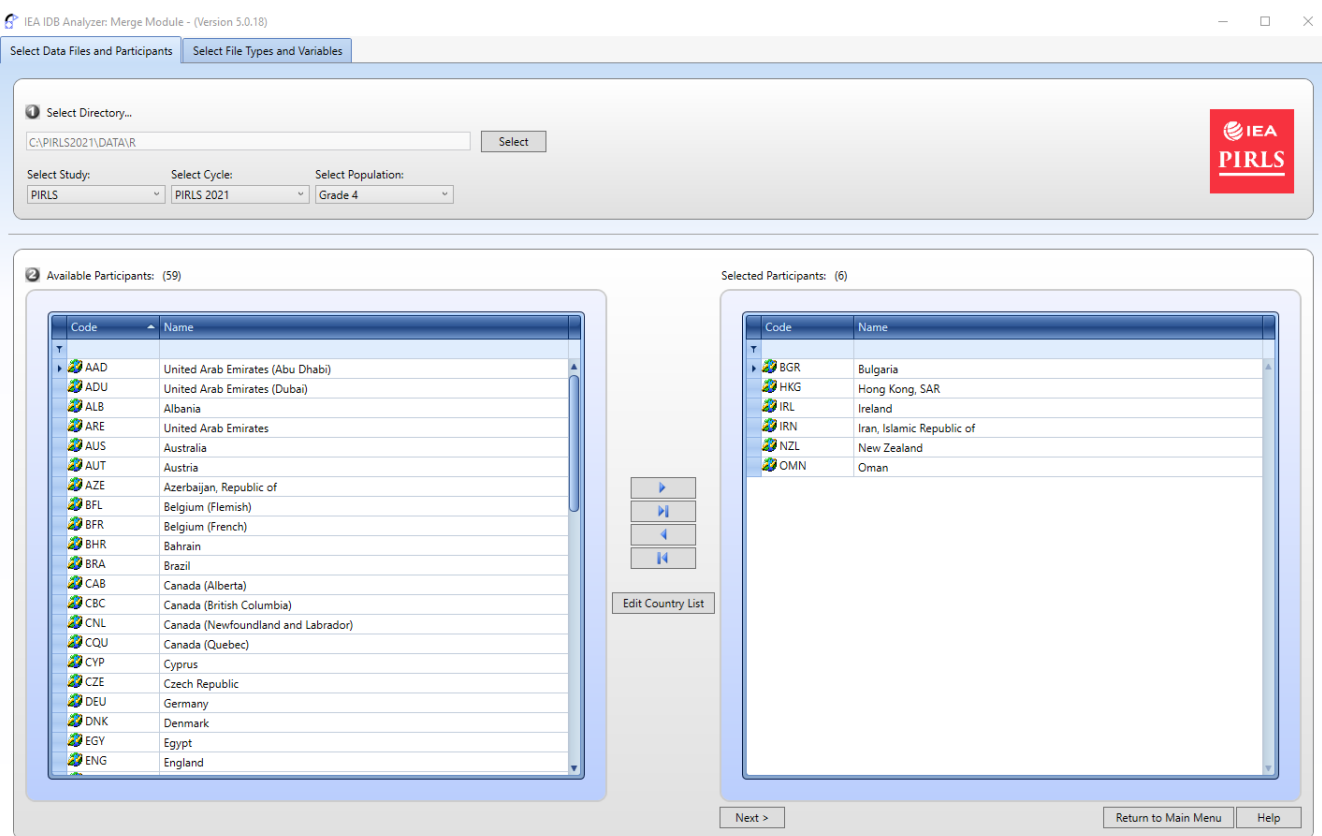
The IEA IDB Analyzer uses the data files available from the TIMSS & PIRLS International Study Center's [PIRLS 2021 International Database webpage](#). The PIRLS 2021 data files are disseminated separately by file type (i.e., data source) and by country. In addition to allowing users to combine like datasets from more than one country for cross-country analyses, the Merge Module allows for the combination of data from different sources (e.g., student, home, school, and teacher) into one R, SPSS, or SAS dataset for subsequent analysis.

Before doing any statistical analysis with the PIRLS 2021 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 within the folder C:\PIRLS2021\DATA\R.

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

1. Start the IEA IDB Analyzer and click the **Merge Module** button.
2. Under the **Select Data Files and Participants** tab and in the **Select Directory** field, browse to the folder where all data files are located. All files must be in the same folder. For example, in Exhibit 1.2, all R data files are located in the folder C:\PIRLS2021\DATA\R. The program will automatically recognize and complete the **Select Study**, **Select Cycle**, and **Select Population** fields and list all countries available in this folder as possible to merge. If the folder contains data from more than one IEA study (e.g., PIRLS, TIMSS) or cycle (e.g., PIRLS 2016, PIRLS 2021, PIRLS 2021 Bridge), users should select the desired combination.

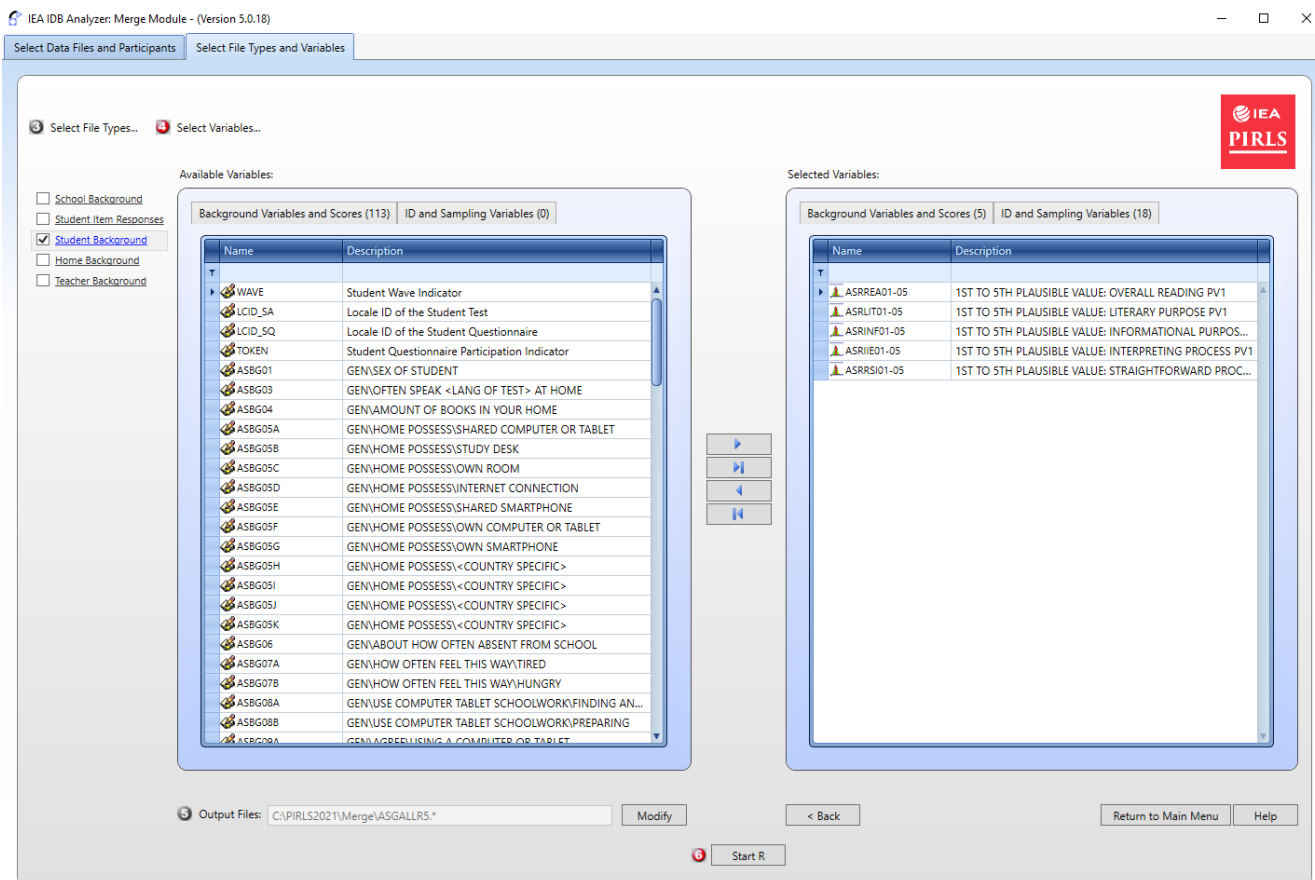
Exhibit 1.2: IEA IDB Analyzer Merge Module – Select Data Files and Participants



3. Click a country of interest from the **Available Participants** list and click the **right arrow** (►) button to move it to the **Selected Participants** panel. Individual countries can be moved directly to the **Selected Participants** panel by double-clicking on the row. To select multiple countries, hold the **Ctrl** key on the keyboard when clicking countries. Click the **tab-right arrow** (►|) button to move all countries to the **Selected Participants** panel. In Exhibit 1.2, Bulgaria, Hong Kong SAR, Ireland, Islamic Republic of Iran, New Zealand, and Oman are selected.

- Click the **Next >** button to proceed to the next step. The software will open the **Select File Types and Variables** tab of the Merge Module, as shown in Exhibit 1.3, to select the file types and the variables to be included in the merged data file.


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



- Select the files for merging by checking the appropriate boxes to the left of the window. For example, in Exhibit 1.3, the box next to **Student Background** is checked, indicating the PIRLS student context data files are selected.²
- Select the variables of interest from the **Available Variables** list in the left panel. The codebook files (described in Chapter 2 of this User Guide) as well as [Supplement 1](#) to this User Guide provide the variable names for storing the data from all questions in the PIRLS 2021 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

² The IEA IDB Analyzer uses the term “background” when referring to context data files or variables.

Selected Variables list. Note that there are two tabs under the **Selected Variables** list: **Background Variables and Scores** and **ID and Sampling Variables**. All achievement scores and all identification, tracking, and sampling variables are selected by default.

7. Specify the desired name for the merged data file and the folder where it will be stored in the **Output Files** field by clicking the **Define** (or **Modify**) button. The IEA IDB Analyzer will create an R script (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the merge. In the example shown in Exhibit 1.3, the R script file ASGALLR5.R and the merged data file ASGALLR5.Rdata both will be created and stored in the folder C:\PIRLS2021\Merge. The merged data file will contain all the variables listed in the **Selected Variables** panel on the right.
8. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** () button (or select **Submit** in the **Run** menu).

Once R, SPSS, or SAS has completed its execution, it is important to check the software output window or log file for possible warnings. If warnings appear, they should be examined carefully, as they might indicate that the merge process was not performed properly and that the resulting merged data file might not be as expected.³

Merging Student and Home Context Data Files

The parents of the students participating in PIRLS 2021 responded to the home questionnaire, also called the “Early Learning Survey.” Their responses are included in the home context data files (named beginning with “ASH”). Although home context variables are located in their own files, they are in essence attributes of the students and must be analyzed in the same manner as student context variables. This will require users to merge the home context data files with the student context data files by selecting both the **Home Background** and **Student Background** file types in the Merge Module of the IEA IDB Analyzer. This is an important step to ensure the proper weights and achievement variables are included for analyses. Variables of interest to be included in the merged data file are chosen separately by

³ For more information on how to use the IEA IDB Analyzer, and for troubleshooting, users should consult the Help manual.

file type, as described below in the section on *Merging Student and Teacher Context Data Files*.

Merging Student and School Context Data Files

Because PIRLS 2021 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, rather than as elements in their own right. Therefore, analyzing school context data should be 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 to be included in the merged data file are selected separately by file type, as described below in the next section: *Merging Student and Teacher Context Data Files*.

Merging Student and Teacher Context Data Files

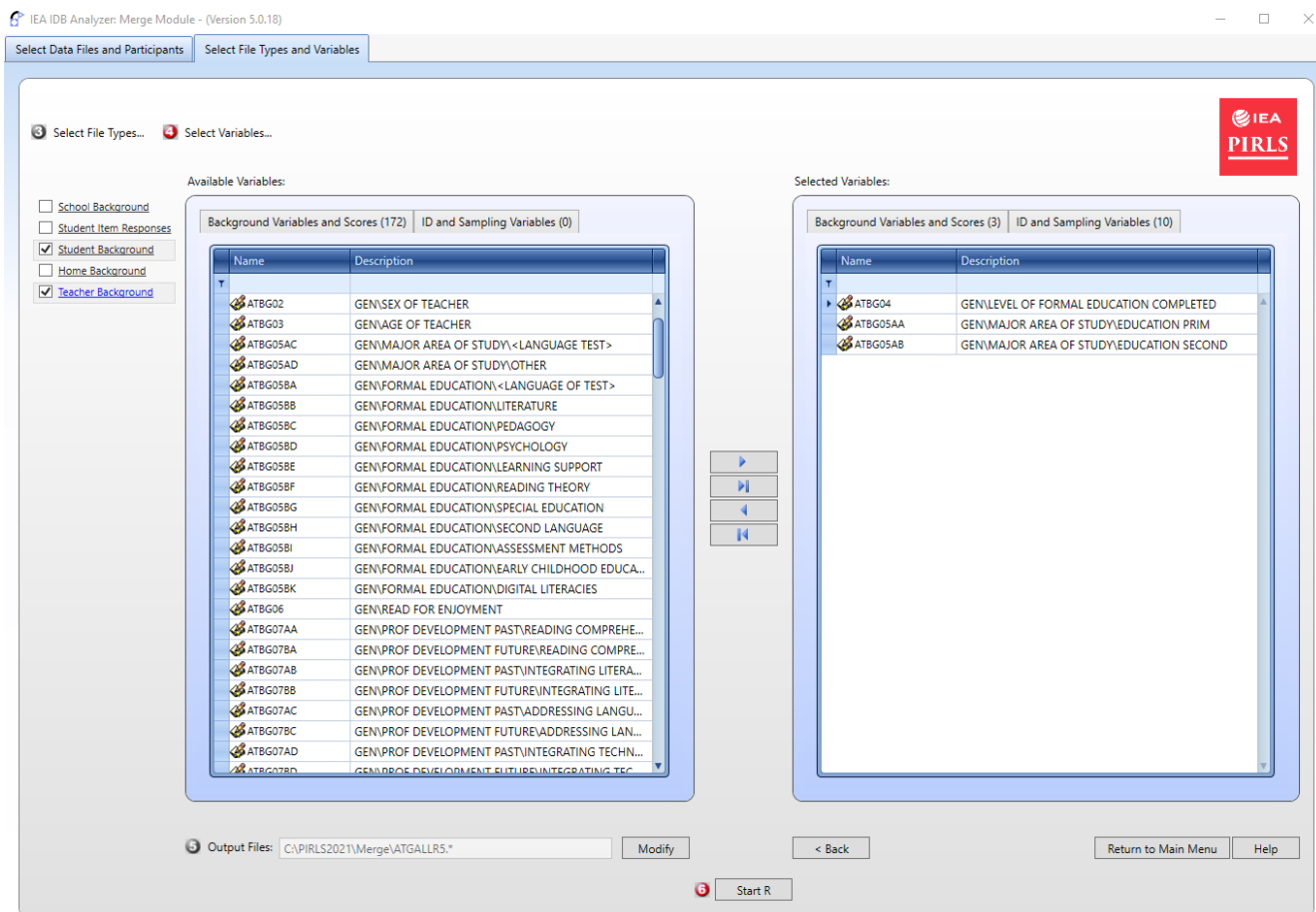
The teachers in the PIRLS 2021 International Database do not constitute representative samples of teachers in the participating countries. Rather, they are the teachers of nationally representative samples of students. Therefore, analyses with teacher data should be made with students as the units of analysis and reported in terms of students who are taught by teachers with a particular attribute. Results of analyses using PIRLS 2021 teacher data should be interpreted with caution because PIRLS 2021 data collection coincided with the COVID-19 pandemic (see [Supplement 2](#) to this User Guide).

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

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



- 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.
- Click the checkbox next to the **Teacher Background** file type, and select the variables of interest from the **Background Variables and Scores** panel on the left in the same manner as in Steps 1 and 2, as shown in Exhibit 1.4.
- Specify the desired name for the merged data file and the folder where it will be stored in the **Output Files** field by clicking the **Define/Modify** button. The IEA IDB Analyzer

also will create an R script (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the merge. In the example shown in Exhibit 1.4, the R script file ATGALLR5.R and the merged data file ATGALLR5.Rdata both will be created and stored in the folder C:\PIRLS2021\Merge. The merged data file will contain all the variables listed in the **Selected Variables** panel on the right.

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

Steps 1 and 2 above are required only if student data (achievement or context) and teacher context data are analyzed simultaneously. It is not recommended to combine both types of files and analyze only student data, as the results may not be correct (see the *Sampling and Weighting Variables* section in Chapter 2 of this User Guide).

Merged Data Files for the User Guide Examples

To conduct the analysis examples presented in this chapter, a number of merged data files were created following the instructions provided above. Because the examples presented in this User Guide are all about PIRLS 2021, merged data files were produced with all countries and benchmarking entities that participated in PIRLS 2021. A full list of countries and their participation in PIRLS 2021 is provided in Chapter 2.

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

ASGALLR5	Merged student context data files with all variables selected for all countries and benchmarking participants
ASHALLR5	Merged home and student context data files with all variables selected for all countries and benchmarking participants
ACGALLR5	Merged school and student context data files with all variables selected for all countries and benchmarking participants
ATGALLR5	Merged teacher and student context data files with all variables selected for all countries and benchmarking participants

Conducting Analyses with the IEA IDB Analyzer

The IEA IDB Analyzer can perform statistical analyses on any files created using the Merge Module. The **Analysis Module** of the IEA IDB Analyzer allows users to specify the type of analysis and select variables from a merged data file as analysis variables. To conduct analyses using plausible values (PVs) for reading achievement, after selecting a **Statistic Type**, users should select the **Use PVs** option from the **Plausible Value Option** drop-down menu.

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

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 bias results. The *PIRLS 2021 International Results in Reading* report annotates results if missing data rates exceed 15% of students. Results are not reported for a country if data are available for fewer than 40% of students, and countries with data available for fewer than 50% are reported at the bottom of the results exhibits. Similarly, achievement estimates can be unreliable if based on small groups of students. Therefore, achievement results are not reported for any context variable categories containing fewer than 2.5% of students.

Statistical Procedures in the IEA IDB Analyzer

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

Percentages and Means

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

Percentages Only

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

Linear Regression

Compute linear regression coefficients for selected independent variables to predict a continuous dependent variable by subgroups defined by grouping variable(s). Plausible

values can be included as dependent or independent variables. This procedure is used in Example 3 of this chapter.

Logistic Regression

Compute logistic regression coefficients for selected independent variables to predict a dichotomous dependent variable by subgroups defined by grouping variable(s). Plausible values can be included as dependent or independent variables. When used as a dependent variable, plausible values will be dichotomized using a specified cutpoint, such as one of the PIRLS 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 PIRLS International Benchmarks, by subgroups defined by grouping variable(s). This procedure is used in Example 4 of this chapter.

Percentiles

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

Definitions of Analysis Variables in the IEA IDB Analyzer

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

Grouping Variables

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

Analysis Variables

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

Plausible Values (PVs)

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

Independent Variables

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

Dependent Variable

This is the variable to be used as the dependent variable when a linear or logistic regression analysis is specified. Only one dependent variable can be listed and can be either a context variable or achievement variables (PVs). To use achievement as the dependent variable, select the **Use PVs** option in the **Plausible Value Option** drop-down menu, click on the **Plausible Values** radio button in the **Dependent Variable** section, and select the achievement scale of interest in the **Plausible Values** field.

Achievement Benchmarks

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

Percentiles

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

Weight Variable

This is the sampling weight variable that will be used in the analysis. The IEA IDB Analyzer automatically selects the appropriate weight variable for analysis based on the file types included in the merged data file. Generally, this will be TOTWGT, but SENWGT and HOUWGT also are available for student-level analyses with student, home, or school context data. For analyzing teacher data, TCHWGT should be used. Chapter 2 of this User Guide provides more information on the PIRLS 2021 sampling weights.

Conducting Analyses with PIRLS Student Achievement and Context Data

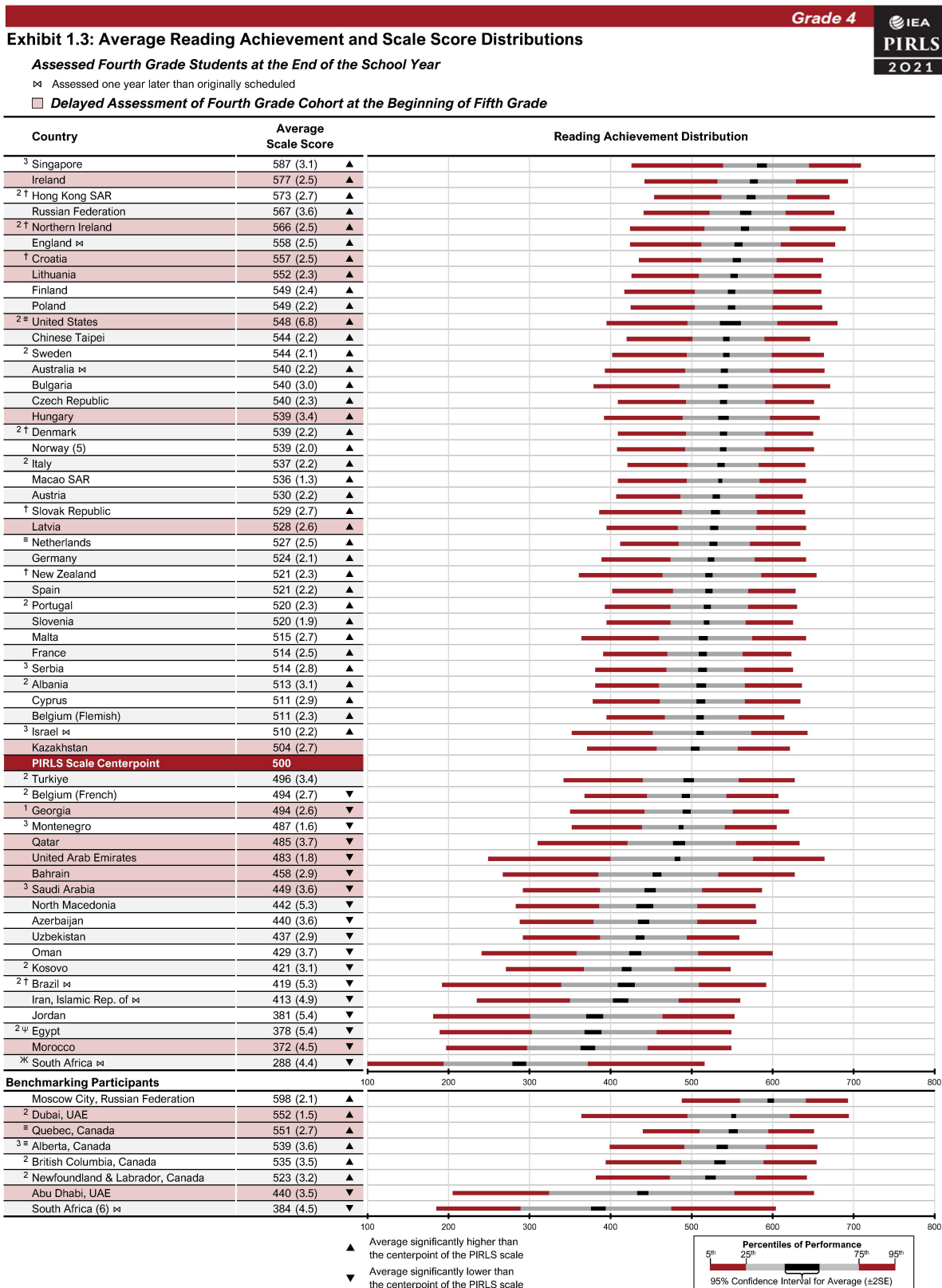
Many types of analyses can be conducted using the student context data files from the PIRLS 2021 International Database. This section presents examples of actual analyses used to produce exhibits in the [*PIRLS 2021 International Results in Reading*](#) report, including percentages and means, linear regression, and International Benchmark analyses.

The examples in this section use the merged PIRLS 2021 student context data file ASGALLR5 described earlier under *Merging Data Files with the IEA IDB Analyzer*, including all countries and benchmarking participants and all available variables. 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 fourth grade students reaching each of the PIRLS 2021 International Benchmarks.

Example 1 – Analysis of Average Reading Achievement

This first example replicates the analysis of the overall distribution of reading achievement, presented in [*Exhibit 1.3*](#) of *PIRLS 2021 International Results in Reading*, repeated below in Exhibit 1.5.

Exhibit 1.5: Exhibit of Example 1 – Analysis of Average Reading Achievement ([Exhibit 1.3](#) of *PIRLS 2021 International Results in Reading*)



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

After creating the merged data file ASGALLR5, the **Analysis Module** of the IEA IDB Analyzer is used to conduct the analysis in the following steps. The completed Analysis Module for this example is shown in Exhibit 1.6.

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

IEA IDB Analyzer: Analysis Module - (Version 5.0.18)

1 Analysis File: C:\PIRLS2021\Merge\ASGALLR5.Rdata [Select]

2 Analysis Type: PIRLS (Using Student Weights) Statistic Type: Percentages and Means Plausible Value Option: Use PVs Number of Decimals: 2 Show Graphs: Yes

3 Select Variables:

Name	Description
ASRLIT01-05	1ST TO 5TH PLAUSIBLE VALUE: LITERARY PURPOSE PV1
ASRINF01-05	1ST TO 5TH PLAUSIBLE VALUE: INFORMATIONAL PURPOSE PV1
ASRIIE01-05	1ST TO 5TH PLAUSIBLE VALUE: INTERPRETING PROCESS PV1
ASRRS01-05	1ST TO 5TH PLAUSIBLE VALUE: STRAIGHTFORWARD PROC PV1

4 Output Files: C:\PIRLS2021\Analysis\REA_Overall.* [Modify] [Start R]

5

Grouping Variables: ☒ Exclude Missing From Analysis

Name	Description
IDCNTRY	Cntry ID

Separate Tables by:

Name	Description

Plausible Values:

Name	Description
ASRREA01-05	1ST TO 5TH PLAUSIBLE VALUE: OVERALL READING PV1

Weight Variable:

Name	Description
TOTWGT	TOTAL STUDENT WEIGHT

[Return to Main Menu] [Help]

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ASGALLR5 as the **Analysis File** by clicking the **Select** button.
3. Select **PIRLS (Using Student Weights)** as the **Analysis Type**.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**.

6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce two graphs in the output file: one graph showing average achievement by country (bar graph in R and SPSS; line graph in SAS), and one bar graph for the weighted percentage of the total students in each country.
8. The IDB Analyzer automatically selects the variable IDCNTY for the **Grouping Variables**. No additional grouping variables are needed for this analysis. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (non-achievement) variables. See the IEA IDB Analyzer Help manual for more information.
10. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASRREA01–05 from the list of available variables in the left panel and move it to the right **Plausible Values** field by clicking the **right arrow** (►) button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis uses student data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define** (or **Modify**) button in the **Output Files** field. The IEA IDB Analyzer also will create an R script (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the analysis. In Exhibit 1.6, the syntax file REA_Overall.R and the output files with the same name will be created and stored in the C:\PIRLS2021\Analysis folder.
13. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** (⌘) button (or select **Submit** in the **Run** menu).

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

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

Exhibit 1.7 displays the results in the R output with six example countries: Bulgaria, Hong Kong SAR, Ireland, Islamic Republic of Iran, Oman, and New Zealand. The results are presented in the “Report” section of the HTML output produced by R.

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

Report											
Analysis for ASRREA by IDCNTRY											
Cntry ID	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASRREA (Mean)	ASRREA (s.e.)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	4043	59248	683.37	3.58	0.10	539.89	3.01	88.00	2.51	0.00	76
Hong Kong, SAR	3830	56443	656.25	3.41	0.09	572.82	2.74	66.52	2.09	0.00	73
Iran, Islamic Republic of	5962	1341065	39450.11	81.12	0.47	412.77	4.88	99.78	3.47	0.00	109
Ireland	4663	68597	1901.17	4.15	0.15	577.33	2.47	76.69	1.36	0.00	75
Oman	5321	64843	1014.79	3.92	0.11	429.48	3.70	109.22	1.89	0.00	110
New Zealand	5557	62946	993.92	3.81	0.11	521.47	2.31	89.32	1.31	0.00	96

Each country’s results are presented on a single line, with countries ordered sequentially according to their numeric ISO code (see Chapter 2). Results for “Table Average” may be produced (not shown), based on all countries included in the data file, including any benchmarking participants. The countries are identified in the first column (Cntry ID) and the second column reports the number of valid cases (N of Cases). The third column reports the sum of weights of the sampled students (Sum of TOTWGT), indicating the estimated total fourth grade population. The fourth column is the standard error of the sum of weights (Sum of TOTWGT (s.e.)). The next two columns report the weighted percentage of students by the grouping variable (Percent), which for this analysis is the percentage of all students in each country out of the total, and its standard error (Percent (s.e.)). The next two columns report the estimated average for the outcome variable, in this case reading achievement (ASRREA (Mean)) and its standard error (ASRREA (s.e.)) (“mnpv” and “mnpv_se” in Excel). The

standard deviation of the achievement scores (Std.Dev.) and its standard error (Std.Dev. (s.e.)) are reported in the next two columns. The last two columns report the percentage of cases with missing data (Percent Missing) and the number of jackknife zones used for computing standard errors (Number of Variance Strata), respectively (see Chapter 13 of [Methods and Procedures: PIRLS 2021 Technical Report](#) for information about jackknifing).

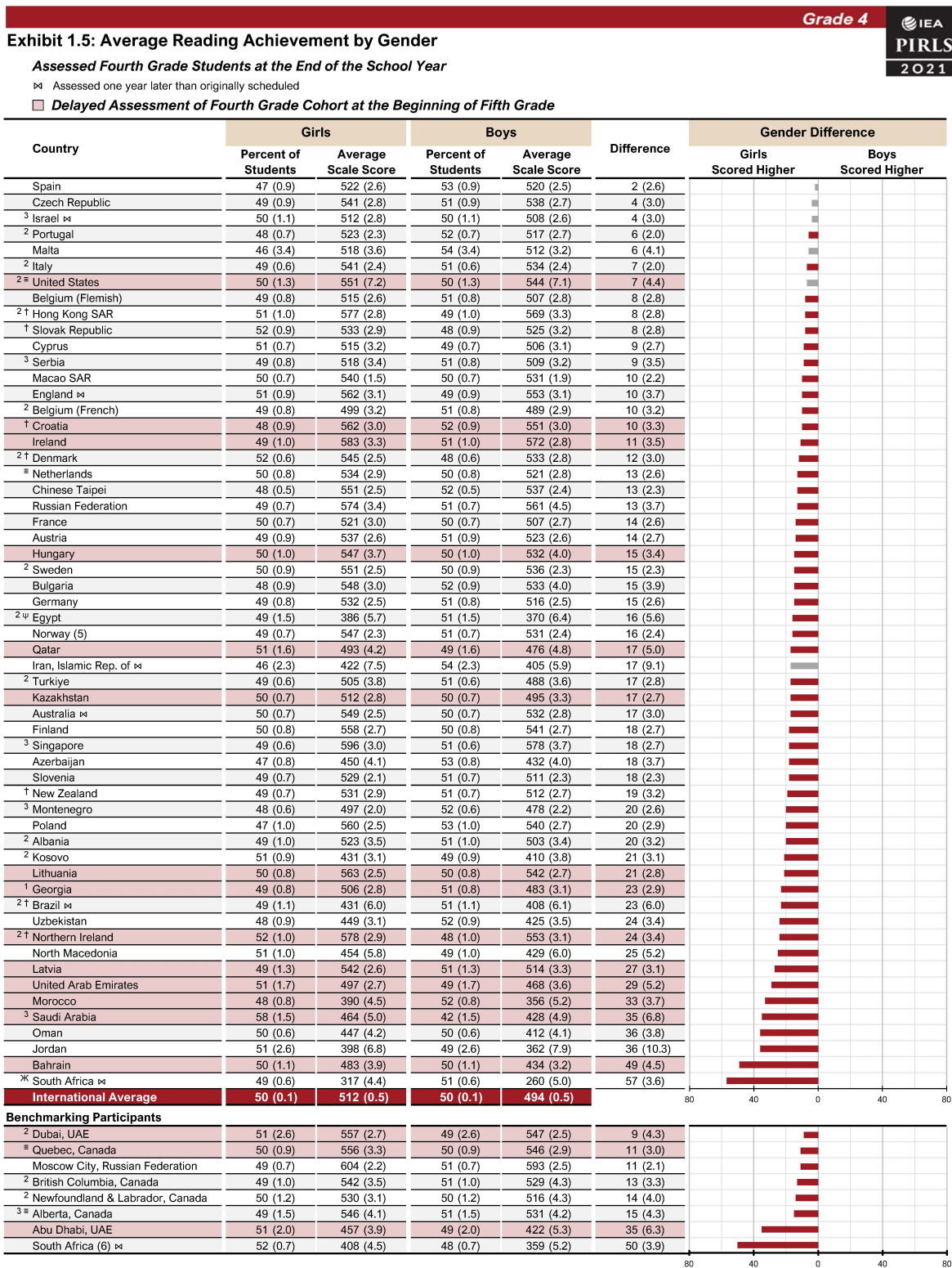
Among the six example countries shown in the output in Exhibit 1.7, Ireland had the highest average achievement. As shown in the fourth line of the table, Ireland had valid data for 4,663 students, and these sampled students represented a population of about 68,597 students, indicated by the sum of the weights. The average reading achievement in Ireland was 577.33 (standard error of 2.47) and its standard deviation was 76.69 (standard error of 1.36). It should be noted that Ireland is one of 14 PIRLS 2021 countries that delayed assessing their fourth grade cohort until the beginning of fifth grade because of COVID-19 disruptions to schooling. Pink highlighting was used to identify these countries throughout the [PIRLS 2021 International Results in Reading](#) report (see Exhibit 1.5) because their students were relatively older compared to other countries and to their samples in previous PIRLS assessments (e.g., PIRLS 2016). Chapter 2 of this User Guide provides a list of all PIRLS 2021 countries and indicates whether data collection occurred on a delayed schedule due to COVID-19.

Example 2 – Analysis of Average Reading Achievement by Gender

The second example using student context data replicates another set of results presented in the [PIRLS 2021 International Results in Reading](#) report. This analysis investigates the relationship between fourth grade students' gender and reading achievement. These results, presented in [Exhibit 1.5](#) of *PIRLS 2021 International Results in Reading*, are repeated below in Exhibit 1.8.

The results of this analysis are based on characteristics of students. In general, before conducting analyses using PIRLS contextual variables, users should refer to the relevant codebook for the data file to identify the appropriate variables and understand the coding scheme. [Supplement 1](#) of this User Guide presents all the context questionnaires administered in PIRLS 2021 and the associated variable names under which the data are saved. [Supplement 2](#) should also be checked for any national adaptations made to the questionnaire items that may impact international comparability.

Exhibit 1.8: Exhibit of Example 2 – Analysis of Average Achievement by Gender (Exhibit 1.5 of PIRLS 2021 International Results in Reading)



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

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

Exhibit 1.9: IEA IDB Analyzer Analysis Module Setup for Example 2 – Analysis of Average Achievement by Gender

IEA IDB Analyzer: Analysis Module - (Version 5.0.18)

1 Analysis File: C:\PIRLS2021\Merge\ASGALLR5.Rdata [Select]

2 Analysis Type: PIRLS (Using Student Weights) | Statistic Type: Percentages and Means | Plausible Value Option: Use PVs | Number of Decimals: 2 | Show Graphs: Yes

3 Select Variables:

Name	Description
ASBG03	GEN\OFTEN SPEAK <LANG OF TEST> AT HOME
ASBG04	GEN\AMOUNT OF BOOKS IN YOUR HOME
ASBG05A	GEN\HOME POSSESS\SHARED COMPUTER OR TABLET
ASBG05B	GEN\HOME POSSESS\STUDY DESK
ASBG05C	GEN\HOME POSSESS\OWN ROOM
ASBG05D	GEN\HOME POSSESS\INTERNET CONNECTION
ASBG05E	GEN\HOME POSSESS\SHARED SMARTPHONE
ASBG05F	GEN\HOME POSSESS\OWN COMPUTER OR TABLET
ASBG05G	GEN\HOME POSSESS\OWN SMARTPHONE
ASBG05H	GEN\HOME POSSESS\<COUNTRY SPECIFIC>
ASBG05I	GEN\HOME POSSESS\<COUNTRY SPECIFIC>
ASBG05J	GEN\HOME POSSESS\<COUNTRY SPECIFIC>
ASBG05K	GEN\HOME POSSESS\<COUNTRY SPECIFIC>
ASBG06	GEN\ABOUT HOW OFTEN ABSENT FROM SCHOOL
ASBG07A	GEN\HOW OFTEN FEEL THIS WAY\TIRED
ASBG07B	GEN\HOW OFTEN FEEL THIS WAY\HUNGRY
ASBG08A	GEN\USE COMPUTER TABLET SCHOOLWORK\FINDING AND...
ASBG08B	GEN\USE COMPUTER TABLET SCHOOLWORK\PREPARING
ASBG09A	GEN\AGREE\USING A COMPUTER OR TABLET
ASBG09B	GEN\AGREE\TYPING

Grouping Variables: ☒ Exclude Missing From Analysis

Name	Description
IDCNTRY	Cntry ID
ITSEX	Sex of Students

Separate Tables by:

Name	Description

Plausible Values:

Name	Description
ASRREA01-05	1ST TO 5TH PLAUSIBLE VALUE: OVERALL READING PV1


Weight Variable:

Name	Description
TOTWGT	TOTAL STUDENT WEIGHT

4 Output Files: C:\PIRLS2021\Analysis\REA_byGender.* [Modify] [Start R] [Return to Main Menu] [Help]

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ASGALLR5 as the **Analysis File** by clicking the **Select** button.

3. Select **PIRLS (Using Student Weights)** as the **Analysis Type**.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce three graphs in the output file: a line graph of the average achievement for each gender by country, a clustered bar graph of average achievement for each gender by country, and a stacked bar graph of average percent of students for each gender by country. R also produces graphs separately for each country.
8. Specify the variable ITSEX as a second grouping variable by first clicking the **Grouping Variables** field to activate it. Then, select ITSEX from the list of variables in the left panel, and move it to the **Grouping Variables** field by clicking the **right arrow** (►) button. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (not achievement) variables. See the Help manual for more information.
10. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASRREA01–05 from the list of available variables in the left panel, and move it to the right **Plausible Values** field by clicking the **right arrow** (►) button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis uses student context data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer also will create an R script (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the analysis. In Exhibit 1.9, the syntax file REA_byGender.R and the output files with the same name will be created and stored in the C:\PIRLS2021\Analysis folder.
13. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open

the **Run** menu and select the **All** menu option. In SAS, click the **Run** () button (or select **Submit** in the **Run** menu).

The IDB Analyzer produces and saves the results output in three file formats within the folder specified in Step 12—an HTML output file (or output in SPSS/SAS), R data file (*.Rdata), and Microsoft Excel Worksheet (*.xlsx). The output files are named using the same name specified for the syntax file in Step 12. The HTML reports produced by R are named with the suffix “_ASRREA” 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 IDCNTY), the IEA IDB Analyzer produces two additional results files in Rdata and xlsx formats. The output file named with the suffix “_sig” reports the significance of the differences between analysis groups—in this case girls and boys—for each country. The output file named with the suffix “_sig2” reports the significance of differences between countries within each of the gender groups.

The results of Example 2 as shown in the R output file are presented in Exhibit 1.10 with example countries Bulgaria, Hong Kong SAR, Ireland, Islamic Republic of Iran, New Zealand, and Oman. The results are presented in the “Report” section of the R output.

Countries are ordered sequentially according to their numeric ISO code (see Chapter 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 (Sex of Students) indicates the category of the grouping variable ITSEX being reported according to the value labels (1: Girl, 2: Boy). The third column reports the number of valid cases (N of Cases), the fourth column reports the sum of weights of the sampled students (Sum of TOTWGT), indicating the estimated total students in the population represented by the sample, and the fifth column is the standard error of the sum of weights (Sum of TOTWGT (s.e.)).

The next two columns report the weighted percentage of students in the particular category of the second grouping variable (Percent), which for this analysis is the percent of students in each category of ITSEX within IDCNTY, 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 reading achievement (ASRREA (Mean)) and its standard error (ASRREA (s.e.)). 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.

Exhibit 1.10: R Output for Example 2 – Analysis of Average Achievement by Gender

Report												
Analysis for ASRREA by IDCNTY ITSEX												
Cntry ID	Sex of Students	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASRREA (Mean)	ASRREA (s.e.)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	Girl	1976	28301	621.20	47.77	0.95	547.82	3.01	86.66	2.36	0.00	76
Bulgaria	Boy	2067	30947	691.63	52.23	0.95	532.64	3.99	88.58	3.34	0.00	76
Hong Kong, SAR	Girl	1955	28623	675.64	50.71	0.97	576.90	2.79	64.04	2.39	0.00	72
Hong Kong, SAR	Boy	1875	27820	604.15	49.29	0.97	568.62	3.34	68.73	2.35	0.00	70
Iran, Islamic Republic of	Girl	2857	621609	27708.88	46.35	2.28	421.81	7.45	98.56	4.59	0.00	61
Iran, Islamic Republic of	Boy	3105	719456	44214.38	53.65	2.28	404.97	5.93	100.15	3.97	0.00	65
Ireland	Girl	2303	33545	1154.95	48.90	1.01	582.99	3.27	77.87	1.99	0.00	71
Ireland	Boy	2360	35052	1199.75	51.10	1.01	571.91	2.81	75.13	1.48	0.00	70
Oman	Girl	2687	32662	672.69	50.37	0.61	447.12	4.20	101.81	2.28	0.00	110
Oman	Boy	2634	32181	615.83	49.63	0.61	411.58	4.08	113.49	2.50	0.00	110
New Zealand	Girl	2731	30717	649.21	48.81	0.72	531.17	2.86	87.10	1.67	0.00	94
New Zealand	Boy	2825	32215	698.13	51.19	0.72	512.22	2.74	90.44	1.65	0.00	96

The results for the Islamic Republic of Iran are interpreted here as an example. In Exhibit 1.8, Iran is one of only a few countries that did not show a statistically significant difference in reading achievement between girls and boys. It should be noted that Iran is annotated with a bowtie (⌘) in Exhibit 1.8, indicating that the country had to delay data collection by one year due to COVID-19 disruptions. Iran assessed the next fourth grade cohort at the end of the school year in 2022. Chapter 2 of this User Guide provides a list of all PIRLS 2021 countries and indicates whether data collection occurred on a delayed schedule due to COVID-19.

From the two lines of results for the Islamic Republic of Iran in Exhibit 1.10, Iran's sample had somewhat more boys than girls: 46.35% of students were girls (standard error of 2.28) and 53.65% were boys (standard error of 2.28). The average reading achievement of girls was 421.81 (standard error of 7.45) and it was 404.97 for boys (standard error of 5.93). Although a 17-point difference would appear to be noteworthy, it is not statistically significant because of the larger standard errors derived from Iran's national sample.

The statistical significance of the gender differences, reported in the "Gender Difference" column of Exhibit 1.8, can be determined by examining the output file named with the suffix "_sig" (REA_byGender_ASRREA_by_ITSEX_sig in this example) provided in R data (*.Rdata)

and Excel (*.xlsx) file formats. This example refers to the xlsx version, which is the same for all software, shown in Exhibit 1.11 with Bulgaria, Hong Kong SAR, and the Islamic Republic of Iran.

Exhibit 1.11: 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
	IDCOUNTRY	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
1	Bulgaria	ASRREA	ITSEX	Girl	Girl	47.77	0.95	47.77	0.95	0.00	0.00	#NUM!	547.82	3.01	547.82	3.01	0.00	0.00	#NUM!
2	Bulgaria	ASRREA	ITSEX	Girl	Boy	47.77	0.95	52.23	0.95	4.47	1.89	2.36	547.82	3.01	532.64	3.99	-15.18	3.86	-3.93
3	Bulgaria	ASRREA	ITSEX	Boy	Girl	52.23	0.95	47.77	0.95	-4.47	1.89	-2.36	532.64	3.99	547.82	3.01	15.18	3.86	3.93
4	Bulgaria	ASRREA	ITSEX	Boy	Boy	52.23	0.95	52.23	0.95	0.00	0.00	#NUM!	532.64	3.99	532.64	3.99	0.00	0.00	#NUM!
5	Hong Kong, SAR	ASRREA	ITSEX	Girl	Girl	50.71	0.97	50.71	0.97	0.00	0.00	#NUM!	576.90	2.79	576.90	2.79	0.00	0.00	#NUM!
6	Hong Kong, SAR	ASRREA	ITSEX	Girl	Boy	50.71	0.97	49.29	0.97	-1.42	1.95	-0.73	576.90	2.79	568.62	3.34	-8.28	2.77	-2.99
7	Hong Kong, SAR	ASRREA	ITSEX	Boy	Girl	49.29	0.97	50.71	0.97	1.42	1.95	0.73	568.62	3.34	576.90	2.79	8.28	2.77	2.99
8	Hong Kong, SAR	ASRREA	ITSEX	Boy	Boy	49.29	0.97	49.29	0.97	0.00	0.00	#NUM!	568.62	3.34	568.62	3.34	0.00	0.00	#NUM!
9	Iran, Islamic Rep	ASRREA	ITSEX	Girl	Girl	46.35	2.28	46.35	2.28	0.00	0.00	#NUM!	421.81	7.45	421.81	7.45	0.00	0.00	#NUM!
10	Iran, Islamic Rep	ASRREA	ITSEX	Girl	Boy	46.35	2.28	53.65	2.28	7.30	4.55	1.60	421.81	7.45	404.97	5.93	-16.85	9.11	-1.85
11	Iran, Islamic Rep	ASRREA	ITSEX	Boy	Girl	53.65	2.28	46.35	2.28	-7.30	4.55	-1.60	404.97	5.93	421.81	7.45	16.85	9.11	1.85
12	Iran, Islamic Rep	ASRREA	ITSEX	Boy	Boy	53.65	2.28	53.65	2.28	0.00	0.00	#NUM!	404.97	5.93	404.97	5.93	0.00	0.00	#NUM!

PIRLS 2021 International Results in Reading labels gender differences as statistically significant based on two-tailed null hypothesis significance tests. For each country, the “sig” output reports the average achievement difference between the reference group (column D) and the comparison group (column E) in column Q, labeled “mnpvdiff.” Dividing this value by its standard error (“mnpvdiff_se” in column R) gives a *t*-statistic (“mnpvdiff_t” in column S) for evaluating whether the estimated difference is significantly different from zero. For an error level (α) of 5%, values greater than +1.96 (the upper critical value) or less than -1.96 (the lower critical value) indicate that the difference between the reference group (girls) average and the comparison group (boys) average is significantly different from zero. Values between -1.96 and +1.96 (the lower and upper critical values for $\alpha = 0.05$) 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 Iran is -1.85, 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 fourth grade reading achievement between girls and boys and determine if it is statistically significant. This example also demonstrates the **Dummy Coding** feature of the IEA IDB Analyzer. Like Example 2, the results of this example

are presented in [Exhibit 1.5](#) of *PIRLS 2021 International Results in Reading* and are shown above in Exhibit 1.8 in the column labeled “Gender Difference.”

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

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

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

The screenshot shows the IEA IDB Analyzer Analysis Module (Version 5.0.18) interface. The Analysis File is set to C:\PIRLS2021\Merge\ASGALLR5.Rdata. The Analysis Type is PIRLS (Using Student Weights), the Statistic Type is Linear Regression, the Plausible Value Option is Use PVs, and the Number of Decimals is 2. The Select Variables section shows a list of variables including ASRLIT01-05, ASRINF01-05, ASRIIE01-05, and ASRRS01-05. The Grouping Variables section is set to IDCNTY. The Independent Variables section includes ITSEX (Sex of Students) with Dummy Coding and 2 categories. The Dependent Variable section is set to ASRREA01-05 (1ST TO 5TH PLAUSIBLE VALUE: OVERALL READING PV1). The Weight Variable section is set to TOTWTG (TOTAL STUDENT WEIGHT). The Output File is C:\PIRLS2021\Analysis\Rea_byGender_LR.*. The Start R button is highlighted.

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ASGALLR5 as the **Analysis File** by clicking the **Select** button.
3. Select **PIRLS (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 ASRREA01–05 from the list of available variables in the left panel, and move it to the right **Plausible Values** field by clicking the **right arrow (►)** button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis uses student context data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer also will create a syntax file of the same name and in the same folder, with the code necessary to perform the analysis. In Exhibit 1.12, the syntax file REA_byGender_LR.R and the output files with the same name will be created and stored in the C:\PIRLS2021\Analysis folder.
13. Click the **Start R** button to create the R script and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard.

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

Exhibit 1.13 displays the main results for this example analysis—the regression coefficients—in the R output file for example countries Bulgaria, Hong Kong SAR, Ireland, Islamic Republic of Iran, New Zealand, and Oman. Countries are ordered numerically according to their numeric ISO code (see Chapter 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.

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

Regression Coefficients								
EqVar	Cntry ID	Variable	Regression Coefficient	Regression Coefficient (s.e.)	Regression Coefficient (t-value)	Stndrdzd. Coefficient	Stndrdzd. Coefficient (s.e.)	Stndrdzd. Coefficient (t-value)
(CONSTANT)	Bulgaria	(CONSTANT)	547.82	3.01	182.10	NaN	NaN	NaN
ITSEX_D2	Bulgaria	ITSEX_D2	-15.18	3.86	-3.93	-0.09	0.02	-4.08
(CONSTANT)	Hong Kong, SAR	(CONSTANT)	576.90	2.79	206.51	NaN	NaN	NaN
ITSEX_D2	Hong Kong, SAR	ITSEX_D2	-8.28	2.77	-2.99	-0.06	0.02	-3.05
(CONSTANT)	Iran, Islamic Republic of	(CONSTANT)	421.81	7.45	56.60	NaN	NaN	NaN
ITSEX_D2	Iran, Islamic Republic of	ITSEX_D2	-16.85	9.11	-1.85	-0.08	0.05	-1.86
(CONSTANT)	Ireland	(CONSTANT)	582.99	3.27	178.46	NaN	NaN	NaN
ITSEX_D2	Ireland	ITSEX_D2	-11.07	3.52	-3.15	-0.07	0.02	-3.16
(CONSTANT)	Oman	(CONSTANT)	447.12	4.20	106.45	NaN	NaN	NaN
ITSEX_D2	Oman	ITSEX_D2	-35.55	3.84	-9.27	-0.16	0.02	-9.47
(CONSTANT)	New Zealand	(CONSTANT)	531.17	2.86	185.55	NaN	NaN	NaN
ITSEX_D2	New Zealand	ITSEX_D2	-18.95	3.18	-5.95	-0.11	0.02	-5.98

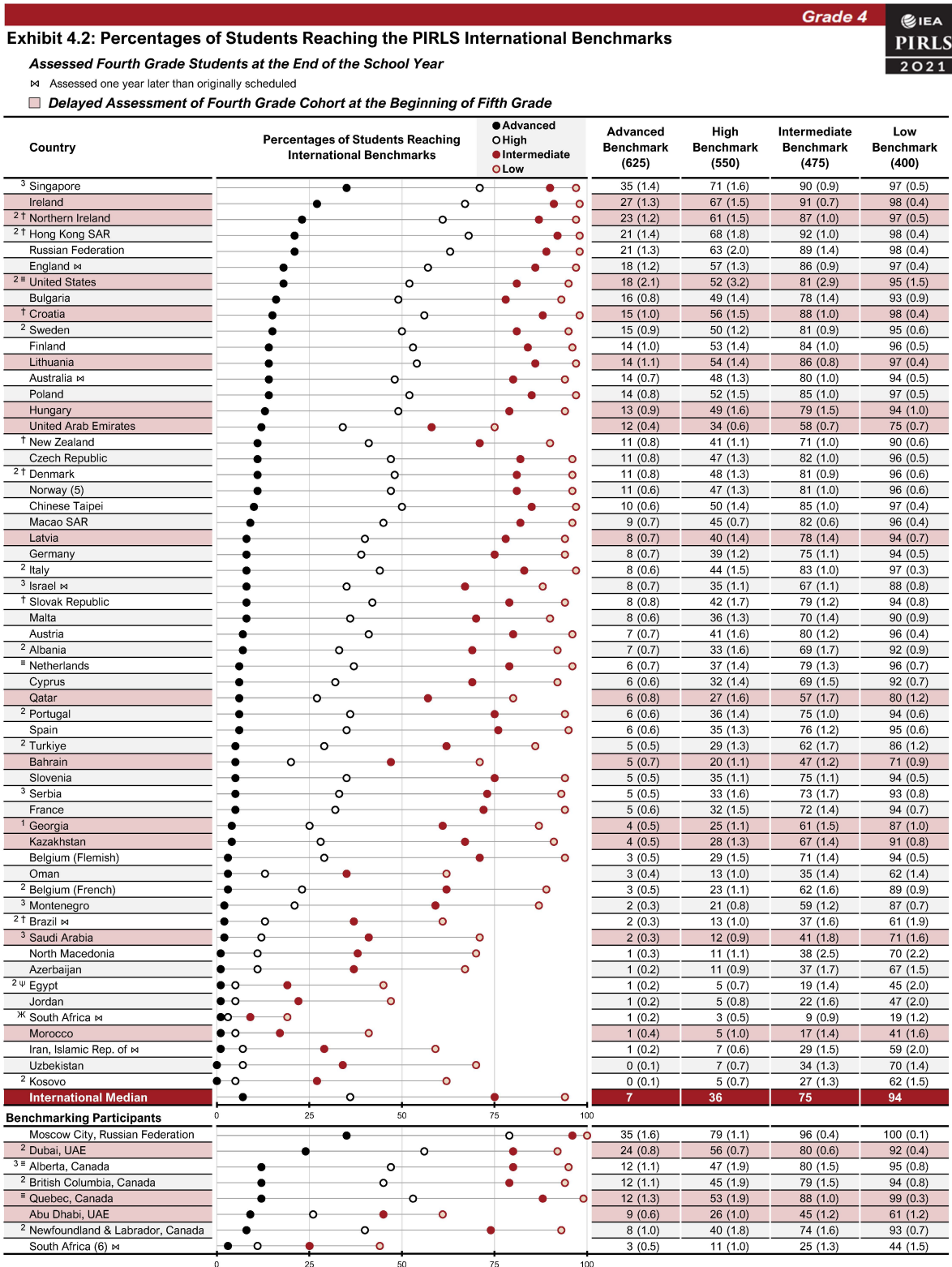
The countries are identified in the second column (Cntry ID) and the third column (Variable) indicates the intercept (CONSTANT) or the regression coefficient being reported. The fourth column reports the “Regression Coefficient” (“b” in Excel), indicating, for the intercept, the average value of the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

In Exhibit 1.13, the first line of results for the Islamic Republic of Iran, labeled “(CONSTANT)” (“Intercept” in SAS), indicates the estimated average reading achievement of fourth grade girls in Iran: 421.81 with a standard error of 7.45. This estimate concurs with the results obtained in the previous example (Exhibit 1.10). From the second line of results labeled “ITSEX_D2,” the fourth grade boys in Iran had a negative average reading achievement difference from girls of –16.85 with an estimated standard error of 9.11. The *t*-value for the coefficient is –1.85, 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 (421.81 – 16.85) yields the estimated average reading achievement of fourth grade boys in Iran, which was 404.97 in Exhibit 1.10.

Example 4 – Analysis of PIRLS International Benchmarks

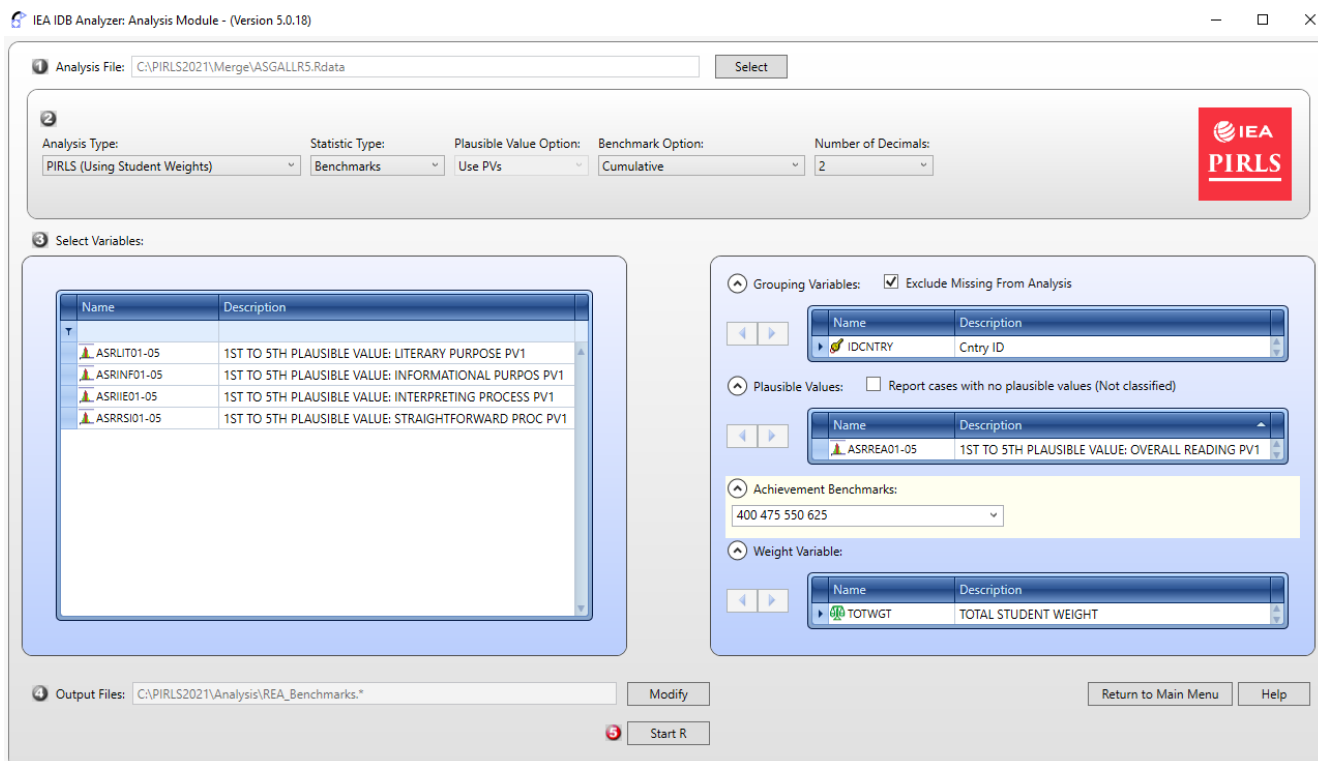
This section describes how to use the IEA IDB Analyzer to perform analyses of student achievement in relation to the PIRLS International Benchmarks of Reading Achievement. This example computes the percentages of students reaching each of the four PIRLS 2021 International Benchmarks of fourth grade reading achievement (Advanced, High, Intermediate, and Low) using the merged ASGALLR5 data file described earlier under *Merging Data Files with the IEA IDB Analyzer*. These results, presented in [Exhibit 4.2](#) of the PIRLS 2021 International Results report, are shown below in Exhibit 1.14.

Exhibit 1.14: Exhibit of Example 4 – Analysis of PIRLS International Benchmarks (Exhibit 4.2 of PIRLS 2021 International Results in Reading)



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

Exhibit 1.15: IEA IDB Analyzer Analysis Module Setup for Example 4 – Analysis of PIRLS International Benchmarks



IEA IDB Analyzer: Analysis Module - (Version 5.0.18)

1 Analysis File: C:\PIRLS2021\Merge\ASGALLR5.Rdata [Select]

2 Analysis Type: PIRLS (Using Student Weights) Statistic Type: Benchmarks Plausible Value Option: Use PVs Benchmark Option: Cumulative Number of Decimals: 2

3 Select Variables:

Name	Description
ASRLIT01-05	1ST TO 5TH PLAUSIBLE VALUE: LITERARY PURPOSE PV1
ASRINF01-05	1ST TO 5TH PLAUSIBLE VALUE: INFORMATIONAL PURPOSE PV1
ASRIIE01-05	1ST TO 5TH PLAUSIBLE VALUE: INTERPRETING PROCESS PV1
ASRRSI01-05	1ST TO 5TH PLAUSIBLE VALUE: STRAIGHTFORWARD PROC PV1

Grouping Variables: ☒ Exclude Missing From Analysis

Name	Description
IDCNTRY	Cntry ID

Plausible Values: ☐ Report cases with no plausible values (Not classified)

Name	Description
ASRREA01-05	1ST TO 5TH PLAUSIBLE VALUE: OVERALL READING PV1

Achievement Benchmarks: 400 475 550 625

Weight Variable:

Name	Description
TOTWGT	TOTAL STUDENT WEIGHT

4 Output Files: C:\PIRLS2021\Analysis\REA_Benchmarks.* [Modify] [Start R] [Return to Main Menu] [Help]

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Specify the data file ASGALLR5 as the **Analysis File** by clicking the **Select** button.
3. Select **PIRLS (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 PIRLS 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 ASRREA01–05 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 PIRLS International Benchmarks in ascending order—400, 475, 550, and 625 (Low, Intermediate, High, and Advanced, respectively). These values can be entered manually with each separated by a blank space, or they can be selected by clicking on the drop-down menu available for this field.
10. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis uses student data.
11. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer will create a syntax file with the code necessary to perform the analysis. In Exhibit 1.15, the syntax file REA_Benchmarks.R and the output files with the same name will be created and stored in the folder C:\PIRLS2021\Analysis.
12. Click the **Start R** button (or Start SPSS/SAS) to create the syntax file and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** (⌘) button (or select **Submit** in the **Run** menu).

The IDB Analyzer produces and saves the results output in three file formats within the folder specified in Step 11—an HTML output file (or output in SPSS/SAS), R data file (*.Rdata), and Microsoft Excel Worksheet (*.xlsx). Graphs are included only in the HTML (or SPSS/SAS) output files. Exhibit 1.16 presents the results of Example 4 as shown in the R output, under the “Report” section. Results are shown for the first four example countries: Bulgaria, Hong Kong SAR, Islamic Republic of Iran, and Ireland.

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

Exhibit 1.16: R Output for Example 4 – Analysis of PIRLS International Benchmarks

Report						
Percentage by Performance Group of ASRREA						
Cntry ID	cutvar	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)
Bulgaria	1. At or Above 400	3842	55029	722.76	92.88	0.94
Bulgaria	2. At or Above 475	3363	46435	917.49	78.37	1.39
Bulgaria	3. At or Above 550	2214	29068	879.16	49.06	1.42
Bulgaria	4. At or Above 625	769	9411	449.47	15.88	0.76
Hong Kong, SAR	1. At or Above 400	3770	55472	727.37	98.28	0.41
Hong Kong, SAR	2. At or Above 475	3549	52093	948.19	92.29	0.99
Hong Kong, SAR	3. At or Above 550	2651	38501	1257.98	68.21	1.84
Hong Kong, SAR	4. At or Above 625	823	11770	834.23	20.85	1.40
Iran, Islamic Republic of	1. At or Above 400	3725	785837	38742.41	58.60	2.01
Iran, Islamic Republic of	2. At or Above 475	1873	385683	26490.50	28.76	1.51
Iran, Islamic Republic of	3. At or Above 550	450	88199	9346.44	6.58	0.62
Iran, Islamic Republic of	4. At or Above 625	36	7589	2061.81	0.57	0.15
Ireland	1. At or Above 400	4562	67076	1844.83	97.78	0.38
Ireland	2. At or Above 475	4242	62345	1728.05	90.89	0.71
Ireland	3. At or Above 550	3160	46097	1507.95	67.20	1.45
Ireland	4. At or Above 625	1284	18480	999.98	26.94	1.34

As shown in the four lines of results for Bulgaria, 92.88% of the fourth grade students in Bulgaria performed at or above the Low International Benchmark of 400 scale score points, with a standard error of 0.94; 78.37% of students reached the Intermediate International Benchmark, with a standard error of 1.39; 49.06% of students reached the High International Benchmark, with a standard error of 1.42; and 15.88% of students reached the Advanced International Benchmark, with a standard error of 0.76.

Ireland had the highest percentage of students reaching the Advanced International Benchmark, with 26.94% of students at or above this benchmark (standard error of 1.34). Again, it should be noted that Ireland had to delay assessing their fourth grade cohort until the beginning of fifth grade because of COVID-19 disruptions to schooling. Their students were notably older, so they are highlighted in pink throughout the *PIRLS 2021 International Results in Reading* report (see Exhibit 1.14). In comparison, only 0.57% of fourth grade students in the Islamic Republic of Iran reached this advanced level (standard error of 0.15). Iran delayed assessing their fourth grade cohort by one year, and is annotated with a bowtie (⋈).

Conducting Analyses with PIRLS Home Context Data

This section presents an analysis conducted using the IEA IDB Analyzer with home context data from the PIRLS 2021 International Database. Home context data were collected from the parents of fourth grade students with the PIRLS 2021 Home Questionnaire, or “Early Learning Survey.” Like the previous section, the example below is an actual analysis used to produce exhibits in the *PIRLS 2021 International Results in Reading* report.

In general, before conducting analyses using PIRLS contextual variables such as those in the home context data files, users should refer to the relevant codebook for the data file to identify the appropriate variables and understand the coding scheme. [Supplement 1](#) of this User Guide presents all the context questionnaires administered in PIRLS 2021 and the associated variable names under which the data are saved. [Supplement 2](#) should also be checked for any national adaptations made to the questionnaire items that may impact international comparability.

Analyzing home context data from the PIRLS 2021 International Database requires that the home context data files (named beginning with “ASH”) be merged with either student achievement files (“ASA”) or student context files (“ASG”) to retrieve the achievement scores and required sample design variables. This example uses home context data merged with student context data described earlier in the chapter under *Merging Data Files with the IEA IDB Analyzer* (ASHALLR5). Example 5 computes the average score on a context questionnaire scale, along with the percentages of students—with their average achievement—for each of the categories of the scale’s corresponding index. The analysis replicates [Exhibit 5.3](#) of *PIRLS 2021 International Results in Reading*, repeated below in Exhibit 1.17.

Exhibit 1.17: Exhibit of Example 5 – Analysis of a Context Questionnaire Scale with Home Context Data ([Exhibit 5.3](#) of PIRLS 2021 International Results in Reading)

Grade 4							IEA PIRLS 2021
Exhibit 5.3: Parents Like Reading							
<i>Students' Results based on Parents' Reports</i>							
Assessed Fourth Grade Students at the End of the School Year							
☒ Assessed one year later than originally scheduled							
☐ Delayed Assessment of Fourth Grade Cohort at the Beginning of Fifth Grade							
Country	Very Much Like Reading		Somewhat Like Reading		Do Not Like Reading		Average Scale Score
	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement	
Serbia	51 (1.2)	532 (3.0)	39 (1.1)	497 (3.6)	10 (1.0)	492 (6.5)	10.4 (0.05)
Montenegro	51 (0.8)	505 (2.1)	43 (0.7)	472 (2.3)	6 (0.4)	455 (5.4)	10.5 (0.03)
Azerbaijan	46 (1.1)	451 (4.2)	46 (1.1)	437 (4.5)	8 (0.6)	409 (7.6)	10.4 (0.04)
Georgia	44 (1.0)	511 (3.0)	50 (1.0)	486 (3.0)	6 (0.6)	471 (12.9)	10.2 (0.04)
Kosovo	42 (1.3)	441 (3.7)	53 (1.2)	410 (3.6)	5 (0.6)	385 (7.6)	10.3 (0.04)
Ireland	42 (1.2)	599 (2.7)	42 (1.1)	574 (2.8)	16 (0.7)	555 (3.8)	10.0 (0.05)
Italy	41 (1.0)	553 (2.6)	45 (0.8)	531 (2.4)	13 (0.6)	520 (3.2)	10.1 (0.04)
North Macedonia	41 (1.2)	467 (4.7)	49 (1.1)	432 (5.5)	10 (1.0)	406 (10.2)	10.1 (0.08)
Bulgaria	41 (1.3)	570 (3.2)	41 (1.2)	537 (4.0)	18 (1.4)	481 (7.0)	9.8 (0.07)
Spain	40 (0.9)	540 (2.7)	44 (0.8)	517 (2.4)	16 (0.6)	498 (3.3)	10.0 (0.03)
Uzbekistan	39 (1.4)	449 (3.1)	57 (1.3)	431 (3.2)	4 (0.3)	404 (8.8)	10.3 (0.04)
Northern Ireland s	39 (1.1)	595 (3.9)	42 (1.2)	569 (4.0)	19 (0.9)	556 (5.0)	9.8 (0.05)
Denmark	39 (1.0)	557 (2.6)	42 (0.9)	539 (2.6)	19 (0.7)	514 (3.4)	9.8 (0.04)
Finland	38 (0.8)	573 (2.6)	44 (0.8)	547 (2.6)	18 (0.8)	522 (3.2)	9.9 (0.04)
Malta r	37 (1.0)	539 (3.7)	47 (1.0)	515 (2.7)	15 (0.9)	514 (4.8)	9.8 (0.04)
Cyprus	37 (0.8)	535 (3.5)	49 (0.9)	505 (2.9)	14 (0.6)	491 (4.2)	9.9 (0.03)
Poland	36 (1.1)	566 (3.1)	47 (1.0)	546 (2.2)	17 (0.9)	529 (4.8)	9.8 (0.05)
Austria	36 (1.1)	559 (2.5)	44 (1.1)	524 (2.6)	20 (0.8)	504 (3.4)	9.7 (0.05)
Albania	35 (1.4)	535 (3.6)	56 (1.3)	504 (3.5)	9 (1.2)	493 (7.6)	10.0 (0.05)
Sweden s	34 (1.2)	575 (3.7)	50 (1.1)	551 (3.6)	16 (0.8)	541 (4.0)	9.7 (0.04)
Germany s	33 (1.0)	562 (3.4)	47 (1.0)	540 (3.4)	20 (0.9)	504 (4.6)	9.6 (0.04)
Czech Republic	33 (0.9)	564 (2.9)	45 (0.8)	543 (2.8)	22 (0.7)	522 (3.1)	9.6 (0.04)
Norway (5)	32 (1.0)	560 (2.4)	48 (0.9)	540 (2.5)	19 (0.8)	520 (2.7)	9.6 (0.04)
Slovak Republic	32 (1.0)	554 (3.1)	47 (1.3)	528 (3.1)	21 (1.6)	504 (6.6)	9.5 (0.09)
Israel ☒ s	32 (1.0)	542 (3.6)	51 (1.1)	503 (3.4)	17 (0.8)	497 (4.9)	9.7 (0.04)
Türkiye	31 (1.2)	531 (3.2)	48 (1.4)	497 (3.4)	21 (1.8)	453 (6.4)	9.5 (0.09)
Hungary r	30 (0.9)	569 (3.6)	50 (1.0)	541 (4.3)	20 (1.0)	509 (6.0)	9.5 (0.04)
Portugal	30 (0.9)	543 (2.2)	52 (0.8)	516 (2.4)	18 (0.6)	501 (3.5)	9.6 (0.03)
Belgium (French) r	28 (0.9)	529 (3.6)	48 (1.1)	496 (3.2)	24 (1.0)	470 (3.4)	9.3 (0.04)
Belgium (Flemish)	27 (1.0)	535 (2.6)	46 (0.9)	513 (2.8)	27 (0.9)	498 (2.9)	9.2 (0.05)
Bahrain	27 (0.8)	488 (4.3)	60 (0.9)	458 (3.4)	14 (0.5)	427 (4.7)	9.6 (0.02)
Croatia	26 (1.3)	575 (3.3)	56 (1.0)	557 (2.8)	18 (1.2)	535 (4.0)	9.5 (0.05)
Slovenia	26 (0.8)	544 (2.8)	57 (0.7)	520 (1.9)	17 (0.7)	501 (3.2)	9.4 (0.03)
Latvia	24 (0.9)	553 (3.4)	54 (0.9)	529 (3.0)	22 (0.9)	507 (4.9)	9.3 (0.04)
South Africa ☒ r	24 (0.8)	342 (6.8)	62 (0.8)	283 (4.8)	15 (0.5)	270 (9.4)	9.6 (0.03)
Iran, Islamic Rep. of ☒	24 (1.0)	449 (4.4)	62 (0.9)	405 (5.5)	14 (0.8)	385 (7.0)	9.5 (0.04)
Saudi Arabia r	23 (0.8)	469 (4.6)	62 (1.0)	451 (3.8)	15 (0.8)	448 (5.7)	9.5 (0.04)
Qatar r	22 (0.9)	512 (5.6)	63 (1.2)	491 (4.7)	15 (0.9)	485 (6.2)	9.4 (0.03)
France	22 (0.8)	546 (3.1)	57 (0.9)	514 (2.8)	21 (0.7)	500 (3.4)	9.3 (0.03)
Oman	22 (0.9)	451 (6.0)	68 (1.0)	428 (3.8)	10 (0.6)	407 (8.7)	9.5 (0.03)
United Arab Emirates s	21 (0.5)	534 (4.0)	65 (0.6)	491 (2.5)	13 (0.4)	488 (3.7)	9.5 (0.01)
Russian Federation	21 (0.9)	587 (2.8)	58 (1.1)	566 (4.3)	21 (0.9)	550 (4.6)	9.2 (0.04)
Singapore	21 (0.6)	622 (3.4)	57 (0.7)	589 (3.2)	23 (0.6)	571 (3.8)	9.2 (0.03)
Morocco	20 (1.1)	410 (5.0)	53 (1.7)	377 (5.6)	27 (2.1)	338 (6.5)	9.1 (0.07)
Brazil ☒	19 (0.9)	456 (7.5)	51 (0.8)	423 (5.8)	29 (1.1)	401 (8.2)	9.0 (0.04)
Macao SAR	17 (0.6)	551 (2.5)	61 (0.7)	535 (1.6)	22 (0.6)	526 (2.2)	9.1 (0.02)
Kazakhstan	17 (0.7)	527 (3.8)	70 (0.7)	501 (2.7)	13 (0.6)	493 (4.6)	9.4 (0.03)
Chinese Taipei	15 (0.6)	564 (3.3)	62 (0.8)	545 (2.2)	23 (0.8)	528 (3.3)	9.1 (0.03)
Jordan	15 (0.9)	416 (6.4)	62 (1.2)	382 (5.8)	23 (1.2)	358 (8.2)	9.1 (0.04)
Hong Kong SAR	14 (0.8)	592 (3.7)	63 (0.9)	573 (2.7)	24 (0.7)	567 (3.8)	9.0 (0.03)
Egypt	14 (0.7)	416 (7.4)	60 (1.4)	379 (5.7)	27 (1.5)	367 (7.7)	8.9 (0.04)
International Average	30 (0.1)	524 (0.5)	52 (0.1)	497 (0.5)	17 (0.1)	478 (0.8)	
New Zealand x	44 (1.4)	563 (3.4)	40 (1.2)	532 (4.5)	17 (0.8)	508 (5.7)	10.0 (0.06)
Netherlands x	39 (1.2)	557 (3.3)	41 (1.3)	541 (3.6)	21 (1.2)	516 (4.5)	9.6 (0.06)
Lithuania y	--	--	--	--	--	--	--
Australia ☒	--	--	--	--	--	--	--
England ☒	--	--	--	--	--	--	--
United States	--	--	--	--	--	--	--
Benchmarking Participants							
Alberta, Canada s	40 (2.0)	573 (4.1)	44 (1.7)	551 (3.8)	15 (1.1)	540 (7.3)	10.0 (0.08)
British Columbia, Canada s	38 (1.7)	572 (3.8)	46 (1.2)	549 (4.7)	16 (1.3)	534 (5.4)	9.9 (0.07)
Newfoundland & Labrador, Canada s	38 (1.4)	549 (3.7)	45 (1.4)	529 (4.7)	18 (1.0)	515 (6.5)	9.9 (0.06)
Moscow City, Russian Federation	32 (0.8)	612 (2.1)	54 (0.7)	595 (2.3)	14 (0.7)	579 (3.9)	9.8 (0.03)
Quebec, Canada r	30 (1.3)	576 (4.1)	48 (1.0)	558 (3.0)	22 (1.1)	543 (3.7)	9.4 (0.06)
South Africa (6) ☒ r	22 (0.8)	436 (6.6)	62 (0.8)	380 (5.1)	16 (0.7)	366 (6.1)	9.5 (0.04)
Abu Dhabi, UAE s	20 (0.7)	513 (6.1)	66 (1.0)	457 (4.1)	14 (0.6)	450 (6.7)	9.4 (0.02)
Dubai, UAE x	26 (0.8)	594 (3.6)	60 (1.0)	555 (2.8)	14 (0.8)	555 (5.6)	9.6 (0.03)

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

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

As described in [Chapter 15](#) of *Methods and Procedures: PIRLS 2021 Technical Report* (Yin & Reynolds, 2023), PIRLS 2021 reports context questionnaire data by creating context questionnaire scales based on Rasch modeling. The context questionnaire scales are available in the PIRLS 2021 International Database for analysis within the context data file corresponding to the respondent of the questionnaire items. Each context questionnaire scale variable is a Rasch score with an international centerpoint of 10 and standard deviation of 2. From each context questionnaire scale, an index is derived that divides the range of scores on that scale into three categories: the most desirable scores (high values), the least desirable scores (low values), and the remaining scores in between.

These context questionnaire scales and their corresponding indices were reported in [PIRLS 2021 International Results in Reading](#). Exhibit 1.17 shows one such example, [Exhibit 5.3](#) of the report, with results for the PIRLS 2021 *Parents Like Reading* scale, based on parents' responses to a set of eight statements on this topic. Results on the Rasch scale are reported for each country as an “Average Scale Score” and its corresponding index is reported as the percentages of students with parents in each category—“very much like reading,” “somewhat like reading,” and “do not like reading”—along with their average reading achievement.

This example replicates both the average scale score column of Exhibit 1.17 and the columns reporting percentages of students and their average achievement in each category. This is accomplished in two analysis steps, both using the merged ASHALLR5 data file.

Step 1: Compute Average Scale Score

The first step of Example 5 will compute the average scale score of the *Parents Like Reading* scale for each country, using the scale variable ASBHPLR. It is conducted in the **Analysis Module** of the IEA IDB Analyzer using the following steps. The completed Analysis Module for this step is shown in Exhibit 1.18.

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

IEA IDB Analyzer: Analysis Module - (Version 5.0.18)

1 Analysis File: C:\PIRLS2021\Merge\ASHALLR5.Rdata Select

2 Analysis Type: PIRLS (Using Student Weights) Statistic Type: Percentages and Means Plausible Value Option: None Used Number of Decimals: 2 Show Graphs: Yes

3 Select Variables:

Name	Description
ASDGHRL	HOME RESOURCES FOR LEARNING/IDX
ASBHSES	HOME SOCIOECONOMIC STATUS/SCL
ASDHSES	HOME SOCIOECONOMIC STATUS/IDX
ASBHELA	EARLY LIT ACTIVITIES BEFORE SCHOOL/SCL
ASDHELA	EARLY LIT ACTIVITIES BEFORE SCHOOL/IDX
ASBHENA	EARLY NUM ACTIVITIES BEFORE SCHOOL/SCL
ASDHENA	EARLY NUM ACTIVITIES BEFORE SCHOOL/IDX
ASBHLEN	EARLY LIT NUM ACTIV BEFORE SCHOOL/SCL
ASDHLEN	EARLY LIT NUM ACTIV BEFORE SCHOOL/IDX
ASBHFLT	EARLY LITERACY TASKS/SCL
ASDHFLT	EARLY LITERACY TASKS/IDX
ASBHPCS	PARENTS PERCEPTIONS OF CHILD SCHOOL/SCL
ASDHPCS	PARENTS PERCEPTIONS OF CHILD SCHOOL/IDX
ASDHPLR	PARENTS LIKE READING/IDX
ASDHAPS	STUDENT ATTENDED PRESCHOOL
ASDHEDUP	PARENTS' HIGHEST EDUCATION LEVEL
ASDHOCPP	PARENTS' HIGHEST OCCUPATION LEVEL
VERSION	Data version
IDPOP	Population ID
IDGRADER	Standardized Grade ID

Grouping Variables: ☒ Exclude Missing From Analysis

Name	Description
IDCNTRY	Cntry ID

Separate Tables by:

Name	Description

Analysis Variables:

Name	Description
ASBHPLR	PARENTS LIKE READING/SCL

Weight Variable:

Name	Description
TOTWGT	TOTAL STUDENT WEIGHT

Output Files: C:\PIRLS2021\Analysis\REA_ASBHPLR_scale.* Modify Start R Return to Main Menu Help

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Specify the data file ASHALLR5 as the **Analysis File** by clicking the **Select** button.
3. Select **PIRLS (Using Student Weights)** as the **Analysis Type**.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **None Used** as the **Plausible Value Option**, because achievement scores are not used for this part of the analysis.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce two bar graphs in the output file: one for average scale score by country, and one for average percent of the total students in each country.

8. The IDB Analyzer automatically selects the variable IDCNTRY for the **Grouping Variables**. No additional grouping variables are needed for this analysis. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (non-achievement) variables. See the IEA IDB Analyzer Help manual for more information.
10. Specify the variable ASBHPLR to be used for the analysis by first clicking the **Analysis Variables** field to activate it. Then, select ASBHPLR from the list of available variables in the left panel, and move it to the right **Analysis Variables** field by clicking the **right arrow** (►) button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis uses student context data combined with home context data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer will create an R, SPSS, or SAS syntax file of the same name and in the same folder, with the code necessary to perform the analysis. In Exhibit 1.18, the syntax file REA_ASBHPLR_scale.R and the output files with the same name will be created and stored in the folder C:\PIRLS2021\Analysis.
13. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu and select the **All** menu option. In SAS, click the **Run** (⌘) button (or select **Submit** in the **Run** menu).

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

Exhibit 1.19 shows the results for the first step of Example 5 in the R output file with six example countries: Bulgaria, Hong Kong SAR, Ireland, Islamic Republic of Iran, New Zealand, and Oman. This step of the analysis produces the same results output files as described in Example 1, with IDCNTRY as the grouping variable and the scale variable ASBHPLR as the outcome instead of ASRREA because no PVs were used. Countries are ordered according to their numeric ISO code (see Chapter 2).

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

Report											
Analysis for ASBHPLR by IDCNTY											
Cntry ID	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASBHPLR (Mean)	ASBHPLR (s.e.)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	4020	58840	692.53	3.70	0.10	9.82	0.07	2.15	0.06	0.69	76
Hong Kong, SAR	3755	55330	777.84	3.48	0.10	9.00	0.03	1.39	0.02	1.97	73
Iran, Islamic Republic of	5874	1324635	40054.24	83.21	0.44	9.51	0.04	1.48	0.04	1.23	109
Ireland	4298	63659	1825.30	4.00	0.15	10.00	0.05	1.97	0.03	7.20	75
Oman	5184	63162	1042.30	3.97	0.12	9.54	0.03	1.26	0.03	2.59	110
New Zealand	2342	26305	875.22	1.65	0.06	10.04	0.06	1.95	0.03	58.21	95

New Zealand is interpreted as an example. On average, students in New Zealand scored 10.04 based on their parents' responses to the items on the *Parents Like Reading* context questionnaire scale, with a standard error of 0.06. This is located approximately at the international centerpoint of 10. Chapter 15 in [Methods and Procedures: PIRLS 2021 Technical Report](#) provides more information about the PIRLS context scales and their interpretation.

When analyzing context data, it is important to check the “Percent Missing” column in the output, which shows that 58.21% of students in New Zealand do not have data for the scale. In Exhibit 1.17, New Zealand is listed below the International Average because they had a high level of missing data on the ASBHPLR variable. The “x” annotation indicates that data were available for only 40–50% of students, so their results should be interpreted with caution. The data are reported but do not contribute to the International Average. Note that the “Percent Missing” column only reports missing data on the outcome variable.

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

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

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

IEA IDB Analyzer: Analysis Module - (Version 5.0.18)

1 Analysis File: C:\PIRLS2021\Merge\ASHALLR5.Rdata Select

2

Analysis Type: PIRLS (Using Student Weights) Statistic Type: Percentages and Means Plausible Value Option: Use PVs Number of Decimals: 2 Show Graphs: Yes

3 Select Variables:

Name	Description
ASBH22	COVLEARNING PROGRESS AFFECTED
ASBGHRL	HOME RESOURCES FOR LEARNING/SCL
ASDGHRL	HOME RESOURCES FOR LEARNING/IDX
ASBHSES	HOME SOCIOECONOMIC STATUS/SCL
ASDHSES	HOME SOCIOECONOMIC STATUS/IDX
ASBHELA	EARLY LIT ACTIVITIES BEFORE SCHOOL/SCL
ASDHELA	EARLY LIT ACTIVITIES BEFORE SCHOOL/IDX
ASBHENA	EARLY NUM ACTIVITIES BEFORE SCHOOL/SCL
ASDHENA	EARLY NUM ACTIVITIES BEFORE SCHOOL/IDX
ASBHELN	EARLY LIT NUM ACTIV BEFORE SCHOOL/SCL
ASDHELN	EARLY LIT NUM ACTIV BEFORE SCHOOL/IDX
ASBHILT	EARLY LITERACY TASKS/SCL
ASDHILT	EARLY LITERACY TASKS/IDX
ASBHPCS	PARENTS PERCEPTIONS OF CHILD SCHOOL/SCL
ASDHPCS	PARENTS PERCEPTIONS OF CHILD SCHOOL/IDX
ASBHPLR	PARENTS LIKE READING/SCL
ASDHAPS	STUDENT ATTENDED PRESCHOOL
ASDHEDUP	PARENTS' HIGHEST EDUCATION LEVEL
ASDHOCPP	PARENTS' HIGHEST OCCUPATION LEVEL
VERSION	Data version

4

Grouping Variables: ☒ Exclude Missing From Analysis

Name	Description
IDCNTRY	Cntry ID
ASDHPLR	PARENTS LIKE READING/IDX

Separate Tables by:

Name	Description
------	-------------

Plausible Values:

Name	Description
ASRREA01-05	1ST TO 5TH PLAUSIBLE VALUE: OVERALL READING PV1

Weight Variable:

Name	Description
TOTWGT	TOTAL STUDENT WEIGHT

Output Files: C:\PIRLS2021\Analysis\REA_ASDHPLR_index.* Modify

Start R Return to Main Menu Help

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Specify the data file ASHALLR5 as the **Analysis File** by clicking the **Select** button.
3. Select **PIRLS (Using Student Weights)** as the **Analysis Type**.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**, because average achievement will be computed by the grouping variable ASDHPLR.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce three graphs in the output file: a line graph of average achievement for each scale category by country, a clustered bar graph of average

achievement for each scale category by country, and a stacked bar graph of average percent of students for each scale category by country. R also produces separate graphs for each country.

8. Specify the variable ASDHPLR as a second grouping variable by first clicking the **Grouping Variables** field to activate it. Then, select ASDHPLR from the list of variables in the left panel, and move it to the **Grouping Variables** field by clicking the **right arrow (►)** button. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (not achievement) variables. See the Help manual for more information.
10. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASRREA01–05 from the list of available variables in the left panel, and move it to the right **Plausible Values** field by clicking the **right arrow (►)** button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because this example analysis combines student context data with home context data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. In Exhibit 1.20, the syntax file REA_ASDHPLR_index.R and the output files with the same name will be created and stored in the C:\PIRLS2021\Analysis folder.
13. Click the **Start R** button to create the R script and open it for execution. The IEA IDB Analyzer will display a warning if it is about to overwrite an existing file in the specified folder. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard.

Exhibit 1.21 shows the results for the second step of Example 5 with six example countries: Bulgaria, Hong Kong SAR, Ireland, Islamic Republic of Iran, New Zealand, and Oman. The results are presented in the “Report” section of the output. This analysis produces the results output files consistent with those described in Example 2, with ASDHPLR as the second grouping variable instead of ITSEX. Each country’s results are displayed on three lines, one for each value of the scale index variable.

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

Report												
Analysis for ASRREA by IDCNTRY ASDHPLR												
Cntry ID	PARENTS LIKE READING/IDX	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASRREA (Mean)	ASRREA (s.e.)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	Very Much Like	1848	24117	759.68	40.99	1.25	569.74	3.24	77.23	3.00	0.00	73
Bulgaria	Somewhat Like	1629	23963	752.19	40.73	1.17	537.06	3.96	84.31	2.82	0.00	76
Bulgaria	Do Not Like	543	10760	862.20	18.29	1.43	480.88	6.97	86.61	4.85	0.00	74
Hong Kong, SAR	Very Much Like	529	7632	463.09	13.79	0.81	592.39	3.68	59.67	2.77	0.00	73
Hong Kong, SAR	Somewhat Like	2337	34675	709.45	62.67	0.93	573.12	2.71	64.64	1.92	0.00	73
Hong Kong, SAR	Do Not Like	889	13023	436.16	23.54	0.73	566.61	3.78	67.71	3.08	0.00	73
Iran, Islamic Republic of	Very Much Like	1396	312618	19092.36	23.60	1.03	449.37	4.40	86.52	2.84	0.00	108
Iran, Islamic Republic of	Somewhat Like	3650	825016	25121.54	62.28	0.89	405.47	5.49	100.89	4.39	0.00	109
Iran, Islamic Republic of	Do Not Like	828	187000	10528.06	14.12	0.77	384.53	7.03	99.75	4.38	0.00	105
Ireland	Very Much Like	1819	26540	1087.95	41.69	1.19	599.35	2.67	70.39	1.67	0.00	75
Ireland	Somewhat Like	1828	26752	948.63	42.02	1.06	574.28	2.80	73.84	2.20	0.00	75
Ireland	Do Not Like	651	10366	582.86	16.28	0.71	555.33	3.79	74.66	3.01	0.00	73
Oman	Very Much Like	1118	13661	667.85	21.63	0.93	451.13	6.02	107.09	3.37	0.00	108
Oman	Somewhat Like	3514	42924	850.55	67.96	0.95	427.83	3.75	107.70	2.15	0.00	110
Oman	Do Not Like	552	6577	387.65	10.41	0.59	406.57	8.70	113.96	5.23	0.00	105
New Zealand	Very Much Like	1033	11468	510.63	43.60	1.41	563.36	3.38	80.85	2.53	0.00	94
New Zealand	Somewhat Like	910	10411	489.67	39.58	1.17	531.89	4.55	87.70	3.20	0.00	94
New Zealand	Do Not Like	399	4427	261.77	16.83	0.84	508.32	5.70	92.45	4.04	0.00	88

As shown in the three lines of the results for New Zealand, 43.60% of fourth grade students in New Zealand had parents who “very much like” reading (standard error of 1.41) and their average reading achievement was 563.36 (standard error of 3.38); 39.58% of students had parents who “somewhat like” reading (standard error of 1.17) and their average reading achievement was 531.89 (standard error of 4.55); and 16.83% of students had parents who “do not like” reading (standard error of 0.84) and their average reading achievement was 508.32 (standard error of 5.70).

In the first step of this example, the “Percent Missing” column indicated that New Zealand had a high proportion of missing data on the context scale outcome. Because in Step 8, the box was checked for **Excluding Missing From Analysis**, any cases missing data on the

grouping variable were excluded and not accounted for in the output. Nevertheless, it is important to consider missing data when interpreting the results of analyses using context data. Users of the data can check for the proportion of missing data on the context variables in the PIRLS 2021 Data Almanacs, described in Chapter 2 of this User Guide.

Conducting Analyses with PIRLS School Context Data

When analyzing school context data from the PIRLS 2021 International Database, it is preferable that they be analyzed to make inferences about students attending schools with a given characteristic, rather than about schools of a given characteristic. Analyzing school context data with student achievement requires that the school context data files be merged with the student context data files to retrieve the achievement scores and required sample design variables.

This section presents an analysis conducted using the IEA IDB Analyzer with school context data collected from principals of schools attended by fourth grade students with the PIRLS 2021 School Questionnaire. Example 6 uses school context data merged with student context data to compute the percentages and average achievement of fourth grade students who attend schools composed of students with different levels of socioeconomic background. The results of this analysis are presented in [Exhibit 6.1](#) of *PIRLS 2021 International Results in Reading*, repeated below in Exhibit 1.22.

Example 6 – Analysis of Average Achievement by School Socioeconomic Composition

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

Before conducting analyses using school context variables, users should refer to the relevant codebook for the data file to identify the appropriate variables related to the school's composition of students by socioeconomic background and understand the coding scheme.

[Supplement 1](#) of this User Guide presents all the context questionnaires administered in PIRLS 2021 and the associated variable names under which the data are saved.

[Supplement 2](#) should also be checked for any national adaptations made to the questionnaire items that may impact international comparability.

Exhibit 1.22: Exhibit of Example 6 – Analysis of Average Achievement by School Socioeconomic Composition (**Exhibit 6.1** of PIRLS 2021 International Results in Reading)

Grade 4						IEA PIRLS 2021
Exhibit 6.1: School Composition by Socioeconomic Background of the Student Body						
Students' Results based on Principals' Reports						
Assessed Fourth Grade Students at the End of the School Year						
<div> <div></div> Assessed one year later than originally scheduled </div>						
<div> <div></div> Delayed Assessment of Fourth Grade Cohort at the Beginning of Fifth Grade </div>						
Country	More Affluent		Neither More Affluent Nor More Disadvantaged		More Disadvantaged	
	Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement
Kazakhstan	78 (3.0)	509 (3.0)	20 (3.1)	481 (5.8)	2 ~	~
Lithuania	78 (4.0)	555 (2.7)	18 (3.5)	532 (5.6)	4 (2.0)	525 (11.1)
Russian Federation	77 (2.8)	576 (3.7)	19 (2.8)	538 (7.1)	5 (1.7)	547 (17.8)
Uzbekistan	73 (3.9)	442 (3.6)	24 (3.8)	430 (5.5)	3 (1.4)	426 (15.0)
Qatar	66 (3.5)	491 (5.0)	26 (3.5)	483 (7.9)	8 (1.9)	461 (12.1)
North Macedonia	64 (5.3)	453 (5.9)	27 (4.3)	436 (14.0)	9 (3.8)	402 (21.2)
Sweden	63 (4.0)	560 (3.7)	26 (4.1)	524 (4.6)	10 (2.5)	514 (8.5)
Croatia	61 (4.1)	561 (3.0)	33 (4.0)	555 (3.9)	6 (1.9)	521 (16.8)
Spain	61 (3.0)	530 (2.4)	28 (3.2)	515 (4.0)	11 (2.1)	483 (6.9)
Denmark	60 (3.9)	550 (2.8)	32 (3.8)	521 (4.3)	8 (1.9)	525 (6.5)
Netherlands	58 (4.4)	536 (3.6)	24 (4.3)	518 (5.0)	19 (3.5)	506 (8.6)
Singapore	57 (0.0)	600 (3.6)	36 (0.0)	573 (5.5)	7 (0.0)	552 (16.6)
Saudi Arabia	57 (5.4)	456 (5.4)	25 (4.7)	441 (10.2)	18 (3.9)	450 (13.4)
United Arab Emirates	56 (1.4)	519 (3.2)	27 (1.5)	510 (4.5)	17 (1.2)	479 (8.6)
Belgium (Flemish)	56 (4.3)	518 (3.1)	29 (4.5)	512 (3.8)	15 (3.4)	479 (5.7)
Hungary	53 (4.0)	563 (3.6)	27 (4.3)	528 (7.0)	20 (3.3)	488 (8.1)
Slovenia	51 (4.8)	524 (3.0)	36 (4.7)	517 (3.1)	13 (3.0)	508 (6.1)
Kosovo	50 (5.6)	423 (5.4)	34 (5.0)	415 (7.2)	16 (4.0)	412 (5.9)
Norway (5)	49 (4.1)	548 (2.7)	45 (4.0)	533 (2.4)	6 (2.1)	513 (12.8)
Czech Republic	48 (3.5)	550 (3.4)	43 (3.5)	535 (3.7)	10 (2.1)	506 (9.0)
Montenegro	47 (0.7)	492 (2.0)	34 (0.8)	486 (3.3)	19 (0.6)	475 (3.6)
Belgium (French)	47 (3.3)	511 (3.7)	26 (3.8)	484 (5.2)	28 (3.4)	475 (5.1)
France	46 (3.5)	530 (3.1)	26 (3.6)	516 (6.0)	27 (3.2)	485 (4.6)
Cyprus	46 (4.0)	529 (4.3)	41 (3.9)	501 (3.4)	13 (2.2)	471 (5.2)
Israel	44 (3.5)	537 (3.7)	26 (3.6)	520 (4.2)	30 (2.9)	465 (5.8)
Ireland	42 (4.8)	595 (3.3)	33 (4.3)	582 (3.9)	25 (3.2)	550 (4.9)
Oman	41 (3.7)	435 (5.9)	36 (3.5)	440 (6.9)	24 (3.4)	403 (8.5)
Australia	40 (3.8)	562 (3.0)	34 (3.6)	540 (3.4)	26 (3.1)	508 (5.6)
Macao SAR	39 (0.1)	546 (1.9)	35 (0.1)	526 (1.7)	26 (0.1)	532 (2.1)
Northern Ireland	38 (4.5)	587 (4.1)	31 (4.8)	564 (4.1)	31 (3.5)	543 (5.2)
New Zealand	38 (3.7)	553 (3.9)	36 (3.8)	528 (4.7)	27 (3.3)	483 (6.5)
Serbia	37 (4.1)	524 (4.2)	43 (4.0)	513 (4.7)	20 (3.5)	494 (7.0)
Albania	37 (4.2)	527 (5.6)	30 (3.6)	516 (5.6)	33 (4.3)	495 (4.6)
Finland	37 (4.3)	561 (3.1)	52 (3.9)	548 (2.6)	11 (2.7)	518 (9.4)
Bulgaria	37 (4.1)	572 (4.1)	45 (4.1)	550 (4.1)	18 (2.7)	483 (9.5)
Italy	35 (4.0)	548 (3.4)	44 (3.8)	540 (3.1)	22 (3.0)	517 (6.3)
Malta	33 (5.4)	538 (4.2)	61 (5.4)	506 (4.0)	6 (2.6)	458 (13.3)
Brazil	32 (4.7)	473 (9.6)	16 (3.9)	438 (14.5)	51 (5.0)	389 (11.5)
Georgia	32 (3.5)	502 (3.8)	36 (3.7)	490 (4.9)	32 (3.8)	489 (3.7)
Latvia	32 (4.0)	531 (5.3)	61 (4.1)	528 (3.3)	7 (1.8)	509 (12.5)
United States	32 (5.0)	587 (8.0)	12 (4.1)	555 (12.2)	56 (5.0)	524 (9.2)
Portugal	32 (2.9)	537 (3.6)	37 (3.4)	517 (2.8)	32 (3.3)	506 (4.5)
England	32 (4.1)	578 (5.0)	25 (4.3)	565 (3.8)	43 (4.4)	539 (3.7)
Austria	30 (3.7)	543 (3.1)	43 (4.1)	537 (3.5)	26 (3.1)	504 (5.0)
Germany	29 (3.6)	543 (3.9)	36 (3.4)	533 (3.5)	35 (2.8)	501 (4.4)
Iran, Islamic Rep. of	29 (3.5)	458 (8.0)	21 (2.8)	422 (6.9)	50 (3.6)	382 (6.7)
Poland	29 (3.9)	555 (3.7)	65 (4.2)	547 (3.0)	7 (2.3)	537 (7.3)
Bahrain	28 (2.6)	494 (8.4)	27 (2.9)	449 (8.5)	44 (3.1)	439 (5.1)
Turkiye	28 (3.3)	530 (5.3)	21 (3.3)	505 (7.3)	51 (3.4)	474 (4.6)
Hong Kong SAR	28 (3.5)	588 (4.3)	28 (4.1)	572 (6.7)	45 (4.3)	564 (4.1)
Egypt	27 (3.3)	410 (7.6)	25 (3.8)	389 (10.7)	49 (3.7)	359 (8.8)
Chinese Taipei	25 (3.2)	558 (3.2)	67 (3.7)	540 (2.4)	8 (2.1)	521 (9.3)
Azerbaijan	13 (2.8)	410 (11.6)	22 (3.3)	450 (10.4)	64 (4.1)	444 (5.2)
Jordan	13 (3.1)	423 (13.3)	33 (4.2)	385 (10.6)	55 (4.3)	367 (9.0)
South Africa	11 (1.9)	420 (22.2)	16 (2.7)	331 (15.2)	73 (3.1)	264 (5.7)
Morocco	7 (1.6)	452 (20.1)	7 (2.0)	402 (15.8)	86 (2.1)	363 (5.7)
International Average	43 (0.5)	521 (0.9)	32 (0.5)	502 (0.9)	25 (0.4)	479 (1.2)
Slovak Republic	- -	- -	- -	- -	- -	- -
Benchmarking Participants						
Quebec, Canada	54 (5.6)	559 (4.1)	30 (4.7)	544 (4.9)	16 (3.7)	540 (6.2)
British Columbia, Canada	49 (4.7)	549 (5.8)	41 (4.5)	535 (5.3)	11 (2.9)	508 (11.7)
Newfoundland & Labrador, Canada	35 (6.8)	536 (5.9)	54 (7.1)	522 (3.8)	11 (3.1)	496 (7.8)
Alberta, Canada	32 (4.8)	561 (5.2)	47 (5.3)	538 (5.7)	21 (3.9)	516 (8.7)
Moscow City, Russian Federation	91 (2.1)	598 (2.2)	7 (2.0)	593 (6.4)	2 ~	~
South Africa (6)	11 (2.6)	465 (32.9)	15 (3.1)	447 (19.5)	74 (3.4)	360 (6.1)
Dubai, UAE	65 (0.4)	581 (2.0)	30 (0.3)	560 (2.7)	5 (0.2)	527 (4.8)
Abu Dhabi, UAE	57 (2.1)	478 (5.9)	27 (2.6)	467 (6.6)	16 (2.9)	470 (16.6)

The codebook for the school context data file indicates that the derived variable ACDGSBC contains information on the socioeconomic composition of schools in three categories represented by the columns of Exhibit 1.22. As described in [Supplement 3](#) to this User Guide, two source variables were used to derive ACDGSBC: ACBG03A for the percentage of students economically disadvantaged and ACBG03B for the percentage of students economically affluent. Schools are characterized as “more affluent,” “more disadvantaged,” or “neither more affluent nor more disadvantaged.”

This example uses the merged data file ACGALLR5 described earlier in this chapter under *Merging Data Files with the IEA IDB Analyzer*. This example analysis is conducted in the **Analysis Module** of the IEA IDB Analyzer using the following steps. The completed Analysis Module is shown in Exhibit 1.23.

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

IEA IDB Analyzer: Analysis Module - (Version 5.0.18)

1 Analysis File: C:\PIRLS2021\Merge\ACGALLR5.Rdata Select

2 Analysis Type: PIRLS (Using Student Weights) Statistic Type: Percentages and Means Plausible Value Option: Use PVs Number of Decimals: 2 Show Graphs: Yes

3 Select Variables:

Name	Description
ACBG03B	GEN\STUDENTS BACKGROUND\ECONOMIC AFFLUEN
ACBG04	GEN\PERCENT OF STUDENTS <LANG OF TEST>
ACBG05A	GEN\HOW MANY PEOPLE LIVE IN AREA
ACBG05B	GEN\IMMEDIATE AREA OF SCH LOCATION
ACBG06A	GEN\INSTRUCTIONAL DAYS PER YEAR
ACBG06B	GEN\TOTAL INSTRUCTIONAL TIME\MINUTES
ACBG06C	GEN\INSTRUCTIONAL DAYS IN 1 CALENDAR WEEK
ACBG07A	GEN\HAVE SCHOOL LIBRARY
ACBG07B	GEN\BOOKS IN LIBRARY
ACBG07C	GEN\BORROW MATERIAL FROM LIBRARY
ACBG08	GEN\ACCESS TO DIGITAL BOOKS
ACBG09	GEN\TOTAL NUMBER COMPUTERS
ACBG10AA	GEN\SHORTAGE\GEN\INSTRUCTIONAL MATERIAL
ACBG10AB	GEN\SHORTAGE\GEN\SUPPLIES
ACBG10AC	GEN\SHORTAGE\GEN\SCHOOL BUILDINGS
ACBG10AD	GEN\SHORTAGE\GEN\HEATING SYSTEMS
ACBG10AE	GEN\SHORTAGE\GEN\INSTRUCTIONAL SPACE
ACBG10AF	GEN\SHORTAGE\GEN\TECHNOLOGICAL STAFF
ACBG10AG	GEN\SHORTAGE\GEN\TECH RES SUPPORT TCH
ACBG10AH	GEN\SHORTAGE\GEN\TECH RES SUPPORT STD

Grouping Variables: ☒ Exclude Missing From Analysis

Name	Description
IDCNTRY	Cntry ID
ACDGSBC	SCHOOL COMPOSITION BY STD BACKGROUND

Separate Tables by:

Name	Description

Plausible Values:

Name	Description
ASRREA01-05	1ST TO 5TH PLAUSIBLE VALUE: OVERALL READING PV1

Weight Variable:

Name	Description
TOTWGT	TOTAL STUDENT WEIGHT

4 Output Files: C:\PIRLS2021\Analysis\REA_ACDGSBC.* Modify


Return to Main Menu Help

Start R

1. Open the **Analysis Module** of the IEA IDB Analyzer.

2. Select the merged data file ACGALLR5 as the **Analysis File** by clicking the **Select** button.
3. Select **PIRLS (Using Student Weights)** as the **Analysis Type**, because the school context data is analyzed as student attributes.
4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**, because average achievement will be computed by the grouping variable ACDGSBC.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce three graphs in the output file: a line graph of average achievement for each category of school composition by country, a clustered bar graph of average achievement for each category of school composition by country, and a stacked bar graph of average percent of students for each category of school composition by country. R also provides separate graphs for each country.
8. Specify the variable ACDGSBC as a second grouping variable by first clicking the **Grouping Variables** field to activate it. Then, select ACDGSBC from the list of variables in the left panel, and move it to the **Grouping Variables** field by clicking the **right arrow (►)** button. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (not achievement) variables. See the Help manual for more information.
10. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASRREA01–05 from the list of available variables in the left panel, and move it to the right **Plausible Values** field by clicking the **right arrow (►)** button.
11. The **Weight Variable** is selected automatically by the software; TOTWGT is selected by default because of the **Analysis Type** selected in Step 3 for this analysis which uses school context data linked to student context data.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer also will create an R syntax file (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the

analysis. In Exhibit 1.23, the syntax file REA_ACDGSBC.R and the output files with the same name will be created and stored in the folder C:\PIRLS2021\Analysis.

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

The results as shown in the R output file are presented in Exhibit 1.24 with four example countries: Bulgaria, Hong Kong SAR, the Islamic Republic of Iran, and Oman. The results are presented in the same manner as in Example 2, with countries identified in the first column and the second column describing the categories of ACDGSBC.

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

Report												
Analysis for ASRREA by IDCNTY ACDGSBC												
Cntry ID	SCHOOL COMPOSITION BY STD BACKGROUND	N of Cases	Sum of TOTWGT	Sum of TOTWGT (s.e.)	Percent	Percent (s.e.)	ASRREA (Mean)	ASRREA (s.e.)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	More Affluent	1605	19094	2175.10	36.74	4.07	571.52	4.10	73.63	1.93	0.00	38
Bulgaria	Neither More Affluent nor More Disadvantaged	1694	23391	2116.41	45.01	4.08	550.35	4.06	81.27	2.66	0.00	42
Bulgaria	More Disadvantaged	402	9489	1527.38	18.26	2.69	483.09	9.54	89.15	5.81	0.00	20
Hong Kong, SAR	More Affluent	1001	14001	1822.78	27.72	3.46	588.26	4.25	63.77	3.62	0.00	29
Hong Kong, SAR	Neither More Affluent nor More Disadvantaged	927	13934	2123.16	27.59	4.13	572.33	6.71	68.27	5.61	0.00	33
Hong Kong, SAR	More Disadvantaged	1478	22567	2166.34	44.69	4.26	564.35	4.12	64.88	2.56	0.00	44
Iran, Islamic Republic of	More Affluent	1531	384682	53877.84	28.95	3.54	458.36	8.00	87.66	7.38	0.00	44
Iran, Islamic Republic of	Neither More Affluent nor More Disadvantaged	1525	284334	37172.85	21.40	2.84	422.09	6.86	93.88	4.43	0.00	42
Iran, Islamic Republic of	More Disadvantaged	2827	659575	44433.22	49.64	3.56	382.04	6.75	98.05	4.67	0.00	74
Oman	More Affluent	1860	24583	2381.10	40.55	3.72	434.78	5.90	106.24	3.22	0.00	63
Oman	Neither More Affluent nor More Disadvantaged	1855	21563	2143.27	35.57	3.49	440.46	6.88	106.98	2.59	0.00	59
Oman	More Disadvantaged	1292	14472	2077.75	23.87	3.41	402.59	8.51	111.81	4.12	0.00	48

In this example, each country's results are presented on three lines, one for each value of the ACDGSBC variable. There are fewer lines if any category does not have any observations. As shown in the three lines of results for Bulgaria, 36.74% (standard error of 4.07) of fourth grade students attended "more affluent" schools and their average achievement was 571.52 (standard error of 4.10); 45.01% (standard error of 4.08) attended "neither more affluent not more disadvantaged" and their average achievement was 550.35 (standard error of 4.06); and 18.26% (standard error of 2.69) attended "more disadvantaged" schools and their average achievement was 483.09 (standard error of 9.54).

Conducting Analyses with PIRLS Teacher Context Data

This section presents an analysis conducted using the IEA IDB Analyzer with teacher context data from the PIRLS 2021 International Database. Analyses with PIRLS teacher context data seek to make inferences about students whose teachers have a given characteristic, attitude, or instructional practice. Because the teachers in PIRLS do not constitute representative samples of teachers, inferences should not be made about teachers themselves.

The *PIRLS 2021 International Results in Reading* report does not include any results of teacher data, and analyses using the data should be interpreted with caution because of the coinciding of PIRLS 2021 data collection with the COVID-19 pandemic. The influence that particular teachers had on students' achievement during disruptions to schooling may be limited, and situations varied greatly within and across countries. The delayed assessments at the beginning of fifth grade had the most impact on the teacher questionnaire data, and in some instances, the data were excluded from the International Database. Particularly if students switched schools between fourth and fifth grade, fifth grade teachers likely were not a reliable source of information about the instruction the students received during their fourth year of schooling. [Supplement 2](#) to this User Guide describes these instances as well as any adaptations made to the questionnaires to accommodate the changed data collection schedule. Most countries tried to contact the previous fourth grade teachers to complete the teacher questionnaire, keeping the prior school year in mind.

As mentioned earlier, users conducting analyses using contextual data should check for any missing data countries have on the context variables, as high levels of missing data could bias results. *PIRLS 2021 International Results in Reading* annotates results if missing data rates exceed 15% of students. Results are not reported for a country if data are available for fewer than 40% of students, and countries with data available for fewer than 50% are reported at the bottom of the results exhibits.

Similarly, achievement estimates can be unreliable if based on small groups of students. Therefore, achievement results are not reported for any context variable categories containing fewer than 2.5% of students.

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

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

This example involves using the **Percentages and Means** statistic type with the **Use PVs** option to estimate the percentages of students with their average reading achievement by reporting categories of teachers' level of formal education.

Before conducting analyses using PIRLS contextual variables, users should refer to the relevant codebook for the data file to identify the appropriate variables related to teachers' education and understand the coding scheme. Also, [Supplement 1](#) of this User Guide presents all the context questionnaires administered in PIRLS 2021 and the associated variable names under which the data are saved. [Supplement 2](#) should also be checked for any national adaptations made to the questionnaire items that may impact international comparability.

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

1. Did not complete <Upper secondary education—ISCED Level 3>
2. <Upper secondary education—ISCED Level 3>
3. <Post-secondary, non-tertiary education—ISCED Level 4>
4. <Short-cycle tertiary education—ISCED Level 5>
5. <Bachelor's or equivalent level—ISCED Level 6>
6. <Master's or equivalent level—ISCED Level 7>
7. <Doctor or equivalent level—ISCED Level 8>

The ATBG04 item was specifically designed to be adapted to each unique education system, indicated by the category labels being contained within brackets in the codebooks, e.g., "<Bachelor's or equivalent level—ISCED Level 6>." Despite the various education

system structures and translations across countries, the variable is made internationally comparable by adhering to the International Standard Classification of Education (ISCED; UNESCO, 2012), which provides an internationally accepted classification scheme for describing levels of schooling across countries. For all such variables, [Supplement 2](#) provides English back translations for the answer categories for each country.

This example uses the merged ATGALLR5 data file described earlier in this chapter under *Merging Data Files with the IEA IDB Analyzer*. The analysis is conducted in the **Analysis Module** of the IEA IDB Analyzer, shown completed in Exhibit 1.25, using the following steps:

Exhibit 1.25: IEA IDB Analyzer Analysis Module Setup for Example 7 – Analysis of Average Achievement by Teachers’ Formal Education

IEA IDB Analyzer: Analysis Module - (Version 5.0.18)

1 Analysis File:

2 Analysis Type: Statistic Type: Plausible Value Option: Number of Decimals: Show Graphs:

3 Select Variables:

Name	Description
ATBG05AB	GEN\MAJOR AREA OF STUDY\EDUCATION SECOND
ATBG05AC	GEN\MAJOR AREA OF STUDY\<LANGUAGE TEST>
ATBG05AD	GEN\MAJOR AREA OF STUDY\OTHER
ATBG05BA	GEN\FORMAL EDUCATION\<LANGUAGE OF TEST>
ATBG05BB	GEN\FORMAL EDUCATION\LITERATURE
ATBG05BC	GEN\FORMAL EDUCATION\PEDAGOGY
ATBG05BD	GEN\FORMAL EDUCATION\PSYCHOLOGY
ATBG05BE	GEN\FORMAL EDUCATION\LEARNING SUPPORT
ATBG05BF	GEN\FORMAL EDUCATION\READING THEORY
ATBG05BG	GEN\FORMAL EDUCATION\SPECIAL EDUCATION
ATBG05BH	GEN\FORMAL EDUCATION\SECOND LANGUAGE
ATBG05BI	GEN\FORMAL EDUCATION\ASSESSMENT METHODS
ATBG05BJ	GEN\FORMAL EDUCATION\EARLY CHILDHOOD EDUCATION
ATBG05BK	GEN\FORMAL EDUCATION\DIGITAL LITERACIES
ATBG06	GEN\READ FOR ENJOYMENT
ATBG07AA	GEN\PROF DEVELOPMENT PAST\READING COMPREHENSION
ATBG07BA	GEN\PROF DEVELOPMENT FUTURE\READING COMPREHENSION
ATBG07AB	GEN\PROF DEVELOPMENT PAST\INTEGRATING LITERACIES
ATBG07BB	GEN\PROF DEVELOPMENT FUTURE\INTEGRATING LITERACIES
ATBG07AC	GEN\PROF DEVELOPMENT PAST\ADDRESSING LANGUAGE NEEDS

Grouping Variables: ☒ Exclude Missing From Analysis

Name	Description
IDCNTRY	Cntry ID
ATBG04	GEN\LEVEL OF FORMAL EDUCATION COMPLETED

Separate Tables by:

Name	Description
------	-------------

Plausible Values:

Name	Description
ASRREA01-05	1ST TO 5TH PLAUSIBLE VALUE: OVERALL READING PV1


Weight Variable:

Name	Description
TCHWGT	WEIGHT FOR RDG TEACHER DATA COMBINED

4 Output Files:

1. Open the **Analysis Module** of the IEA IDB Analyzer.
2. Select the merged data file ATGALLR5 as the **Analysis File** by clicking the **Select** button.
3. Select **PIRLS (Using Teacher Weights)** as the **Analysis Type**.

4. Select **Percentages and Means** as the **Statistic Type**.
5. Select **Use PVs** as the **Plausible Value Option**, because average achievement will be computed by the grouping variable ATBG04.
6. The default value in the **Number of Decimals** drop-down menu is **2**. Changing this value affects only the number of visible decimal places in the output files.
7. The default value selected in the **Show Graphs** menu is **Yes**. For this analysis, selecting **Yes** will produce three graphs in the output file: a line graph of average achievement for each category of teachers' education level by country, a clustered bar graph of average achievement for each category of teachers' education level by country, and a stacked bar graph of average percent of students for each category of teachers' education level by country. R also produces separate graphs by country.
8. Specify the variable ATBG04 as a second grouping variable by first clicking the **Grouping Variables** field to activate it. Then, select ATBG04 from the list of variables in the left panel, and move it to the **Grouping Variables** field by clicking the **right arrow (►)** button. The IEA IDB Analyzer automatically checks the **Exclude Missing From Analysis**, which excludes cases with missing values on the grouping variables from the analysis. This box should be checked for this analysis.
9. The **Separate Tables by** field should be empty for this analysis. This field is used to separately analyze several grouping variables or several continuous dependent (not achievement) variables. See the Help manual for more information.
10. Specify the achievement scores to be used for the analysis by first clicking the **Plausible Values** field to activate it. Then, select ASRREA01–05 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; TCHWGT is selected by default because of the **Analysis Type** selected in Step 3.
12. Specify the desired name for the output files and the folder they will be stored in by clicking the **Define/Modify** button in the **Output Files** field. The IEA IDB Analyzer will create an R syntax file (*.R), SPSS syntax file (*.SPS), or SAS syntax file (*.SAS) of the same name and in the same folder, with the code necessary to perform the analysis. In Exhibit 1.25, the syntax file REA_ATBG04.R and the output files with the same name will be created and stored in the C:\PIRLS2021\Analysis folder.
13. Click the **Start R** button (or Start SPSS/SAS) to create the R script (or SPSS/SAS syntax file) and open it for execution. The R script can be executed by clicking the **Source** button or pressing **Ctrl+Alt+R** on the keyboard. In SPSS, open the **Run** menu

and select the **All** menu option. In SAS, click the **Run** () button (or select **Submit** in the **Run** menu).

This analysis produces the results output files in the same manner as described for Example 2, with countries identified in the first column and the second column describing the categories of the analysis variable ATBG04. Exhibit 1.26 shows the results as shown in the R output file with three example countries: Bulgaria, Hong Kong SAR, and Oman.

Exhibit 1.26: R Output for Example 7 – Analysis of Average Achievement by Teachers’ Formal Education

Report												
Analysis for ASRREA by IDCNTRY ATBG04												
Cntry ID	GENOF FORMAL EDUCATION COMPLETED	N of Cases	Sum of TCHWGT	Sum of TCHWGT (s.e.)	Percent	Percent (s.e.)	ASRREA (Mean)	ASRREA (s.e.)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing	Number of Variance Strata
Bulgaria	<Post-secondary, non-tertiary education—ISCED Level 4>	18	566	387.73	0.96	0.65	475.01	44.58	92.58	15.97	0.00	2
Bulgaria	<Short-cycle tertiary education—ISCED Level 5>	66	1164	681.63	1.96	1.15	504.38	24.62	65.14	5.91	0.00	4
Bulgaria	<Bachelor’s or equivalent level—ISCED Level 6>	701	12559	1557.39	21.20	2.55	519.38	8.21	90.18	4.50	0.00	30
Bulgaria	<Master’s or equivalent level—ISCED Level 7>	3233	44744	1532.18	75.52	2.71	547.04	3.78	86.32	3.09	0.00	68
Bulgaria	<Doctor or equivalent level—ISCED Level 8>	25	215	214.69	0.36	0.36	614.71	NaN	59.89	NaN	0.00	1
Hong Kong, SAR	<Short-cycle tertiary education—ISCED Level 5>	25	191	191.05	0.34	0.34	574.33	NaN	56.70	NaN	0.00	1
Hong Kong, SAR	<Bachelor’s or equivalent level—ISCED Level 6>	2315	35652	2308.80	63.16	4.00	570.81	3.52	66.73	2.75	0.00	63
Hong Kong, SAR	<Master’s or equivalent level—ISCED Level 7>	1517	20600	2279.11	36.50	4.02	576.30	4.31	66.10	2.86	0.00	45
Oman	<Upper secondary education—ISCED Level 3>	30	288	287.65	0.46	0.46	402.18	NaN	87.94	NaN	0.00	1
Oman	<Post-secondary, non-tertiary education—ISCED Level 4>	92	1278	660.05	2.02	1.04	461.12	41.13	115.39	25.20	0.00	4
Oman	<Short-cycle tertiary education—ISCED Level 5>	330	4371	1297.80	6.92	2.04	444.59	8.59	100.42	5.43	0.00	13
Oman	<Bachelor’s or equivalent level—ISCED Level 6>	4333	51889	2017.53	82.16	2.73	422.51	4.33	109.27	1.89	0.00	107
Oman	<Master’s or equivalent level—ISCED Level 7>	364	5108	1346.38	8.09	2.14	464.85	11.68	96.76	4.24	0.00	14
Oman	<Doctor or equivalent level—ISCED Level 8>	25	221	220.64	0.35	0.35	439.95	NaN	81.59	NaN	0.00	1

Each country's results are displayed on up to seven lines, one for each value of the ATBG04 variable. There are fewer lines if any category does not have any observations. For example, in Exhibit 1.26, Bulgaria has only five lines of results, because no teacher responded in the first two categories; Hong Kong SAR has only three lines of results because no teacher responded in the first, second, third, or seventh categories; and Oman only has six lines because no teacher responded in the first category. A country with no valid data on this variable would not appear in the results output.

As shown in the five lines of results for Bulgaria, 0.96% (standard error of 0.65) of fourth grade students in Bulgaria were taught by teachers who completed post-secondary education; 1.96% (standard error 1.15) were taught by teachers who completed short-cycle tertiary education; 21.20% (standard error 2.55) were taught by teachers who completed a bachelor's degree; 75.52% (standard error 2.71) were taught by teachers who completed a master's degree; and 0.36% (standard error 0.36) were taught by teachers who completed a doctoral degree.

For categories with very few students, standard errors associated with achievement means may appear unusually high because achievement cannot be estimated reliably for small numbers of students. In *PIRLS 2021 International Results in Reading*, average achievement is not reported for groups of students representing less than 2.5% of a given country. For the two categories with at least 2.5% of Bulgarian students, average achievement was higher for students taught by teachers with a master's degree (547.04, standard error of 3.78) compared to students taught by teachers with a bachelor's degree (519.38, standard error of 8.21).

Additional Analyses Conducted with the IEA IDB Analyzer

Computing Correlations

In addition to the analyses described above, the IEA IDB Analyzer is able to compute **Correlations** among a set of variables. Thus, it can compute correlations between context variables (such as the context questionnaire scales *Students Like Reading* and *Home Socioeconomic Status*), between achievement scores (such as Literary and Informational reading subdomains), and between a combination of both (such as students' age and their reading achievement). While these types of analyses are not demonstrated here, the steps for conducting them are similar to 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 PIRLS 2021 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’s Help manual. Note that this analysis is not yet available for R in IEA IDB Analyzer version 5.0.

References

- Fishbein, B., Foy, P., & Yin, L. (2021). *TIMSS 2019 User Guide for the International Database* (2nd ed.). Boston College, TIMSS & PIRLS International Study Center. <https://timss2019.org/international-database/>
- Foy, P. (2017). *PIRLS 2016 User Guide for the International Database*. Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/pirls2016/international-database/>
- Foy, P., & Almaskut, A. (2023). Estimating standard errors in the PIRLS 2021 results. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 13.1-13.31). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb1576>
- 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). (2023). IEA IDB Analyzer (Version 5.0) [Computer software]. Hamburg, Germany: IEA Hamburg. Retrieved from <https://www.iea.nl/data>
- Mullis, I. V. S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., & Wry, E. (2023). *PIRLS 2021 International Results in Reading*. Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb5342>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>

- RStudio Team (2020). 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.
- UNESCO. (2012). *International Standard Classification of Education (ISCED) 2011*. Montreal, Quebec: UNESCO Institute of Statistics. Retrieved from
<http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>
- Yin, L., & Reynolds, K. A. (2023). Creating and interpreting the PIRLS 2021 context questionnaire scales. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 15.1-15.161). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb6994>

CHAPTER 2

Contents and Structure of the PIRLS 2021 International Database

Overview

This chapter describes the contents of the PIRLS 2021 International Database with a special emphasis on the actual data files used in Chapter 1 of this User Guide to analyze the PIRLS 2021 assessment results with the IEA IDB Analyzer (IEA, 2023). The various data files and related materials included in the database are described, in accordance with the structure summarized in Exhibit 2.1. Descriptions of data files include the conventions for naming the various file types and variables as well as codes for missing values.

The PIRLS 2021 International Database and all accompanying documentation are available on the Boston College, TIMSS & PIRLS International Study Center website: <https://pirls2021.org/data>. Through [IEA's Data Repository](#), zip files containing all data and documents can be downloaded for each data format (R, SPSS, SAS).

Exhibit 2.1: Contents of the PIRLS 2021 International Database

Contents	Description
User Guide	User Guide and Supplements
Achievement Items Documentation	
Item Information	Table describing characteristics of each item in the PIRLS 2021 assessment
IRT Parameters	Item parameters and transform constants from the PIRLS 2021 achievement scaling
Percent Correct Statistics	Percent correct statistics for all PIRLS 2021 achievement items
International Database	
R Data	PIRLS 2021 student, home, school, and teacher data files in R format
SPSS Data	PIRLS 2021 student, home, school, and teacher data files in SPSS format
SAS Data	PIRLS 2021 student, home, school, and teacher data files in SAS format
Curriculum Data	PIRLS 2021 Curriculum Questionnaire data Excel files
Codebooks	Codebook files describing all variables in the PIRLS 2021 International Database
Data Almanacs	Summary statistics for all PIRLS 2021 achievement items and context variables
Special Programs	R, SPSS, and SAS programs to score achievement items

The PIRLS 2021 International Database contains student achievement data as well as student, home, school, and teacher context data collected in the 57 countries and 8 benchmarking participants that took part in PIRLS 2021, including the PIRLS 2021 Bridge. Exhibit 2.2 lists all of the countries and benchmarking participants included in the PIRLS 2021 International Database, along with the identification codes used in the data files comprising the database.

Exhibit 2.2: Countries in the PIRLS 2021 International Database

Country	Identification Codes		PIRLS 2021 Participation			Data Collection Disruption
	Alpha	Numeric	Paper	Digital	Bridge	
Albania	ALB	8	●			
Australia	AUS	36	●			One year later (2021)
Austria	AUT	40	●			
Azerbaijan	AZE	31	●			
Bahrain	BHR	48	●			Delayed to fifth grade
Belgium (Flemish)	BFL	956		●	●	
Belgium (French)	BFR	957	●			
Brazil	BRA	76	●			One year later (2021)
Bulgaria	BGR	100	●			
Chinese Taipei	TWN	158		●	●	
Croatia	HRV	191		●	●	Delayed to fifth grade
Cyprus	CYP	196	●			
Czech Republic	CZE	203		●	●	
Denmark	DNK	208		●	●	
Egypt	EGY	818	●			
England	ENG	926	●			One year later (2022)
Finland	FIN	246		●	●	
France	FRA	250	●			
Georgia	GEO	268	●			Delayed to fifth grade
Germany	DEU	276		●	●	
Hong Kong SAR	HKG	344	●			
Hungary	HUN	348		●	●	Delayed to fifth grade
Iran, Islamic Rep. of	IRN	364	●			One year later (2022)
Ireland	IRL	372	●			Delayed to fifth grade
Israel	ISR	376		●	●	One year later (2022)
Italy	ITA	380		●	●	
Jordan	JOR	400	●			
Kazakhstan	KAZ	398		●	●	Delayed to fifth grade
Kosovo	XKX	411	●			
Latvia	LVA	428	●			Delayed to fifth grade
Lithuania	LTU	440		●	●	Delayed to fifth grade
Macao SAR	MAC	446	●			
Malta	MLT	470		●	●	
Montenegro	MNE	499	●			

Exhibit 2.2: Countries in the PIRLS 2021 International Database (continued)

Country	Identification Codes		PIRLS 2021 Participation			Data Collection Disruption
	Alpha	Numeric	Paper	Digital	Bridge	
Morocco	MAR	504	●			Delayed to fifth grade
Netherlands	NLD	528	●			
New Zealand	NZL	554		●	●	
North Macedonia	MKD	807	●			
Northern Ireland	NIR	928	●			Delayed to fifth grade
Norway (5)	NOR	578		●	●	
Oman	OMN	512	●			
Poland	POL	616	●			
Portugal	PRT	620		●	●	
Qatar	QAT	634		●	●	Delayed to fifth grade
Russian Federation	RUS	643		●	●	
Saudi Arabia	SAU	682		●	●	Delayed to fifth grade
Serbia	SRB	688	●			
Singapore	SGP	702		●	●	
Slovak Republic	SVK	703		●	●	
Slovenia	SVN	705		●	●	
South Africa	ZAF	710	●			One year later (2021)
Spain	ESP	724		●	●	
Sweden	SWE	752		●	●	
Turkiye	TUR	792	●			
United Arab Emirates	ARE	784		●	●	Delayed to fifth grade
United States	USA	840	●			Delayed to fifth grade
Uzbekistan	UZB	860	●			
Benchmarking Participants						
Alberta, Canada	CAB	9134		●		
British Columbia, Canada	CBC	9135		●		
Newfoundland & Labrador, Canada	CNL	9130		●		
Quebec, Canada	CQU	9133		●		Delayed to fifth grade
Moscow City, Russian Federation	RMO	643001		●	●	
South Africa (6)	ZA6	7106	●			One year later (2021)
Abu Dhabi, UAE	AAD	7842		●		Delayed to fifth grade
Dubai, UAE	ADU	7841		●		Delayed to fifth grade

Disruptions to school operations during the COVID-19 pandemic necessitated delaying data collection in a number of participating countries, resulting in the PIRLS 2021 data collection occurring over a two-year period instead of only a few months as is typical. The far right-hand column in Exhibit 2.2 indicates whether the country administered PIRLS 2021 on a delayed schedule due to disruptions to schooling because of the COVID-19 pandemic. Delays to test administration in some countries introduced some complexities for analyzing and interpreting the PIRLS 2021 data. Most countries managed to administer the PIRLS 2021

assessment to the targeted fourth grade students at the end of the school year in 2020 or 2021, according to the original plan. However, six countries and one benchmarking participant assessed the next fourth grade cohort of students one year later (listed as “one year later” in Exhibit 2.2). In 14 countries and three benchmarking entities, assessment of the fourth grade cohort was delayed until the beginning of the next school year in the fifth grade (listed as “Delayed to fifth grade”). The latter group of countries had samples of markedly older students compared to their PIRLS 2016 counterparts, and require caution in making appropriate comparisons.

More details about the national samples and test administration dates are provided in Chapter 8 in [*Methods and Procedures: PIRLS 2021 Technical Report*](#) (Almaskut et al., 2023a). [Supplement 2](#) to this User Guide describes any modifications made to the PIRLS 2021 Context Questionnaires because of these delays. A section in the International Report on [Impacts of Modifying the Assessment Schedule on Students’ Achievement](#) discusses the potential impacts that delayed testing had on the PIRLS 2021 results (Mullis et al., 2023). Chapter 12 in [*Methods and Procedures: PIRLS 2021 Technical Report*](#) discusses the potential impact in relation to the digital mode of administration (Yin et al., 2023a).

The PIRLS 2021 International Database also includes data files and documentation for the bridge assessment. The PIRLS 2021 Bridge data were collected in countries participating in digitalPIRLS. About half of the participating countries in 2021 chose to administer the digital assessment and the other half retained the traditional paper-based administration—referred to as paperPIRLS. To form a link between digitalPIRLS and paperPIRLS and maintain comparability of trend measurements over time, digitalPIRLS also administered a set of paper booklets (“bridge booklets”) consisting of paperPIRLS trend texts and items to randomly selected, equivalent groups of students from the same student populations. As such, the PIRLS 2021 Bridge data allowed for linking the two modes of administration through population-based linking (Bezirhan et al., 2023).

Supplements to the PIRLS 2021 User Guide

The User Guide is accompanied by the following supplements:

- **[Supplement 1](#): International Versions of the PIRLS 2021 Context Questionnaires**—Supplement 1 includes the international version of all context questionnaires administered in PIRLS 2021, including the Curriculum Questionnaire. This supplement serves as a reference of all items included in the context questionnaires and the associated names under which the responses are recorded in the International Database.

- **Supplement 2: National Adaptations to the PIRLS 2021 Context Questionnaires**—Supplement 2 provides details on national adaptations that were made to the national versions of the PIRLS 2021 Context Questionnaires. This supplement also includes national adaptations arising from delayed data collection in countries that administered PIRLS 2021 at the beginning of fifth grade. Users should refer to this supplement for any special national adaptations to context variables that could potentially affect the interpretation of analysis results.
- **Supplement 3: Variables Derived from the PIRLS 2021 Student, Home, School, and Teacher Context Data**—Supplement 3 describes how the derived context variables were computed to produce measures for exhibits in *PIRLS 2021 International Results in Reading* (Mullis et al., 2023).

Restricted Use Version of the PIRLS 2021 International Database

There are two versions of the PIRLS 2021 International Database. The public use version is available for immediate access from the [PIRLS 2021 International Database webpage](#) and the [IEA Data Repository](#). A number of variables were removed from the public use version in order to minimize the risk of disclosing confidential information. Exhibit 2.3 lists the variables removed from the public use version of the PIRLS 2021 International Database that are available in the restricted use version.

Exhibit 2.3: Variables Available in the Restricted Use Version of the PIRLS 2021 International Database

Variable	Description
ITBIRTHY / ITBIRTHM	Students' year and month of birth from the tracking forms
ITDATE	PIRLS 2021 testing date from the tracking forms
ITMODE_x	Mode of administration for the PIRLS assessments and context questionnaires
ITDEV	Type of device used for the digitalPIRLS assessment
ASBG02A / ASBG02B	Students' year and month of birth from the student questionnaire
ACBG01	Total school enrollment from the school questionnaire
ACBG02	School enrollment in the target grade from the school questionnaire

More details about the restricted use variables are available in the appropriate 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](#) to obtain permission

and access to the restricted use version of the PIRLS 2021 International Database. Note that the variable reporting students' age remains available in the public use version of the International Database.

Achievement Items Documentation

A number of documents related to the PIRLS 2021 achievement items are available for download along with the PIRLS 2021 International Database. They include summary information on the PIRLS 2021 items, the IRT item parameters estimated from the PIRLS 2021 item calibrations and their associated linear transformation constants, and percent correct statistics for the PIRLS 2021 achievement items.

Item Information

Item information includes characteristics of each achievement item in the PIRLS 2021 assessment. The PIRLS 2021 International Database provides item information in a single Excel file, with separate tabs for paperPIRLS ("P21 Paper"), digitalPIRLS ("P21 Digital"), and bridge ("P21Br") assessments. The file includes the following information for all items in the PIRLS 2021 assessments:

- **Item ID**, the item's unique identifier corresponding to the achievement variable name
- The item's **Text Name**, indicating the name of the reading text or text to which the item pertains
- The **Item Sequence**, indicating the sequential location of the item within the text
- An indicator for the assessment **Cycle** when the item and its text was first presented (see later section: *Item Variable Naming Convention*)
- The item's **Secure Status**, indicating whether the item and text are available for *restricted use* after the 2021 assessment, or *secure* for use in future assessments
- **Scaling Status**, indicating whether the item was included in the IRT scaling
- The **Reading Purpose** and **Comprehension Process** assessed by the item, according to the [PIRLS 2021 Reading Assessment Framework](#) (Mullis & Martin, 2019)
- The item's **Maximum Points** value
- The **Item Type**, either multiple-choice or constructed-response
- The number of **Options** for multiple-choice items

- The correct response **Key** for multiple-choice items
- A **Label** for the item

IRT Item Parameters and Scale Transformation Constants

The International Database includes Excel files with the IRT item parameters estimated for all PIRLS 2021 and PIRLS 2016 items from the concurrent item calibrations and the subsequent calibrations for digitalPIRLS items through the link with the PIRLS 2021 Bridge data. The [PIRLS 2021 International Database webpage](#) provides separate Excel files for item parameters from the PIRLS 2021 paper concurrent calibration and item parameters from the PIRLS 2021 digital and ePIRLS calibrations. These item parameters are presented in Chapter 11 of [Methods and Procedures: PIRLS 2021 Technical Report](#) (Yin et al., 2023b).

In addition to item parameters, an Excel file is provided with the linear scale transformation constants that were used to set the PIRLS 2021 achievement scores on the PIRLS trend scale. The constants in the “Paper” tab were used to set all paperPIRLS and bridge data on the PIRLS trend scale, and the constants in the “Digital” tab were used to set all digitalPIRLS data on the trend scale.

Item Percent Correct Statistics

The International Database includes percent correct statistics for all PIRLS 2021 items, including items from the PIRLS 2021 Bridge data. These statistics are available in Excel and PDF format for paperPIRLS data (“pP21”), digitalPIRLS data (“dP21”), and Bridge data (“P21Br”), respectively.

PIRLS 2021 Data Files

The PIRLS 2021 International Database includes the actual data from all instruments administered to the students and their parents, school principals, and teachers. This includes the student responses to the achievement items and the responses to the student, home, school, and teacher context questionnaires. These data files also include the achievement scores estimated for participating students, as well as context variables derived for reporting in [PIRLS 2021 International Results in Reading](#). National Research Coordinators’ responses to the PIRLS 2021 Curriculum Questionnaire also are part of the International Database and are described later in this chapter.

The next few sections describe the format and contents of the PIRLS 2021 data files, including the PIRLS 2021 Bridge data. With the exception of the curriculum data files, the PIRLS 2021 data files are provided in R (R Core Team, 2020) format (*.Rdata), SPSS (IBM Corporation, 2016) format (*.SAV) and SAS (SAS Institute, 2016) format (*.SAS7BDAT). Data

files are provided for each country that participated in PIRLS 2021 and the PIRLS 2021 Bridge. The PIRLS 2021 Bridge data files are included in all PIRLS 2021 data downloads from the [PIRLS 2021 International Database webpage](#).

The file names given to the various data file types in PIRLS 2021 are shown in Exhibit 2.4. The PIRLS 2021 data files are named with the suffix “R5” corresponding to the fifth cycle of PIRLS, and contain data for paperPIRLS and digitalPIRLS. The Bridge data files are named with the suffix “A5.” All PIRLS files are named beginning with “A,” corresponding to the fourth grade target population. The second and third characters indicate the data file type, and the fourth through sixth characters indicate the country according to the ISO alpha code (see Exhibit 2.2). For example, ASGAUTR5.Rdata is an R data file that contains Austria’s PIRLS 2021 fourth grade student context questionnaire data.

Exhibit 2.4: PIRLS 2021 Data File Names

Assessment	File Name	Description
PIRLS 2021	ACG●●●R5	School context data files
	ASA●●●R5	Student achievement data files
	ASP●●●R5	Student process data files
	ASR●●●R5	Within-country scoring reliability data files
	ASG●●●R5	Student context data files
	ASH●●●R5	Home context data files
	AST●●●R5	Student-teacher linkage files
	ATG●●●R5	Teacher context data files
PIRLS 2021 Bridge	ACG●●●A5	School context data files
	ASA●●●A5	Student achievement data files
	ASR●●●A5	Within-country scoring reliability data files
	ASG●●●A5	Student context data files
	ASH●●●A5	Home context data files
	AST●●●A5	Student-teacher linkage files
	ATG●●●A5	Teacher context data files

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

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

Student Achievement Data Files (ASA)

The PIRLS 2021 student achievement data files contain the student responses to the individual achievement items in the PIRLS 2021 assessments. The student achievement data files are best suited for performing item-level analyses. Achievement variables (plausible values) for all of the PIRLS 2021 achievement scales and sampling and weighting variables are available in the student achievement data files, as well as in the student context data files and student-teacher linkage data files (described in a later section).

As described in the [PIRLS 2021 Assessment Design](#) (Martin et al., 2019), students who participated in paperPIRLS were administered one of 18 assessment booklets, each containing one “literary” text and one “informational” text. Students who participated in digitalPIRLS were administered one of 18 “regular” booklets under the same booklet design as paperPIRLS, one of 20 “ePIRLS” booklets containing two ePIRLS tasks, or one of 45 “hybrid” booklets containing one ePIRLS task and one paper-equivalent informational text. For the PIRLS 2021 Bridge, students were administered one of eight paper booklets that contained only trend texts and items.

As PIRLS 2021 marked the transition from paper-based assessment to computer-based assessment, the PIRLS 2021 student achievement data files (ASA●●●R5) include separate sets of items by mode of administration. As such, all student achievement data files include variables for all paperPIRLS and all digitalPIRLS items. However, paperPIRLS countries have responses for the paperPIRLS items only, and digitalPIRLS countries for digitalPIRLS items only. The PIRLS 2021 Bridge student achievement data files include only variables for the paper trend items administered in the bridge booklets.

In all the booklets administered as part of PIRLS 2021, some of the items were multiple-choice and some were constructed-response. The student achievement data files contain the actual responses to the multiple-choice questions (e.g., where 1 corresponds to the first option, 2 to the second option, and so on), and the score codes assigned to the constructed-response items based on the PIRLS 2021 scoring guides. Score codes correspond to score point values 0, 1, 2, or 3, depending on the maximum points each item is worth.

Item Variable Naming Convention

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

- The first character is either “R” for regular PIRLS items or “E” for ePIRLS items.
- The second character indicates the assessment mode or type. The letter “P” indicates paperPIRLS, “E” indicates digitalPIRLS, and “0” is used for ePIRLS.

- The third character indicates the assessment cycle when the item was first presented in PIRLS. The code “1” was used for items introduced in PIRLS 2001. The items in the PIRLS 2021 assessment have “2” for items introduced in 2006, “3” for items introduced in 2011, “4” for items introduced in 2016, or “5” for new items in 2021.
- The fourth character is “1” for all PIRLS items, corresponding to the fourth grade target population.
- The fifth character is a unique letter used to identify each text. The item information Excel files show the characters used in 2021.
- The sixth and seventh characters are two-digit sequential numbers that indicate the sequential order in which the item is presented.
- The eighth character indicates the item part, and appears only when required. It is generally a letter from “A” to “F,” depending on how many parts there are to a particular item.

As an example of an item variable name, RE51T05A is the first part (part A) of a digitalPIRLS item introduced in 2021, that is the fifth item in the passage *Ostrich and the Hat*.

Item Response Codes

A series of conventions were adopted to code the item responses included in the PIRLS 2021 student achievement data files. The value assigned to each item response depends on the item format.

For multiple-choice items, numerical values 1, 2, 3, 4, etc., are used to correspond to the response options A, B, C, D, etc., respectively. For these items, the correct response key is included in the item information files and as part of the variable label in the achievement codebook files (described in a later section). R, SPSS, and SAS programs are included as part of the PIRLS 2021 International Database to derive correctness scores (score points) for these items based on their item response codes and key (see Chapter 3).

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

Derived Items

For some items, 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. These multi-part items can be worth 1, 2, or 3 score points, according to their scoring guide. A list of all derived items in the PIRLS

2021 assessments and the rules for awarding item response codes is provided in Appendix 9G of [*Methods and Procedures: PIRLS 2021 Technical Report*](#) (Bristol et al., 2023).

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

For example, derived item variable RE51N02 contains the combined score for its five parts labeled RE51N02A through RE51N02E. This derived item response was assigned 2 score points (complete comprehension) if all five parts were answered correctly, 1 score point (partial comprehension) if four parts were answered correctly, and 0 points otherwise.

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 (R: NA; SPSS: sysmis; SAS: .A)

Special codes were given to items that were *Not Administered* to distinguish these cases from data that were missing due to student non-response. In general, the Not Administered code was used when an item was not administered, either by design—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 PIRLS 2021 received one of 18 available test booklets for paperPIRLS, one of 83 booklets for digitalPIRLS, and one of eight test booklets for Bridge. All variables corresponding to items that were not present in a student’s assigned booklet 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 relevant 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 (R: 9; SPSS: 9; SAS: .)

Omitted response codes 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 (R: 6; SPSS: 6; SAS: .R)

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.

Modifications to Achievement Item Data

Some modifications were applied to the achievement item data based on quality checks and psychometric analyses conducted for the PIRLS 2021 psychometric scaling. For example, a 2-point item may be collapsed into a 1-point item, or the data for a particular country may be deleted. In a few rare cases, achievement item variables were excluded from the International Database because the item was deemed not to provide accurate and reliable measurement of reading achievement. The analyses conducted and the criteria for making these changes is described in Chapter 9 of [Methods and Procedures: PIRLS 2021 Technical Report](#) (Bristol et al., 2023). [Appendix 9F](#) includes a list of all modifications reflected in the PIRLS 2021 International Database files.

PIRLS Achievement Variables

PIRLS 2021 achievement scales were produced for overall reading as well as for the reading subdomains defined in the [PIRLS 2021 Reading Assessment Framework](#) (Mullis & Martin, 2019): the two purposes for reading (“Literary” and “Informational”) and the two overarching comprehension processes (“Retrieving and Straightforward Inferencing” and “Interpreting, Integrating, and Evaluating”).

The PIRLS 2021 achievement scales are listed in Exhibit 2.5. A total of five achievement scales were produced for both PIRLS 2021 and PIRLS 2021 Bridge.

Exhibit 2.5: PIRLS 2021 Achievement Scales

Achievement Scale		Description
Overall	REA	Overall Reading
Purposes for Reading	LIT	Literary Experience
	INF	Acquire and Use Information
Comprehension Processes	RSI	Retrieving and Straightforward Inferencing
	IIE	Interpreting, Integrating, and Evaluating

For each achievement scale, the PIRLS 2021 International Database provides five imputations of achievement for each student given their item responses and context data. The five imputations are known as “plausible values,” and the variability between them reflects the uncertainty inherent in the data and the estimation. A detailed description of the PIRLS 2021 scaling approach and how these achievement scales were created is available in Chapter 11 of [Methods and Procedures: PIRLS 2021 Technical Report](#) (Yin et al., 2023b).

The plausible values for any given scale are the best available measures of student achievement on that scale in the PIRLS 2021 International Database and should be used as the outcome measure in any study of student achievement. It is important to note that these plausible values are not suitable measures of individual student achievement, as explained in Chapter 10 of [Methods and Procedures: PIRLS 2021 Technical Report](#) (Bezirhan et al., 2023). Plausible values can be analyzed readily using the IEA IDB Analyzer (see Chapter 1 of this User Guide).

The achievement (plausible value) variable names are based on an 8-character alphanumeric code, which adheres to the following rules:

- The first character is always “A” for the fourth grade
- The second character is always “S” to indicate it is a student achievement (score) variable
- The third character is “R” to indicate a reading scale, whether it is overall reading, reading purposes, or comprehension processes
- The fourth through sixth characters are a three-character code describing the achievement scale, as shown in Exhibit 2.5
- The seventh and eighth characters are a two-digit number indicating the plausible value—01, 02, 03, 04, or 05

For example, ASRINF01 is the first plausible value on the fourth grade Informational reading achievement subscale.

PIRLS International Benchmarks of Reading Achievement

To help users of the PIRLS 2021 International Database, as well as readers of [PIRLS 2021 International Results in Reading](#) (Mullis et al., 2023), understand what performance on the overall reading achievement scale signifies in terms of the reading skills and strategies demonstrated by fourth grade students, PIRLS identified four points on the overall reading scale to serve as International Benchmarks. As shown in Exhibit 2.6, the PIRLS International Benchmark scores are 625, 550, 475, and 400, which correspond to the Advanced International Benchmark, the High International Benchmark, the Intermediate International Benchmark, and the Low International Benchmark, respectively.

Exhibit 2.6: PIRLS 2021 International Benchmarks of Reading Achievement

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

PIRLS 2021 used a technique known as *scale anchoring* to summarize and describe student achievement at these four points on the scale (see Chapter 14 of [Methods and Procedures: PIRLS 2021 Technical Report](#); Wry et al., 2023). The [PIRLS 2021 International Results in Reading](#) report presents the results of this scale anchoring, and reports the percentage of students in each country reaching each of the PIRLS International Benchmarks.

The PIRLS 2021 International Database contains a set of variables indicating which International Benchmark the students have reached. There are five benchmark variables—one for each plausible value of the overall reading scale. The International Benchmark variables follow the achievement score variable naming convention where the fourth through sixth positions have the letters “IBM.” Thus, **ASRIBM01-05** are the five benchmark variables for fourth grade overall reading in PIRLS. The codes defined for the benchmark variables are described in Exhibit 2.7.

Exhibit 2.7: PIRLS 2021 International Benchmark Variable Codes

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

Non-Response Indicator Variables

The PIRLS 2021 student achievement data files include several variables indicating whether students had missing achievement item responses (Omitted or Not Reached) for subgroups of items according to item type and PIRLS reading subdomain. These non-response indicator variables were utilized in generating PIRLS plausible values of student achievement (the imputations described in the previous *PIRLS Achievement Variables* section). The non-response indicator variables use the numerical value 1 if the student answered all items in the subset, or 0 if the student had at least one missing item response in the subset. The non-response indicator variable names are based on an eight-character scheme as follows:

- The first character is either “R” for groups of regular PIRLS items or “E” for groups of ePIRLS items.
- The second character indicates the assessment mode or type of the group of items. The letter “P” indicates paperPIRLS items, “E” indicates digitalPIRLS items, and “0” is used for ePIRLS items.
- The third and fourth characters indicate the item type of the group of items. “MC” is used for groups of multiple-choice items, and “CR” is used for groups of constructed-response items.
- The fifth through seventh characters indicate the PIRLS subdomain for the group of items according to the [PIRLS 2021 Reading Assessment Framework](#) (Mullis & Martin, 2019). The three characters correspond to the subdomains according to the achievement scales listed in Exhibit 2.5. Some students in countries that administered the digital assessment received “ePIRLS” booklets or “hybrid” booklets with items only for the Informational purpose. Therefore, indicator variables also are included according to booklet part for these students, “PT1” for part 1 and “PT2” for part 2.

For more information about how these variables were used in achievement scaling, see Chapter 11 of [Methods and Procedures: PIRLS 2021 Technical Report](#) (Yin et al., 2023b).

Student Process Data Files (ASP)

The PIRLS 2021 student process data files contain variables associated with students' navigation in the digital PIRLS 2021 assessment. Data for derived process variables are available only for countries that participated in the digital assessment. It is recommended that users of the process data files have a general understanding of how the PIRLS 2021 digital assessment functions. Video examples of digitalPIRLS passages and ePIRLS tasks can be viewed in [PIRLS 2021 International Results in Reading](#). ePIRLS tasks can be tried out on [IEA's website](#).

The PIRLS 2021 process data files include three types of process variables associated with the achievement items:

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

Variables for total time on item and number of item visits are only provided for digitalPIRLS items. Time on first item visit is provided for both digitalPIRLS and ePIRLS items. Codebooks published with the database list all derived variables included in the data files and describe codes for missing values in the process data.

The process variables for each item are named according to the item variable naming convention (described in the previous *Item Variable Naming Convention* section for the student achievement data files), followed by the suffix “_S” for total time, “_R” for time on first item visit, and “_F” for the number of item visits. There are occasional items that include multiple parts recorded in the achievement data files. In these instances, the process data variable names consist of the first seven common characters of the corresponding item names. For example, RE51R17A and RE51R17B are two items with separate responses but share the same stem and are presented together. The process data variable names associated with that screen are RE51R17_S for the total response time, RE51R17_R for the time on first visit, and RE51R17_F for the frequency of visits for the two item parts together.

The student process data variables were derived from the raw student response data files generated through the PIRLS 2021 digital assessment. RM Results ([RMresults.com](#)) was responsible for the development of the PIRLS 2021 assessment software, including the production of raw data files. The raw RM response data files contain system-computed duration values for each visit on an item, rounded down to the nearest whole second. The system-computed duration values exclude page-load time. The item-level timing values in the student process data files are not fully representative of all student time spent during the test

session. RM's duration values may be affected by any technical glitches that occurred during the test session, such as server timeouts or device malfunctions. The data in the student process data files were cleaned for inconsistencies, outliers, and any known technical issues in the production of the raw RM data files.

In addition to the derived process variables, the student process data files include basic identifying and tracking information for each student (see later sections on *Identification Variables* and *Tracking Variables*), plus additional tracking and indicator variables about student participation and cleaning of the process variables, described below.

ITPARTPTx

ITPARTPTx indicates whether the student has valid item responses (non-missing) for each of the two parts of the assessment, where “x” can take the values 1 (part 1) or 2 (part 2). For these variables, a value of 1 indicates the student has valid item responses in the part, and 2 indicates no valid item responses in the part. Process data is not available for a part if no valid item responses are available.

FLAGPTx

FLAGPTx indicates whether the student had process data removed for each of the two parts of the assessment, where “x” can take the values 1 (part 1) or 2 (part 2). For these variables, a value of 1 indicates the student had process data deleted for the part due to a known technical issue, and a value of 2 indicates the student had process data deleted for the part due to an outlier. A value of 0 indicates that no issues were detected for the part.

Within-Country Scoring Reliability Data Files (ASR)

The PIRLS 2021 within-country scoring reliability data files contain data that can be used to investigate the reliability of the PIRLS item scoring for human-scored constructed-response items. The scoring reliability data files contain one record for each student whose responses to constructed-response items, in whole or in part, were double scored during the within-country scoring reliability exercise (see Chapter 9 in [Methods and Procedures: PIRLS 2021 Technical Report](#); Bristol et al., 2023). For each constructed-response item requiring human scoring, the following three variables are included in the scoring reliability data files:

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

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

It should be noted that the Second Score data were used only to evaluate within-country scoring reliability and were not used in computing the achievement scores included in the International Database and presented in [*PIRLS 2021 International Results in Reading*](#).

Scoring Reliability Variable Naming Convention

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

- The **Original Score** variable follows the item variable naming convention described earlier. The second character has the letter “P” for a paperPIRLS item, the letter “E” for a digitalPIRLS item, or the number “0” for an ePIRLS item.
- The **Second Score** variable has the letter “R” added to the Original Score variable after the second character (e.g., RPR51D01).
- The **Score Agreement** variable has the letter “I” added to the Original Score variable after the second character (e.g., RPI51D01).

Scoring Reliability Codes

The values contained in both the Original Score and Second Score variables are the scores assigned using the PIRLS 2021 scoring guides. The Score Agreement variable may have one of two values, indicating either agreement or disagreement between the score codes assigned by the two scorers, as described in Exhibit 2.8.

Exhibit 2.8: PIRLS 2021 Score Agreement Variable Codes

Code	Description
1	Identical codes (Agreement)
0	Different codes (Disagreement)

Context Data Files

This section describes the PIRLS 2021 context data files and the conventions for naming the various files and variables and coding the data. There are six types of PIRLS 2021 context data files: the first four context data files correspond to the four types of context questionnaires administered in PIRLS 2021 (student, home, school, and teacher); the fifth data file serves to link the student and teacher context data; and the sixth data file

corresponds to the PIRLS 2021 Curriculum Questionnaire administered to the National Research Coordinators of each participating country. The Curriculum Data are provided separately from the other data files in the PIRLS 2021 International Database. Comprising mostly descriptive, narrative information about the national education systems, these data are provided in Excel format (see later section).

The four context data files for student, home, school, and teacher questionnaires contain the responses to the questions asked in their respective context questionnaires, plus some additional derived variables used for reporting (see [Supplement 3](#) to this User Guide), as well as variables for the PIRLS 2021 Context Questionnaire scales (see Chapter 15 in [Methods and Procedures: PIRLS 2021 Technical Report](#); Yin & Reynolds, 2023).

Student Context Data Files (ASG)

All students who participated in PIRLS 2021 were administered a context questionnaire with questions related to their home context, school experiences, and attitudes toward reading. Regardless of whether they were participating in paperPIRLS, digitalPIRLS, or bridge, students were administered a questionnaire at the end of their testing session. The student context data files contain students' responses to all of these questions. They also contain students' reading achievement variables (plausible values) to facilitate analyses of relationships between student characteristics and achievement.

The student context data files also contain a number of identification variables, tracking variables, sampling and weighting variables, and derived variables that were used for producing exhibits in [PIRLS 2021 International Results in Reading](#). These variables are described later in this chapter (see later section on *Structure and Design Variables*).

Home Context Data Files (ASH)

PIRLS 2021 included a home questionnaire (also called the Early Learning Survey), which was completed by the students' parents or guardians. It asked questions about preparations for primary schooling, including attendance in preschool and literacy activities in the home before the child began school, such as reading books, singing songs, or writing letters or words. Parents answered questions about home resources in addition to information about their highest level of education and their employment status. Analyzing data from the home questionnaire requires that the home context data files be merged with the student context data files using the country and student identification variables (see later section on *Structure and Design Variables*). Details of the merging procedure with the IEA IDB Analyzer are described in Chapter 1 of this User Guide.

School Context Data Files (ACG)

The school context data files contain principals' responses to the questions in the PIRLS 2021 School Questionnaire. Although school-level analyses where the schools are the units of analysis can be performed, it is preferable to analyze school-level variables as attributes of students. To perform student-level analyses with school data, the school context data files must be merged with the student context data files using the country and school identification variables (see later section on *Structure and Design Variables*). Details of the merging procedure with the IEA IDB Analyzer are described in Chapter 1 of this User Guide.

The COVID-19 pandemic required some adaptations to the school questionnaire in countries that delayed PIRLS 2021 data collection until the beginning of fifth grade. See [Supplement 2](#) to this User Guide for details.

Teacher Context Data Files (ATG)

The teachers of the students that were sampled in PIRLS 2021 were administered a questionnaire with questions pertaining to their teaching context, attitudes, and teaching practices in the classes of the sampled students. Each teacher was asked to respond to a questionnaire for each class taught that contained sampled students. The teacher context data files contain one record for each of the classes taught by a teacher. If a teacher taught more than one sampled class, they were expected to complete multiple questionnaires, responding only once to general context questions and multiple, separate class-specific questions for each class they taught. In some cases, although a teacher was expected to respond to more than one questionnaire, responses to only one were obtained. In these cases, there were as many records entered in the teacher context data file as sampled classes taught by the teacher, and the general context information from the completed questionnaire was entered into these teacher records.

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

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

representative sample of students. The teacher data, therefore, should be thought of as attributes of the students to which they are linked, and should be analyzed only in conjunction with the student-teacher linkage data files. Chapter 1 of this User Guide describes student-level analyses combining the teacher data and the student-teacher linkage data files with the IEA IDB Analyzer software.

As discussed in Chapter 1, in the *Conducting Analyses with PIRLS Teacher Context Data* section, analyses using PIRLS 2021 teacher data should be interpreted with caution because of the PIRLS 2021 data collection coinciding with the COVID-19 pandemic. The influence that particular teachers had on students' achievement during disruptions to schooling may have been impacted, and situations varied greatly within and across countries. [Supplement 2](#) to this User Guide describes these issues as well as any adaptations made to the questionnaires to accommodate changes to countries' data collection schedules.

Student-Teacher Linkage Data Files (AST)

The PIRLS 2021 student-teacher linkage data files contain information required to link the student and teacher data files. The student-teacher linkage data files contain one entry per student-teacher linkage combination in the data. For instance, if three teachers are linked to a student, the file has three entries corresponding to that student. The sole purpose of the student-teacher linkage data files is to link teacher-level data with student-level data in order to perform appropriate student-level analyses where teacher characteristics are considered as attributes of the students. The student-teacher linkage data files also include sampling and weighting information and achievement scores to facilitate the analyses of teacher data (see later section on *Structure and Design Variables*).

Curriculum Data

The PIRLS 2021 curriculum data contain the responses provided by the PIRLS National Research Coordinators of the participating countries to the PIRLS 2021 Curriculum Questionnaire. The data are available to download in an Excel file from the [PIRLS 2021 International Database webpage](#).

Context Variable Naming Convention

The context variable naming convention for the variables in the student, home, school, and teacher context data files uses a 7- or 8-character string. The following rules are applied in naming the context variables:

- The first character is always “A” for fourth grade data.

- The second character indicates the type of respondent. The letter “C” identifies data from the school principals, the letter “T” for teacher data, and the letter “S” for student and parent data.
- The third character is used to indicate the source of the data. The letter “B” is used for all context variables reporting responses to the context questionnaires. The letter “D” is used for variables derived from responses in the context questionnaires. In addition, the letter “B” is used for the Rasch scores to context questionnaire scales derived from questionnaire data, and the letter “D” is used for the index variables constructed from these context questionnaire scale Rasch scores.⁴
- The fourth character is used to indicate the subject or topic to which a context question refers. The following letters are used:
 - G—General questions (not subject specific)
 - H—Home questionnaire questions
 - R—Questions related to reading
- The fifth through eighth characters of all context variables represent the sequential numbering of the questions as presented in their respective questionnaires.

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

- GEN—General questions (not subject specific)
- READ—Questions related to reading instruction
- COVID—Questions related the COVID-19 pandemic

The remaining characters in the curriculum data variable naming convention refer to the question location, as shown in [Supplement 1](#) to this User Guide, and summarized in Exhibit 2.9.

⁴ The context questionnaire scales are described in Chapter 15 of [Methods and Procedures: PIRLS 2021 Technical Report](#) (Yin & Reynolds, 2023).

Exhibit 2.9: PIRLS 2021 Context Variable and Question Location Naming Convention

Questionnaire	Question Location Name	Context Variable Name	Description
Student Questionnaire	SQG-●●●	ASG●●●	General questions
	SQR-●●●	ASR●●●	Reading questions
Home Questionnaire	HQ-●●●	ASH●●●	All questions
School Questionnaire	ScQ-●●●	ACG●●●	All questions
Teacher Questionnaire	TQG-●●●	ATG●●●	General questions
	TQR-●●●	ATR●●●	Reading questions
Curriculum Questionnaire	CQG-●●●	GEN●●●	General questions
	CQR-●●●	READ●●●	Reading questions
	CQC-●●●	COVID●●●	COVID-19 questions

Context Question Location Convention

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

Context Variable Response Codes

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

Codes for Missing Values

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.

High levels of missing data could bias results. *PIRLS 2021 International Results in Reading* (Mullis et al., 2023) annotates results when missing data rates exceed 15% of students. Results are not reported for a country if data are available for fewer than 40% of students, and countries with data available for fewer than 50% are reported at the bottom of the results exhibits. Similarly, achievement estimates can be unreliable if based on small groups of students. Therefore, achievement results are not reported for any context variable categories containing fewer than 2.5% of students.

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

Special codes were given to items that were *Not Administered* to distinguish these cases from data that were missing due to non-response. In general, the Not Administered code was used when an entire questionnaire was not completed. The Not Administered code also was used in the following cases:

- Question was removed: Variables corresponding to questions in the student, home, school, or teacher context questionnaires that were considered not appropriate in some countries were not included in the national versions of the questionnaires. These questions were coded as Not Administered.
- Question left out or misprinted: When a particular question (or a whole page) was misprinted, or otherwise not available to the respondent, the corresponding variables were coded as Not Administered.
- Question mistranslated or not internationally comparable: In some cases, questions in the international version of the questionnaires were mistranslated or modified to fit the national context. Whenever possible, modified questions were recoded to match the international version as closely as possible. When this was not possible, modified questions were coded as Not Administered.

[Supplement 2](#) of this User Guide reports all instances of questions removed or questions modified such that the data were not internationally comparable.

Omitted Response Codes (R: 9, 99, 999, ...; SPSS: 9, 99, 999, ...; SAS: .)

Omitted response codes were used for questions that a student, parent, teacher, or school principal should have answered but did not. The length of the Omitted response code given to a variable in the R and SPSS data files depends on the number of characters needed to represent the variable. In all cases, the space necessary to represent the variable is filled with 9's. No distinction is made between items left blank and items with invalid answers, such as

checking two or more response options in a categorical question, or unreadable or uninterpretable responses to open-ended questions. In a few cases, data received from a country in an invalid or inconsistent manner also were coded as Omitted.

Not Applicable Response Codes (R: 6, 96, 996, ...; SPSS: 6, 96, 996, ...; SAS: .B)

Not Applicable response codes were used for the context questionnaire items for which responses were dependent upon a filter question. Generally, a “No” response to a filter question leads to any follow-up questions being coded as Not Applicable because there were no appropriate responses to these follow-up questions. For example, in the school questionnaire, if a principal answered “No” to being asked if the school had a school library (ACBG07A), all items corresponding to having a library, such as the number of books in the library (ACBG07B), were coded as Not Applicable.

Context Questionnaire Scales and Derived Variables

In the PIRLS 2021 Context Questionnaires, there are instances where several questions are asked about various aspects of a single construct. In these cases, responses to the individual items were combined to create a score, using Rasch scaling, which provided a more comprehensive interpretation of the construct of interest than the individual variables could on their own. These context questionnaire scales also were categorized, usually into three groups, to create an index. The context questionnaire scales and their indices are included in the PIRLS 2021 International Database context data files and described in Chapter 15 of [*Methods and Procedures: PIRLS 2021 Technical Report*](#) (Yin & Reynolds, 2023).

Additional variables were derived from responses to multiple questions to provide more pertinent information for analysis and reporting. Parents’ education is an example where responses from both parents were combined into a single variable to report a single educational level. [Supplement 3](#) to the User Guide provides a description of the derived variables included in the PIRLS 2021 International Database.

Sampling and Weighting Variables

Several sampling and weighting variables are included in the PIRLS 2021 data files—they are listed and described in Exhibit 2.10. Exhibit 2.11 indicates the location of the various sampling and weighting variables among the different types of data files in the PIRLS 2021 International Database. It is important to note that the teacher context data files, home context data files, and scoring reliability data files do not have any sampling and weighting variables.

Exhibit 2.10: PIRLS 2021 Sampling and Weighting Variables

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

As a general rule, TOTWGT—the overall student sampling weight—is the preferred sampling weight to use when analyzing student-level data. Although TOTWGT has desirable properties, it may have drawbacks for some analyses. Because TOTWGT sums to equal the student population size in each country, analyses using TOTWGT that combine countries will have proportionately more students from larger countries and fewer from smaller countries, which may not be desirable for some purposes. For cross-country analyses in which countries should be treated equally, PIRLS provides SENWGT, a transformation of TOTWGT, that results in a weighted sample size of 500 in each country. Additionally, because TOTWGT inflates sample sizes to estimate the population size, software systems that use the actual sample size to perform significance tests may give misleading results for analyses weighted by TOTWGT. HOUWGT, another transformation of TOTWGT, ensures that the weighted sample corresponds to the actual sample size in each country.

Exhibit 2.11: Locations of Sampling and Weighting Variables in PIRLS 2021 Data Files

Variable	Data File			
	ASA	ASG	AST	ACG
TOTWGT	•	•		
SENWGT	•	•		
HOUWGT	•	•		
SCHWGT				•
STOTWGTU				•
TCHWGT			•	
JKZONE	•	•	•	
JKREP	•	•	•	
JKCZONE				•
JKCREP				•
WGTFAC1	•	•		•
WGTADJ1	•	•		•
WGTFAC2	•	•		
WGTADJ2	•	•		
WGTFAC3	•	•		
WGTADJ3	•	•		

Exhibit 2.4 describes the file name conventions.

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

The weight variable TCHWGT is specifically designed for using teacher context data in student-level analyses and is based on TOTWGT. This teacher weight is located in the student-teacher linkage file (AST), not in the actual teacher context data file (ATG). Analyses with teacher data will be weighted properly by merging the teacher files with the student-teacher linkage files and using TCHWGT.

All weighting variables beginning with the letters “WGT” provide insight into the multi-stage sampling and weighting methodology applied to the PIRLS data. These weights are described in detail in Chapter 3 of [Methods and Procedures: PIRLS 2021 Technical Report](#) (Almaskut et al., 2023b).

The sampling variables beginning with the letters “JK” are used to compute standard errors based on the jackknife repeated replication methodology (see Chapter 13 in [Methods and Procedures: PIRLS 2021 Technical Report](#); Foy & Almaskut, 2023).

A word of caution to prospective users of the PIRLS 2021 International Database is in order. As tempting as it seems to merge all file types into one all-encompassing data file, analyses with student, home, and school context data may not produce correct results. This problem arises from the student-teacher linkage and how it apportions student weights (TOTWGT) to the student-teacher linkage records in the teacher weights (TCHWGT). Only the use of TCHWGT will produce correct results in this context, even for analyses that do not include teacher data.

Structure and Design Variables

Besides the variables used to store responses to the context questionnaires and achievement booklets, the PIRLS 2021 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 PIRLS 2021 data files, several identification variables are included that provide information to identify countries, students, teachers, or schools. These variables also are used to link, or merge, cases between the different data file types. The identification variables have the prefix “ID” and are described below.

IDCNTRY

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

IDSCHOOL

IDSCHOOL is a four-digit identification code that uniquely identifies the participating schools within each country. The school codes are generated and assigned specifically for PIRLS 2021 and are not meant to represent actual school identifiers in the participating countries. They are not unique across countries. Schools across countries can be identified uniquely only with the IDCNTRY and IDSCHOOL combination of linking variables.

IDCLASS

IDCLASS is a six-digit identification code that uniquely identifies the sampled classrooms within a country. The variable IDCLASS has a hierarchical structure and is formed by concatenating the IDSCHOOL variable and a two-digit sequential number identifying the

sampled classrooms within a school. Classrooms can be identified uniquely in the database by the combination of IDCNTRY and IDCLASS as linking variables.

IDSTUD

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

IDTEACH

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

IDLINK

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

IDTEALIN

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

IDGRADE

IDGRADE identifies the target grade of the participating students. In PIRLS 2021, the usual value is “4.” Some countries and benchmarking participants can have the values “5” or “6.” Countries that delayed assessment until the beginning of fifth grade have the value “5.”

IDBOOK

IDBOOK identifies the specific assessment booklet that was administered to each student. In the PIRLS 2021 data, the booklets are given a numerical value from “1” through “18” for paperPIRLS countries and from “1” through “83” for digitalPIRLS countries. In the PIRLS 2021 Bridge data, books have a value from “1” through “8.”

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

Exhibit 2.12: Location of Identification Variables in PIRLS 2021 Data Files

Variable	Data File						
	ASA	ASP	ASG	AST	ATG	ACG	ASH
IDCNTRY	•	•	•	•	•	•	•
IDSCHOOL	•	•	•	•	•	•	•
IDCLASS	•	•	•	•			•
IDSTUD	•	•	•	•			•
IDTEACH				•	•		
IDLINK				•	•		
IDTEALIN				•	•		
IDGRADE	•	•	•	•	•	•	•
IDBOOK	•	•	•	•			

Shading indicates combinations of variables that allow for uniquely identifying records across data files.

Exhibit 2.4 describes the file name conventions.

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

Tracking Variables

Tracking variables are used to store information about students, teachers, and schools provided by the survey tracking forms.⁵ Tracking variables also include information about the test administration. Many of these variables have the prefix “IT.” All tracking variables are included in the student achievement data files (ASA) and the student context data files (ASG). ITSEX, ITLANG, and WAVE are included in the student process data files (ASP). ITLANG and ITMODE are included in the home, school, and teacher context files.

⁵ Survey tracking forms are lists of students, teachers, and schools used for sampling and administrative purposes.

ITSEX

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

ITBIRTHM and ITBIRTHY

ITBIRTHM and ITBIRTHY indicates 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 PIRLS 2021 International Database.

ITDATE

ITDATE indicates the testing date for each student. This variable is available only in the restricted use version of the PIRLS 2021 International Database.

ASDAGE

ASDAGE indicates the age of each student at the time of testing. It is derived from ITBIRTHM, ITBIRTHY, and ITDATE. While the source variables are only included in the restricted use version of the PIRLS 2021 International Database, ASDAGE is available in the public use version.

WAVE

WAVE was created especially for PIRLS 2021 to indicate the data collection period in which the students were administered the PIRLS assessment. Cases with the value “1” on this variable took the assessment according to the original schedule, in October–December 2020 for Southern Hemisphere countries or in February–July 2021 for Northern Hemisphere countries. The value “2” is used for testing in August–December 2021, in which Southern Hemisphere countries tested “one year later” and Northern Hemisphere countries “delayed to fifth grade” (see Exhibit 2.2). The value “3” is used for countries that tested in April–July 2022 (Northern Hemisphere countries that tested “one year later”).

ITLANG_x

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

ITMODE_x

ITMODE_x indicates the mode of administration (paper or digital) for the PIRLS assessments and context questionnaires, where “x” can take the values “SA,” “SH,” “SQ,” “CQ,” and “TQ” to denote the various PIRLS survey instruments. This variable is available only in the restricted use version of the PIRLS 2021 International Database.

ITDEV

ITDEV indicates the type of digital device used for the digitalPIRLS assessment. This variable is available only in the restricted use version of the PIRLS 2021 International Database.

ITADMINI

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

Codebooks

All information related to the structure of the PIRLS 2021 data files, as well as the source, format, descriptive labels, and response option codes for all variables, is contained in codebook files. Codebooks can be downloaded as Excel files for PIRLS 2021 (paperPIRLS and digitalPIRLS countries) and for the PIRLS 2021 Bridge data.

In the codebook Excel files, there is a tab for each appropriate data file type in the PIRLS 2021 International Database (see Exhibit 2.4 for the naming convention). These tabs describe the contents and structure of the individual PIRLS 2021 data files. Important codebook fields include LABEL, which contains extended textual information for all variables, QUESTION LOCATION, which provides the location of questions and achievement items within their respective survey instruments, and VALUE SCHEME DETAILED, which lists the acceptable responses allowed for each variable.

Users should note that the codebooks include the variables that are available only in the PIRLS 2021 restricted use data 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 PIRLS 2021 data files. There are three basic types of data almanacs: achievement data almanacs for the achievement items, context data almanacs for the context variables, and enjoyment almanacs for the variables indicating how much the students liked reading the PIRLS texts. On the [PIRLS 2021 International Database webpage](#), users can download almanac files for achievement, context, and enjoyment 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 PIRLS 2021 and PIRLS 2021

Bridge assessments. The achievement data almanac files available in the International Database are listed in Exhibit 2.13. Within PIRLS 2021, separate almanacs are provided for paperPIRLS items (“pP21”) and digitalPIRLS items (“dP21”). Then, separate files are also provided for Literary items (“LIT”) and Informational items (“INF”).

Exhibit 2.13: PIRLS 2021 Achievement Data Almanacs

Assessment		File Name	Description
PIRLS 2021	Digital Items	dP21_INF_ItemAlmanac	Almanac for digitalPIRLS (and ePIRLS) Informational items
		dP21_INF_TrendItemAlmanac	Trend almanac for ePIRLS items
		dP21_LIT_ItemAlmanac	Almanac for digitalPIRLS Literary items
	Paper Items	pP21_INF_ItemAlmanac	Almanac for paperPIRLS Informational items
		pP21_INF_TrendItemAlmanac	Trend almanac for paperPIRLS Informational items
		pP21_LIT_ItemAlmanac	Almanac for paperPIRLS Literary items
PIRLS 2021 Bridge	Paper Items	pP21_LIT_TrendItemAlmanac	Trend almanac for paperPIRLS Literary items
		pP21_INF_ItemAlmanac	Almanac for paperPIRLS Informational items
		pP21_INF_TrendItemAlmanac	Trend almanac for paperPIRLS Informational items
		pP21_LIT_ItemAlmanac	Almanac for paperPIRLS Literary items
		pP21_LIT_TrendItemAlmanac	Trend almanac for paperPIRLS Literary items

For each item, the achievement data **Item Almanacs** display its classification in the reading purpose and comprehension processes, the text or task (assessment block) to which it belongs, a brief description of the item, its variable name, whether it is a multiple-choice or constructed-response item, its point value, and the correct response key if it is a multiple-choice item.

The **Trend Item Almanacs** provide summary statistics for achievement items used in both the 2016 and 2021 assessments. There is no trend item almanac for digitalPIRLS Literary items, because only ePIRLS Informational tasks were administered in 2016 as part of the ePIRLS assessment.

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

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

- N: The number of students to whom the item was administered

- **DIFF:** The percentage of students that responded correctly to the item. In the case of constructed-response items worth 2 or 3 points, the percentages are weighted across each of the possible score point categories. Omitted and Not Reached responses are excluded from this calculation.
- **A, B, C, D, etc.:** The percentage of students choosing each one of the response options for a multiple-choice item
- **Scoring Guide Codes (e.g., 0, 1, 2, 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.

Enjoyment Almanacs

The enjoyment almanacs provide weighted summary statistics for each participating country on how much the students enjoyed reading each text in PIRLS 2021. At the end of each text and item set in the assessment, students were asked to indicate how much they liked reading it, with four options “I liked it a lot,” “I liked it a little,” “I didn’t like it very much,” and “I didn’t like it at all.”

There are two enjoyment almanacs in the International Database: one for PIRLS 2021 (“P21”), including both paperPIRLS and digitalPIRLS data, and one for PIRLS 2021 Bridge (“P21Br”). The enjoyment almanacs report, for each text, the number of students to whom the text was administered along with the percentage of students responding in each category together with their average reading achievement. The enjoyment almanacs also display the international averages for each text, with each country weighted equally. The benchmark participants, listed below the international averages, are not included in the calculation of international averages.

Context Data Almanacs

Context data almanac files contain weighted summary statistics for each participating country on each variable in the student, home, school, and teacher context questionnaires, including the context questionnaire scales and their indices and the derived variables based on context variables. Among the statistics reported is mean reading achievement by response category. The context data almanacs also display for each variable the question as it was asked, its

location in the corresponding questionnaire, and its variable name in the data files. The context data almanac files available in the International Database are listed in Exhibit 2.14.

Exhibit 2.14: PIRLS 2021 Context Data Almanacs

Assessment	File Name
PIRLS 2021	P21_StudentAlmanac
	P21_HomeAlmanac
	P21_SchoolAlmanac
	P21_TeacherAlmanac
PIRLS 2021 Bridge	P21Br_StudentAlmanac
	P21Br_HomeAlmanac
	P21Br_SchoolAlmanac
	P21Br_TeacherAlmanac

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

The PIRLS 2021 context almanacs include paperPIRLS and digitalPIRLS countries together since the same context questionnaires were administered regardless of mode of administration. There is a separate set of context almanacs for the PIRLS 2021 Bridge data.

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

- The sample size (number of students, parents, schools, or teachers included in the sample)
- The number of valid cases (number of students, parents, schools, or teachers for whom valid data were obtained)
- The weighted percentages of students corresponding to each valid response option (percentages based only on the students with valid data, as well as Not Applicable codes when used)
- The weighted percentages of students for whom none of the valid response options were selected, coded as Not Administered or Omitted (percentages based on all students)

- The weighted mean achievement values of students corresponding to each valid response option, as well as the Not Administered and Omitted codes
- In cases where a variable can be coded as Not Applicable because of an earlier filter question, the weighted percentage of students for whom the variable is coded as Not Applicable also is displayed, along with the corresponding weighted mean achievement

The almanac display for continuous variables includes the following:

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

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

Special R, SPSS, and SAS Programs

The PIRLS 2021 International Database includes R, SPSS, and SAS syntax files which will recode the responses to individual items from the achievement data files to their appropriate score levels. These programs are described in Chapter 3 of this User Guide.

References

- Almaskut, A., LaRoche, S., & Foy, P. (2023a). Sample implementation in PIRLS 2021. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 8.1-8.154). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb2743>
- Almaskut, A., LaRoche, S., & Foy, P. (2023b). Sample design in PIRLS 2021. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 3.1-3.32). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb9560>
- Bezirhan, U., Foy, P., & von Davier, M. (2023). PIRLS 2021 achievement scaling methodology: Item Response Theory, population models, and linking across modes. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 10.1-10.19). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb3131>
- Bristol, J., Mullis, I. V. S., Fishbein, B., & Foy, P. (2023). Reviewing the PIRLS 2021 achievement item statistics. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 9.1-9.45). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb5892>
- Foy, P., & Almaskut, A. (2023). Estimating standard errors in the PIRLS 2021 results. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 13.1-13.31). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb1576>
- 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). (2023). IEA IDB Analyzer (Version 5.0) [Computer software]. Hamburg, Germany: IEA Hamburg. Retrieved from <https://www.iea.nl/data>
- Mullis, I. V. S., & Martin, M. O. (2019). PIRLS 2021 reading assessment framework. In I. V. S. Mullis & M. O. Martin (Eds.), *PIRLS 2021 Assessment Frameworks*. Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/pirls2021/frameworks/>
- Martin, M. O., von Davier, M., Foy, P., & Mullis, I. V. S. (2019). PIRLS 2021 assessment design. In I. V. S. Mullis & M. O. Martin (Eds.), *PIRLS 2021 Assessment Frameworks*. Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/pirls2021/frameworks/>
- Mullis, I. V. S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., & Wry, E. (2023). *PIRLS 2021 International Results in Reading*. Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb5342>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>

- SAS Institute. (2016). SAS for Windows (Version 9.4) [Computer software]. Cary, NC: SAS Institute, Inc.
- Wry, E., Fishbein, B., & von Davier, M. (2023). Using scale anchoring to interpret the PIRLS 2021 achievement results. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 14.1-14.32). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb8403>
- Yin, L., Bezirhan, U., Fishbein, B., & Foy, P. (2023b). Implementing the PIRLS 2021 achievement scaling methodology. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 11.1-11.82). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb3067>
- Yin, L., Fishbein, B., Bezirhan, U., Foy, P., & von Davier, M. (2023a). Examining country-level differences between digitalPIRLS data and bridge data. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 12.1-12.33). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb9281>
- Yin, L., & Reynolds, K. A. (2023). Creating and interpreting the PIRLS 2021 context questionnaire scales. In M. von Davier, I. V. S. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and Procedures: PIRLS 2021 Technical Report* (pp. 15.1-15.161). Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.tr2103.kb6994>

CHAPTER 3

Special R, SPSS, and SAS Programs

Overview

This chapter describes special R (R Core Team, 2020), SPSS (IBM Corporation, 2016), and SAS (SAS Institute, 2016) programs available in the PIRLS 2021 International Database. The programs recode the responses to the individual PIRLS 2021 items from the achievement data files to their appropriate score levels and apply customized values for item non-response. One program is provided for the PIRLS 2021 student achievement (ASA) files and another is provided for the PIRLS 2021 Bridge files.

Scoring the PIRLS 2021 Items

Student achievement in PIRLS 2021 is represented by sets of five plausible values for overall reading and the reading purpose and comprehension process subdomains. They are the preferred scores for any analysis of student achievement. However, analyzing performance on individual items may be of interest to some users. Carrying out such analyses may require that the individual items in the PIRLS 2021 International Database be assigned their correctness score levels, rather than the actual response options selected by students for multiple-choice items. Users may also wish to apply certain values to item-level non-response codes for Omitted and Not Reached (described in Chapter 2). The International Database provides R, SPSS, and SAS programs to perform this task.

For multiple-choice items and some item parts of digital item types, the numbers 1 through 4 are used to represent response options A through D, respectively, in the PIRLS 2021 achievement data files.⁶ These responses must be converted to their appropriate score level (“1” for correct and “0” for incorrect) based on each multiple-choice item’s correct response key. Constructed-response items, worth a total of 1, 2, or 3 score points, already are coded as the appropriate point values of the students’ responses.

⁶ Some digital items offer more than 4 response options and therefore will use numbers beyond 4 to represent them. This information is provided in the codebooks and the item information spreadsheets.

For all items, special codes are set aside to represent missing data as *Not Administered*, *Omitted*, or *Not Reached*. These special missing codes also may be recoded in order to carry out specific item-level analyses. By default, the Not Administered, Omitted, and Not Reached response codes are left as missing. These default settings can be modified within the score programs, depending on the requirements of the item-level analyses. For example, users may wish to treat Omitted responses as incorrect, instead of missing.

The PIRLS 2021 International Database includes two R scripts, two SPSS programs, and two SAS programs to recode the responses to individual items from the achievement data files (ASA) to their appropriate score levels. The ASASCRR5 programs score the paperPIRLS and digitalPIRLS (including ePIRLS) items for the data files with the “R5” suffix. The ASASCRA5 programs score the paper trend items for the data files with the “A5” suffix associated with the PIRLS 2021 Bridge data.

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

Exhibit 3.1 shows a condensed version of the ASASCRR5.R program to score the individual PIRLS 2021 items. All R, SPSS, and SAS programs have a similar structure (see Exhibits 3.2 and 3.3). To score each individual PIRLS 2021 item, the program code in the R score program must be adapted by completing the following steps:

1. Open ASASCRR5.R or ASASCRA5.R with R or RStudio (RStudio Team, 2020)
2. At the end of the program, specify the folder where the merged R data file of student achievement data files is located in the argument **indir**
3. Specify the folder where the merged R data file of scored achievement items will be located in the argument **outdir**
4. Specify the name of the merged student achievement R data file in the argument **infile**
5. Submit the edited code for processing by R

In this example, the merged R data file of student achievement data files is called ASAALLR5, located in the folder C:\PIRLS2021\Data, and contains the achievement item variables. The resulting data file of scored achievement items will be called ASAALLR5_SCR and saved in the location specified in the **outdir** argument. In this case, the same folder was used for both **indir** and **outdir**. Note that in R, (forward) slashes (/) should be used instead of

backslashes in the file path, so the folder path in Exhibit 3.1 is given as: C:/PIRLS2021/Data/. If using SPSS or SAS, backslashes should be used.

Exhibit 3.1: The ASASCRR5.R Program (Condensed)

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

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

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

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

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

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

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

  . . .
}

doit(indir = "C:/PIRLS2021/Data/",
     outdir = "C:/PIRLS2021/Data/",
     infile = "ASAALLR5")
```

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

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

```
. == NR ~ NaN, # Not Reached
. == OM ~ NaN, # Omitted
```

with these statements:

```
. == NR ~ 0, # Not Reached
. == OM ~ 0, # Omitted
```

Users should be careful to use the score program appropriate for the student achievement data files under consideration: the ASASCRR5 program for the items in suffix “R5” files, and the ASASCRA5 program for the items in suffix “A5” files.

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

Exhibit 3.2: The ASASCRR5.SPS Program (Condensed)

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

!ENDDDEFINE .

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

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

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

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

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

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

!ENDDDEFINE .

DOIT INDIR = C:\PIRLS2021\Data /
  OUTDIR = C:\PIRLS2021\Data /
  INFILE = ASAALLR5 .
```

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

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

with these statements:

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

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

Exhibit 3.3: The ASASCRR5.SAS Program (Condensed)

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

. . .

%MEND SCOREIT ;

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

. . .

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

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

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

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

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

. . .

%MEND DOIT ;

%DOIT (INDIR = C:\PIRLS2021\Data ,
      OUTDIR = C:\PIRLS2021\Data ,
      INFILE = ASAALLR5 ) ;
```

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

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

with these statements:

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

References

- IBM Corporation. (2016). IBM SPSS Statistics for Windows (Version 24.0) [Computer software]. Armonk, NY: IBM Corporation.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>
- RStudio Team (2020). 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.

APPENDIX

Organizations and Individuals Responsible for PIRLS 2021

Introduction

PIRLS 2021 was a collaborative effort involving hundreds of individuals around the world. This appendix acknowledges the key individuals and organizations for their contributions. Given that the work on PIRLS 2021 spanned more than five years and involved so many people and organizations, this is not a comprehensive list of everyone who contributed. PIRLS 2021 also acknowledges the students, parents, teachers, and school principals who contributed their time and effort to the study. This report would not be possible without them.

Management and Coordination

PIRLS is a major undertaking of IEA, and together with the Trends in International Science and Mathematics Study (TIMSS), comprises the core of IEA's regular cycles of studies. The PIRLS assessment at the fourth grade complements TIMSS, which regularly assesses mathematics and science achievement at the fourth and eighth grades.

PIRLS 2021 was conducted by IEA's TIMSS & PIRLS International Study Center, which is responsible for the overall direction and management of TIMSS and PIRLS, including design, development, and implementation. Headed by Executive Directors Drs. [Ina V.S. Mullis](#) and [Matthias von Davier](#), the study center is located in the Lynch School of Education and Human Development at Boston College. The TIMSS & PIRLS International Study Center worked closely with IEA Amsterdam, which was responsible for country participation, verification of all translations produced by the participating countries, and coordinating school visits for the International Quality Control Monitors. In addition, staff at IEA Hamburg worked with participating countries to organize sampling and data management operations and to check all data for accuracy and consistency within and across countries. Statistics Canada in Ottawa was responsible for school and student sampling activities.

The Project Management Team, comprising the study directors and representatives from the TIMSS & PIRLS International Study Center, IEA Amsterdam, IEA Hamburg, and Statistics Canada met twice a year throughout the study to discuss the study's progress, procedures, and schedule. In addition, the study directors met with members of IEA's Technical Executive Group twice each year to review technical issues.

To work with the international team and coordinate within-country activities, each participating country designates an individual to be the PIRLS National Research Coordinator (NRC). The NRCs have the challenging task of implementing PIRLS in their countries in accordance with the PIRLS guidelines and procedures. In addition, the NRCs contribute to the development of PIRLS and provide input throughout the course of the project. The quality of the PIRLS assessment and data depends on the work of the NRCs and their colleagues in carrying out the complex sampling, data collection, and scoring tasks. The PIRLS 2021 NRCs worked tirelessly to support PIRLS 2021 data collection despite the difficult circumstances presented by the COVID-19 pandemic. They performed their many tasks with dedication, competence, energy, and goodwill. PIRLS 2021 owes its success to the commitment and determination of the NRCs and their colleagues.

Funding

Funding for PIRLS 2021 was provided primarily by the participating countries. Boston College also is gratefully acknowledged for its generous financial support and stimulating educational environment.

TIMSS & PIRLS International Study Center at Boston College

Ina V.S. Mullis, *Executive Director*

Matthias von Davier, *Executive Director*

Michael O. Martin, *Executive Director (through 2021)*

Pierre Foy, *Senior Director, Sampling, Psychometrics, and Data Analysis*

Ieva Johansone, *Senior Director, Operations and Quality Control*

Ann Kennedy, *Senior Director, Project Management (from 2022)*

Paul Connolly, *Director, Graphic Design and Publications*

Dana Kelly, *Director, Development and Reporting (through 2021)*

Sonia Baron, *Graduate Assistant*

Ummugul Bezirhan, *Senior Research Specialist, Measurement and Data Analysis*

Marcie Bligh, *Manager, Events and Administration*

Alicia Bouchard, *Administrative Assistant*

Jessie Bristol, *Research Specialist, Measurement and Data Analysis*

Susan Farrell, *Lead Web and Database Designer*

Bethany Fishbein, *Assistant Director, Psychometrics and Data Analysis*

Susan Flicop, *Research Specialist, Operations and Quality Control*

Christine Hoage, *Manager of Finance (through 2021)*

Joan King, *Manager of Finance (from 2022)*

Maya Komakhidze, *Graduate Assistant*

Dihao Leng, *Graduate Assistant*

Jenny Liu, *Research Specialist, Questionnaire Development and Policy Research (through 2019)*

Tianzheng Mao, *Research Associate, Measurement and Data Analysis (from 2022)*

Anne McCarthy, *Front-End Web Developer (through 2022)*

Mario Pita, *Lead Graphic Designer*

Sean Quinn, *Senior Front-End Web Developer (from 2022)*

Katherine Reynolds, *Assistant Director, Questionnaire Development and Policy Research*

Ruthanne Ryan, *Senior Graphic Designer*

Erin Shaw, *Research Associate, Reporting and Communications*

Steven A. Simpson, *Senior Graphic Designer (through 2021)*

Nicole St. Louis, *Research Associate, Operations and Quality Control*

Lillian Tyack, *Research Specialist, Measurement and Data Analysis*

Erin Wry, *Senior Research Specialist, PIRLS Coordinator*

Liqun Yin, *Senior Research Psychometrician*

IEA Amsterdam

Dirk Hastedt, *Executive Director*

Andrea Netten, *Director*

Paulína Koršňáková, *Senior Research and Liaison Adviser*

Roel Burgers, *Financial Director (through 2021)*

Jan-Peter Broek, *Financial Director*

Isabelle Braun-Gémin, *Senior Financial Officer*

Sandra Dohr, *Research Officer (through 2021)*

David Ebbs, *Senior Research Officer*

Philippa Elliott, *Publications Manager*

Katie Hill, *Head of Communications*

Marta Moreno Hidalgo, *Junior Research Officer*

IEA Hamburg

Juliane Hencke, *Director*

Heiko Sibberns, *Senior Research Advisor*

Oliver Neuschmidt, *Senior Research Analyst, Head of International Studies Unit*

Mark Cockle, *Research Analyst, International Data Manager - PIRLS*

Viktoria Böhm, *Research Analyst, International Data Manager - PIRLS*

Christine Busch, *Research Analyst, Team Lead QUADIS*

Elma Cela, *Research Analyst (through 2021)*

Limao Duan, *Software Developer*

Eugenio Gonzalez, *Senior Research Analyst*

Deepti Kalamadi, *Software Developer (through 2021)*

Svenja Kalmbach, *Research Analyst*

Alec Kennedy, *Senior Research Analyst*

Kamil Kowolik, *Research Analyst*

Lorelia Lerps, *Research Analyst*

Guido Martin, *Senior Research Analyst, Head of Coding Unit*

Sabine Meinck, *Senior Research Analyst, Head of Research, Analysis and Sampling*

Yagmur Rizvi, *Research Analyst (through 2020)*

Duygu Savaşçı, *Research Analyst, Sampling Unit*

Katharina Sedelmayr, *Research Analyst, Coding Unit*

Hannah Smith, *Research Analyst, Sampling Unit (through 2021)*

Xiao Sui, *Research Analyst*

Milena Taneva, *Senior Research Analyst*
 Sabine Tieck, *Research Analyst, Team Lead Sampling*
 Meng Xue, *Head of Software Unit*
 Neha Yadav, *Software Developer (through 2020)*

Statistics Canada

Ahmed Almaskut, *Senior Methodologist*
 Sylvie LaRoche, *Senior Methodologist (through 2020)*

Sampling Referee

Keith Rust, *Senior Vice President and Senior Statistical Fellow, Westat, Inc.*

PIRLS 2021 Reading Development Group

Julian Fraillon
 IEA Amsterdam
Australia

Jan Mejding
 Aarhus University
 Department of Education (DPU)
Denmark

Liz Twist
 National Foundation for Educational
 Research
England

Marc Colmant
 Ministère de l'Éducation nationale et de la
 Jeunesse
 Direction de l'Évaluation, de la Prospective
 et de la Performance (DEPP)
France

Galina Zuckerman
 Psychological Institute
 Russian Academy of Education
Russian Federation

Elizabeth Pang
 Curriculum Planning and Development
 Division
 Ministry of Education
Singapore

Verónica Díez Girado
 National Institute for Educational
 Assessment
 Ministry of Education
Spain

Jenny Wiksten Folkeryd
 Uppsala University
 Department of Education
Sweden

Latifa Alfalasi
 Ministry of Education
United Arab Emirates

Karen Wixson
 Educational Testing Service
United States

PIRLS 2021 Item Development and Scoring Working Group

Ina V.S. Mullis, *TIMSS & PIRLS International Study Center Executive Director*
 Michael O. Martin, *TIMSS & PIRLS International Study Center Executive Director*
 Erin Wry, *Senior Research Specialist, PIRLS Coordinator*
 Julian Fraillon, *PIRLS Consultant (IEA Amsterdam)*
 Liz Twist, *PIRLS Consultant (National Foundation for Educational Research)*
 Karen Wixson, *PIRLS Consultant (Educational Testing Service)*

PIRLS 2021 Questionnaire Development Group

Anna Matoul
 Université de Liège
Belgium (French)

Beti Lameva
 National Examination Center
North Macedonia

Hwa Wei Ko
 Graduate Institute of Learning and
 Instruction
 National Central University
Chinese Taipei

Fathia Mohammed Amour Al Mawali
 Ministry of Education
Oman

Kaisa Leino
 Finnish Institute for Educational Research
 University of Jyväskylä
Finland

Surette van Staden
 Centre for Evaluation and Assessment
 (CEA)
 University of Pretoria
South Africa

Megan Chamberlain
 Ministry of Education
 Educational Measurement and Assessment
 Team
New Zealand

Bridget Dalton
 University of Colorado – Boulder
 School of Education
United States

PIRLS 2021 National Research Coordinators

Albania

Aurora Balliu
Rezana Vrapı (through 2021)
*National Agency of Examinations (NAE)
Center for Educational Services*

Australia

Sue Thomson
Australian Council for Educational Research

Austria

Juliane Schmich
*Federal Institute for Quality Assurance of the
Austrian School System (IQS)*

Azerbaijan

Nermine Aliyeva
*Ministry of Education of the Republic of
Azerbaijan*

Bahrain

Samah Mohamed Sulaiman Alajjawi
Huda Al-Awadi (through 2021)
Ministry of Education

Belgium (Flemish)

Katrijn Denies
*KU Leuven
Centre for Educational Effectiveness and
Evaluation*

Belgium (French)

Anne Matoul
Université de Liège

Brazil

Lorena Pimenta de Andrada
Viviane Pinto (through 2022)
INEP, Ministry of Education

Bulgaria

Marina Vasileva Mavrodieva
*Center for Assessment in Pre-School and
School Education (CAPSE)*

Canada

Kathryn O'Grady
Tanya Scerbina
*Council of Ministers of Education, Canada
(CMEC)*

Chinese Taipei

Yu-Wen Chang
National Taipei University of Education

Croatia

Ines Elezović
*National Centre for External Evaluation of
Education*

Cyprus

Yiasemina Karagiorgi
*Center of Educational Research and
Evaluation
Cyprus Pedagogical Institute*

Czech Republic

Zuzana Janotová
Czech School Inspectorate

Denmark

Simon Skov Fougat
Jan Mejding (through 2019)
Aarhus University

Egypt

Hanaa Mohamed Aly Makhoulf
*National Center for Examinations and
Educational Evaluation*

England

Grace Grima
Pearson UK

Finland

Kaisa Leino
*Finnish Institute for Educational Research
University of Jyväskylä*

France

Marc Colmant
*Ministère de l'Éducation nationale et de la
Jeunesse
Direction de l'Évaluation, de la Prospective
et de la Performance (DEPP)*

Georgia

Giorgi Ratiani
Irine Samsonia (through 2021)
*National Assessment and Examinations
Center*

Germany

Nele McElvany
*Institut für Schulentwicklungsforschung IFS
Center for Research on Education and
School Development
TU Dortmund University*

Hong Kong SAR

Wai Ip Josephine Lam
*Faculty of Education
The University of Hong Kong*

Hungary

Ildikó Balázs
Péter Balkányi
*Department of Assessment and Evaluation
Educational Authority*

Iran, Islamic Republic of

Masoud Kabiri
Abdol'azim Karimi (through 2021)
*The Organization for Educational Research
and Planning*

Ireland

Emer Delaney
Educational Research Centre

Israel

Georgette Hilu
Inbal Ron-Kaplan
Noa Schori Eyal
*National Authority for Measurement and
Evaluation in Education (RAMA)*

Italy

Laura Palmerio
Margherita Emiletti
*Istituto Nazionale per la Valutazione del
Sistema Educativo di Istruzione e di
Formazione (INVALSI)*

Jordan

Khattab M. A. Abulibdeh
*National Center for Human Resources
Development*

Kazakhstan

Nazym Smanova
Aigul Baigulova (through 2020)
*JSC "Information-Analytical Center"
Ministry of Education of the Republic of
Kazakhstan*

Kosovo

Nizafete Kutlllovci-Bardhi
*Ministry of Education, Science and
Technology of Kosovo*

Latvia

Antra Ozola
*Faculty of Education, Psychology and Art
University of Latvia*

Lithuania

Ramutė Skripkienė
*Research and Analysis Unit
National Agency for Education*

Macao SAR

Napier Kin Mou Wong
Ao In Heng (through 2021)
*Education and Youth Development Bureau
(DSEDJ)*

Malta

Louis Scerri
*Directorate for Learning and Assessment
Programmes
Ministry for Education, Sport, Youth,
Research and Innovation*

Montenegro

Marina Radović
Milanka Izgarevic (through 2019)
Examination Centre of Montenegro

Morocco

Mohammed Sassi
*Centre National de l'Evaluation, des
Examens et de l'Orientation
Ministere de l'Education Nationale et de la
Formation Professionnelle*

Netherlands

Joyce Gubbels
Dutch Center for Language Education

New Zealand

Megan Chamberlain
*Educational Measurement and Assessment
Team
Ministry of Education*

North Macedonia

Biljana Mihajlovska
Tanja Andonova Mitrevska (through 2020)
State Examination Center

Northern Ireland

Juliet Sizmur
Rachel Classick
*National Foundation for Educational
Research*

Norway

Åse Kari H. Wagner
*Norwegian Reading Centre
University of Stavanger*

Oman

Zuwaina Saleh AlMaskari
Ministry of Education

Poland

Joanna Kaźmierczak
Educational Research Institute

Portugal

Anabela Serrão
Vanda Lourenço (through 2019)
Instituto de Avaliação Educativa, I. P.

Qatar

Badriya Salman Al-Mohannadi
*Student Assessment Office
Ministry of Education and Higher Education*

Russian Federation

Galina Sidorova
Galina Kovaleva
*Federal Institute for the Strategy of
Education
Development of the Russian Academy of
Education
Centre for Evaluating the Quality of
Education*

Russian Federation (continued)

Sergey Stanchenko
*Federal Institute for the Evaluation of the
Education Quality*

Saudi Arabia

Abdullah Aljouiee
*Education and Training Evaluation
Commission
National Center of Assessment*

Serbia

Branislav Randjelovic
Institute for Education Quality and Evaluation

Singapore

Elizabeth Pang
*Curriculum Planning and Development
Division
Ministry of Education*

Foo Seau Fah
Chan Lee Shan
Vanessa Chua
*Research and Management Information
Division
Research and Evaluation Branch
Ministry of Education*

Slovak Republic

Gabriella Kopas
Kristína Čevorová (through 2021)
*National Institute for Certified Educational
Measurement (NUCEM)*

Slovenia

Eva Klemencic
Educational Research Institute

South Africa

Surette van Staden
*Centre for Evaluation and Assessment (CEA)
University of Pretoria*

Mark Chetty
*National Assessment Department of Basic
Education*

Spain

Verónica Díez Girado
*National Institute for Educational Assessment
Ministry of Education*

Sweden

Cecilia Stenman
*Swedish National Agency for Education
(SKOLVERKET)*

Turkiye

Muhsin Polat
*The General Directorate of Measurement,
Evaluation and Examination Services
The Ministry of National Education*

United Arab Emirates

Shaikha Ali Al Zaabi
Hessa Al Wahhabi
Ministry of Education

United States

Mary Coleman
Sheila D. Thompson (through 2021)
*National Center for Education Statistics
U.S. Department of Education*

Uzbekistan

Abduvali Ismailov
*National Center for Conducting International
Studies on the Assessment of Quality of
Education*

Benchmarking Participants

Ontario, Canada

Jeannette Amio

Jennifer Hove

Laurie McNelles (through 2021)

Education Quality and Accountability Office

Quebec, Canada

Laurence Harvey

Ministère de l'Éducation et de

l'Enseignement supérieur

Moscow City, Russian Federation

Elena Zozulya

Moscow Center for Quality of Education

Abu Dhabi, United Arab Emirates

Nada Abu Baker Husain Ruban

Maryam Al Hawwai

Ministry of Education UAE

Dubai, United Arab Emirates

Mariam Al Ali

Knowledge and Human Development

Authority



BOSTON
COLLEGE

timssandpirls.bc.edu

