IEA

LaNA

LITERACY AND NUMERACY ASSESSMENT

LaNA 2023

LINKING STUDY RESULTS

Matthias von Davier, Lale Khorramdel,
Katherine Reynolds, Charlotte E. A. Aldrich,
Allison Bookbinder, Ann Kennedy,
and Eugenio Gonzalez

IEA TIMSS & PIRLS BOSTON COLLEGE

LaNA 2023 Linking Study Results
Matthias von Davier, Lale Khorramdel, Katherine Reynolds, Charlotte E. A. Aldrich,
Allison Bookbinder, Ann Kennedy, and Eugenio Gonzalez

Suggested Citations

Full Report:

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

Report Appendices:

Aldrich, C. E. A. & Khorramdel, L. (2025). Appendix A: Instrument development and the LaNA 2023 linking study design. In M. von Davier, L. Khorramdel, K. Reynolds, C. E. A. Aldrich, A. Bookbinder, A. Kennedy, & E. Gonzalez, *LaNA 2023 Linking Study Results* (pp. A.1-A.13). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

Savaşcı-Smith, D., Leyton, R., & Tieck, S. (2025). Appendix B: Sample implementation. In M. von Davier, L. Khorramdel, K. Reynolds, C. E. A. Aldrich, A. Bookbinder, A. Kennedy, & E. Gonzalez, *LaNA 2023 Linking Study Results* (pp. B.1-B.14). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

Neuschmidt, O. & Beyer, C. (2025). Appendix C: Survey operations and international database. In M. von Davier, L. Khorramdel, K. Reynolds, C. E. A. Aldrich, A. Bookbinder, A. Kennedy, & E. Gonzalez, *LaNA 2023 Linking Study Results* (pp. C.1-C.22). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

Yin, L., Bezirhan, U., & Khorramdel, L. (2025). Appendix D: Psychometric analysis. In M. von Davier, L. Khorramdel, K. Reynolds, C. E. A. Aldrich, A. Bookbinder, A. Kennedy, & E. Gonzalez, *LaNA 2023 Linking Study Results* (pp. D.1-D.21). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

Khorramdel, L., Aldrich, C. E. A., & Kennedy, A. (2025). Appendix E: Scale anchoring and description of new basic LaNA benchmarks. In M. von Davier, L. Khorramdel, K. Reynolds, C. E. A. Aldrich, A. Bookbinder, A. Kennedy, & E. Gonzalez, *LaNA 2023 Linking Study Results* (pp. E.1-E.6). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

For more information about LaNA 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

Foreword

Part of the mission of IEA (the International Association for the Evaluation of Educational Achievement) is to support education systems worldwide with data-driven insights for evidence-based policies and improved student learning outcomes. Together with the TIMSS & PIRLS International Study Center at Boston College, LaNA (Literacy and Numeracy Assessment) has been developed to offer a flexible, accessible, and thorough assessment of reading and mathematics skills at the end of primary education. It has a targeted assessment design to more accurately measure foundational literacy and numeracy skills. In this way, LaNA allows education systems to gain deeper insights into their students who are in the process of developing these necessary skills, as well as those who are achieving high results. This study has been designed to connect LaNA to the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS) to continue their tradition of high-quality assessment while also adapting to various contexts in order to better meet the diverse needs countries have.

This 2023 LaNA Linking Study is a special administration made possible with the participation of six education systems from around the world: Burkina Faso, Egypt, Nigeria, Pakistan, the Palestinian National Authority, and Senegal. It marks a crucial milestone to further establish an understanding of achievement ranges connected to the TIMSS mathematics and PIRLS reading benchmarks.

These education systems' involvement and the subsequent analyses supported the establishment of a psychometrically sound assessment. In this report, links between LaNA's numeracy and literacy sections and the respective TIMSS mathematics and PIRLS reading scales are represented using appropriate confidence intervals. This allowed for the introduction of a new Basic International Benchmark, while also contributing to participating countries being able to evaluate their students' proficiency against internationally recognized benchmarks. Furthermore, the results of LaNA provide essential tools for tracking progress toward UNESCO's SDG (Sustainable Development Goal) 4.1.1b, which focuses on achieving minimum proficiency levels in these critical areas. This assessment is able to provide valuable insights into student achievement and backgrounds within important global contexts.

The study's findings highlight the diversity and complexity of education systems across the world. Beyond reporting on student achievement, the assessment delves into the contextual factors that influence learning, such as access to resources and school climate. Such findings provide policymakers, educators, and researchers with a comprehensive understanding of the unique educational environments within participating countries, as well as the common challenges faced globally.

The LaNA Linking Study would not have been possible without the collaboration and support of many individuals and institutions. We extend our deepest gratitude to the NRCs (National Research Coordinators) and their teams for their commitment and execution of this multifaceted

study. This special assessment of LaNA was graciously made possible by the financial support from participating countries, the World Bank, IEA, and Boston College.

We also acknowledge the contributions of the TIMSS & PIRLS International Study Center (ISC) at Boston College's Lynch School of Education and Human Development, whose expertise in assessment design, implementation, and analysis formed the foundation of LaNA's success. The dedication of the ISC staff in their endeavors to provide high-quality data that addresses the diverse needs of education globally has been pivotal in the development of LaNA. The ISC's collaborative efforts with IEA exemplify the importance of partnership in tackling global education challenges. Further thanks extend to the staff at IEA's Amsterdam and Hamburg offices for their comprehensive work in seeing this study through to its high-quality completion, especially to the Capacity Building team for their instrumental coordination and management.

A special thank you goes to the many officials, teachers, schools, and students who participated in the assessment, as all of you provided the data that make these analyses meaningful. Your participation helps illuminate both the strengths and areas for improvement within education systems and build evidence that assists with informing educational policy.

We hope to see regular administrations of LaNA continue around the world, supported by the benchmarks established in this inaugural Linking Study. In addition to providing high-quality data on an education system, this assessment also serves as a capacity-building initiative, equipping national teams with experience in assessment planning, design, implementation, and data analysis. For countries considering future participation in TIMSS and PIRLS, LaNA offers a foundation to establish national assessment networks and engage in further international large-scale studies. Publications dedicated to the use of LaNA's data and opportunities to join the study can be found on the IEA website.

As we celebrate the achievements of the LaNA Linking Study, we are reminded of the critical role that collaboration, innovation, and shared vision play in improving education for all. We hope this report serves as a valuable resource for advancing student learning and inspiring continued progress toward the SDGs.

IEA remains committed to fostering innovation and capacity building through initiatives like LaNA, and we look forward to ongoing collaboration with more countries choosing to administer this important study as we work together to improve education systems worldwide.

Dirk Hastedt

IEA Executive Director

Introduction

IEA's Literacy and Numeracy Assessment (LaNA)

The IEA's Literacy and Numeracy Assessment (LaNA), developed by the TIMSS & PIRLS International Study Center at Boston College, is an international assessment designed to measure emerging reading and mathematics skills at the primary school level. It builds on the robust frameworks of IEA's flagship studies, the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS), administered in over 70 countries. Compared to TIMSS and PIRLS, LaNA offers a less demanding assessment tailored for students at the end of primary school for whom the TIMSS and PIRLS assessments are likely to be too difficult. The LaNA design includes fewer test booklets than TIMSS and PIRLS, consequently leading to fewer resources needed for translation and layout verification, among other things, and can therefore be administered off-cycle in a shorter timeline. Currently available in a paper-and-pencil format, LaNA can be implemented in fourth, fifth, or sixth grade, depending on the educational context.

LaNA employs a rotated booklet design to assess reading and mathematics comprehensively. Each student responds to a subset of questions from the total item pool. The assessment includes a 45-minute reading comprehension section, a 45-minute mathematics section (both composed of multiple-choice items), and a 30-minute background questionnaire section.

The reading assessment consists of five distinct text passages, each accompanied by 10–11 comprehension questions. These passages reflect the dual purposes of reading in PIRLS: reading for literary experience (stories) and reading to acquire and use information. Each student reads two passages and answers the accompanying questions.

The mathematics assessment includes 80 items, with each student completing 40. These items are organized into content domains (number, measurement and geometry, data) and cognitive domains (knowing, applying, and reasoning), closely mirroring those in TIMSS. Students engage in tasks such as recognizing and comparing simple fractions, performing whole-number computations, and interpreting graphs. More information about the assessment design can be found in Appendix A of this report.

In addition to the cognitive assessments, LaNA gathers contextual information about factors associated with learning, such as resource availability, students' learning experiences, and the broader learning environment. These data are collected through context questionnaires administered to students and school administrators. Appendix A of this report provides further details on the LaNA instruments and their development.

LaNA offers countries reliable data to inform educational policy. Its results can highlight strengths and weaknesses in education systems and identify characteristics of successful students and schools, providing a foundation for national policy interventions and evidence-based strategies. Repeated administrations of LaNA can also help evaluate the effects of educational policies on student outcomes over time. Moreover, LaNA offers new Basic

International Benchmarks for mathematics and reading that extend the TIMSS and PIRLS International Benchmarks. This allows for connecting student performance to TIMSS and PIRLS achievement scales which provide an internationally recognized reference, acknowledged by UNESCO as contributing to the monitoring of progress toward the UN's Sustainable Development Goal (SDG) 4.1.1b.

Beyond the assessment results, participation in LaNA provides countries and national teams with valuable experience in assessment planning, sampling, design, implementation, and standardized test administration. It also enhances capabilities in result interpretation and reporting. Thus, LaNA serves as a stepping stone for participation in future TIMSS and PIRLS cycles or as capacity building for the development of national assessments.

The LaNA 2023 Linking Study

The LaNA 2023 Linking Study was conducted in six participating countries—Burkina Faso, Egypt, Nigeria, Pakistan, Palestinian National Authority, and Senegal—to establish new basic international benchmarks for mathematics and reading on the TIMSS and PIRLS scales. These benchmarks provide a valuable opportunity for countries where TIMSS and PIRLS are too challenging to evaluate their students' reading and mathematics performance against international standards.

To establish the connection between LaNA and TIMSS and PIRLS, the study employed a special linking design that combined four LaNA-specific booklets with four carefully designed linking booklets. The linking booklets included easier TIMSS and PIRLS blocks consisting of multiple-choice and constructed-response items to ensure alignment.

For nationally representative reading and mathematics outcomes, the target sampling design in each country required 4,500 students to be assessed in intact classes from at least 100 schools. IEA sampling experts guided the selection of classrooms and schools. Burkina Faso, Nigeria, and Senegal implemented the study at the beginning of Grade 6, while Egypt, Pakistan, and the Palestinian National Authority administered it at the beginning of Grade 5. Appendices A and B provide detailed information about the linking design and the sample implementation for each country.

The administration of the linking study followed best practices adopted from TIMSS and PIRLS, including detailed test administrator and data processing guidelines, as well as scoring guides and scoring training. More information is provided in Appendix C.

The psychometric analysis for establishing the link to TIMSS and PIRLS used item response theory (IRT) scaling, which utilized TIMSS 2019 and PIRLS 2021 item parameters in a fixed item parameter linking and latent regression population modeling for calculating plausible values for mathematics and reading provided in the international database. Appendix C provides information about the creation of the international database, and Appendix D provides information about the psychometric analysis.

Using the results from the IRT scaling, the same approach to scale anchoring as used in TIMSS and PIRLS was applied to develop descriptions for the new LaNA Basic International Benchmarks for Mathematics and Reading. These descriptions detail the skills and competencies students demonstrate upon reaching these benchmarks. They complement the existing TIMSS 2019 and PIRLS 2021 benchmark descriptions, extending the ability to describe the lower end of the proficiency scales based on the types of tasks covered in LaNA. Further details on the scale anchoring can be found in Appendix E.

Outcomes and Structure of this Report

The LaNA 2023 Linking Study Report is organized into the international results and technical documentation that consists of several appendices. The international results include the average achievement ranges in reading and mathematics within and across participating countries, the percentage of students reaching the new LaNA Basic International Benchmarks and the TIMSS and PIRLS International Benchmarks, and examine the relationship between these achievement outcomes and contextual variables gathered through the student and school context questionnaires.

The appendices to this report provide the following technical documentation:

- Appendix A: Instrument Development and the LaNA 2023 Linking Study Design
- Appendix B: Sample Implementation
- Appendix C: Survey Operations and International Database
- Appendix D: Psychometric Analysis
- Appendix E: Scale Anchoring and Description of New Basic LaNA Benchmarks
- Appendix F: Organizations and Individuals Responsible for LaNA 2023 Linking Study

In addition to the report, countries receive a database containing comparable LaNA data. This database is accompanied by a User Guide and analysis software to support further analyses for national educational planning and policymaking.

The LaNA Link: Reporting Results Using Confidence Intervals

LaNA is less complex than TIMSS and PIRLS and is intended for countries where TIMSS and PIRLS may be too difficult for most students. LaNA can be administered off-cycle in a shorter timeline than TIMSS or PIRLS and at any time during the school year. While based on the TIMSS mathematics and PIRLS reading assessment frameworks, the design of LaNA features fewer booklets, fewer items, and easier content compared to TIMSS and PIRLS. Additionally, while the linking booklets use TIMSS and PIRLS items, some of which are open ended and need human scoring, the newly developed LaNA booklets include only selected-response items that do not require human scoring. These features streamline the translation of the LaNA booklets into different languages and the scoring of responses collected with the LaNA booklets. However, this less complex, easier to use design consequently leads to a reduced coverage of the TIMSS and PIRLS frameworks, resulting in reduced measurement precision and the inability to report on specific content or cognitive domain subscales.

To reflect this difference in precision, this report presents achievement results as confidence intervals (CIs), instead of the estimated averages that are presented in TIMSS and PIRLS reports. CIs show the interval of a distribution where the true value of a statistic is likely to be found. This report uses 95% CIs, calculated assuming variances are estimated with sufficient degrees of freedom. A 95% coverage probability of confidence intervals indicates that, under repeated sampling, these intervals have a 95% chance of containing the true parameter value, provided the variance estimates are reliable. A 99% confidence interval would necessarily be wider than the 95% interval to gain this higher level of coverage probability, while a 90% confidence would be more narrow, but would mean that on average, among many replications, 10% of the intervals would not include the population average.

CIs have several useful qualities for characterizing achievement and achievement differences between groups. CIs are characterized by an upper and lower boundary that depend on the confidence level. These boundaries provide an easy way to gauge if differences between groups may be noteworthy and warrant further investigation. CIs can be compared with each other by checking for overlap between them. If two CIs do not overlap, this can be taken as an indication that there is very likely a difference between the two groups. If two CIs do overlap, groups may or may not differ. Comparing CIs based on their overlap is more conservative and leads to fewer "false positives" (e.g., incorrectly describing groups as significantly different), than statistical significance testing with an alpha error of 5%.

This report uses CIs to reap the benefit of a more conservative test of differences across different reporting groups. Throughout the report, readers will find commentary on CIs that do not overlap to draw attention to groups that are likely to differ in mathematics or reading achievement. However, lack of overlap does not necessarily mean that there is a difference of practical significance. This is a matter of further investigation, which requires triangulation of these results with additional evidence. Similarly, if CIs do overlap, this does not necessarily mean that there is no difference between groups. In these cases, examining the upper and lower bounds of CIs for monotonically increasing or decreasing patterns across the groups can still be fruitful in suggesting broad relationships between, for example, groups defined by context variables and their achievement in mathematics or reading.

Mathematics and Reading Achievement

This section presents the mathematics and reading achievement results for the six countries in the LaNA 2023 Linking Study. First, Exhibits 1.1.1 and 1.1.2 display the distributions of LaNA mathematics achievement by country and by gender, followed by Exhibits 1.2.1 and 1.2.2 which show the distributions of LaNA reading achievement by country and by gender.

The first set of exhibits is followed by descriptions of the new LaNA 2023 Basic Benchmarks established in the linking study (Exhibit 1.1.3a for mathematics and Exhibit 1.1.4a for reading). In addition, the summary descriptions of the TIMSS 2019 International Benchmarks in mathematics and the descriptions of the PIRLS 2021 International Benchmarks in reading are provided to contextualize the Basic Benchmarks (Exhibits 1.1.3b and 1.2.3b, respectively).

Subsequent exhibits present the percentages of students in LaNA countries reaching the new LaNA 2023 Basic Benchmarks for mathematics and reading, as well as the percentages reaching the TIMSS and PIRLS International Benchmarks based on the linking study results (Exhibit 1.1.4a for mathematics and Exhibit 1.2.4a for reading). These are followed by summaries of TIMSS 2019 and PIRLS 2021 countries' achievement at the International Benchmarks, presented in Exhibits 1.1.4b (mathematics) and 1.2.4b (reading).

As described in the introduction of this report, mathematics and reading achievement results are presented as 95% confidence intervals (CIs) around the estimate of average achievement. Readers are encouraged to refer back to the introduction for information about interpretation of these CIs. A brief description accompanies each exhibit to summarize main findings.

Mathematics Achievement

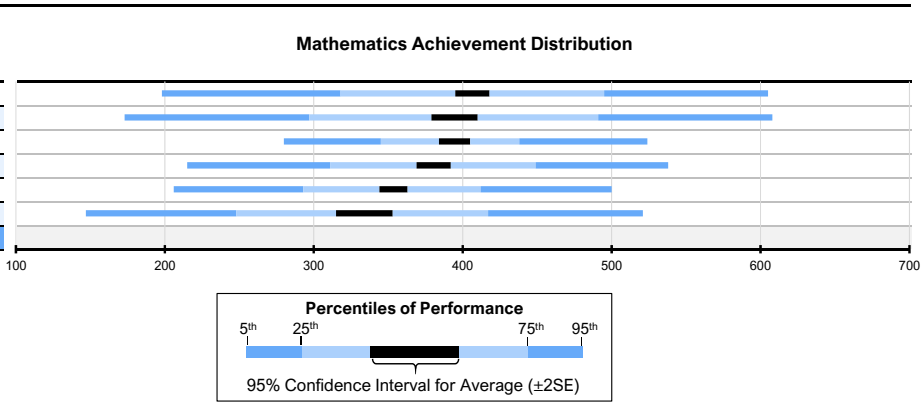
Mathematics Scale Score Distributions

Exhibit 1.1.1 shows the 95% confidence intervals for average mathematics achievement at the fifth and sixth grades for the six participating countries in the LaNA 2023 Linking Study. Additionally, it includes a visualization of the achievement distribution, showing the percentiles of the distribution and the 95% confidence interval for the average (LaNA 2023 Average) of the participating countries.

The six countries are presented sorted by the location of their confidence intervals. The confidence intervals of three participants (Palestinian National Authority, Egypt, and Burkina Faso) cover the 400 score point mark, while the upper boundary of the confidence intervals of Pakistan, Senegal, and Nigeria are below 400 (391, 362, and 353, respectively). There is some overlap among the confidence intervals for the Palestinian National Authority, Egypt, Burkina Faso, and Pakistan. The upper boundaries of the intervals for Senegal and Nigeria are 362 and 353, respectively, below the lowest confidence interval boundaries in the other four countries (368 in Pakistan).

Exhibit 1.1.1: Mathematics Scale Score Distributions

Country	95% Confidence Interval for Average Mathematics Scale Score
[1] Palestinian Nat'l Auth.	395 - 418
Egypt	379 - 410
Burkina Faso	384 - 405
† Pakistan	368 - 391
Senegal	343 - 362
[2] Nigeria	315 - 353
LaNA 2023 Average	372 - 382



LaNA mathematics scale scores are presented on the TIMSS trend scale established in 1995 with a centerpoint of 500 located at the mean of the combined achievement distribution. The units of the scale were chosen so that 100 scale score points corresponded to the standard deviation of the distribution. The LaNA link is based on a limited number of easy TIMSS trend blocks. See Appendix A for more information. See Appendix B.2 for population coverage notes 1 and 2. See Appendix B.6 for sampling guidelines and the sampling the participation note †.

While comparing the estimated locations of country averages is informative, the LaNA 2023 Linking Study data also provides information about the range of student achievement that was observed. The achievement distribution on the right-hand side of Exhibit 1.1.1 shows the range of student achievement observed in each country. The Palestinian National Authority and Egypt have the widest observed distributions, ranging from below 200 to above 600 for the fifth and ninety-fifth percentiles. The twenty-fifth and seventy-fifth percentile scores for these two countries range from about 300 to 500 scale score points. Somewhat narrower achievement ranges are observed in the other four countries; the top five percent of students (those above the 95% percentile) in all four countries reach 500 on the mathematics scale, while the lower end of the achievement range is somewhat more staggered across the score range, from about 150 to 280 scale score points.

Later sections of the report further enrich and contextualize the results by presenting the associations between mathematics achievement and various contextual factors, providing a deeper understanding of how achievement is related to other factors in each participating country.

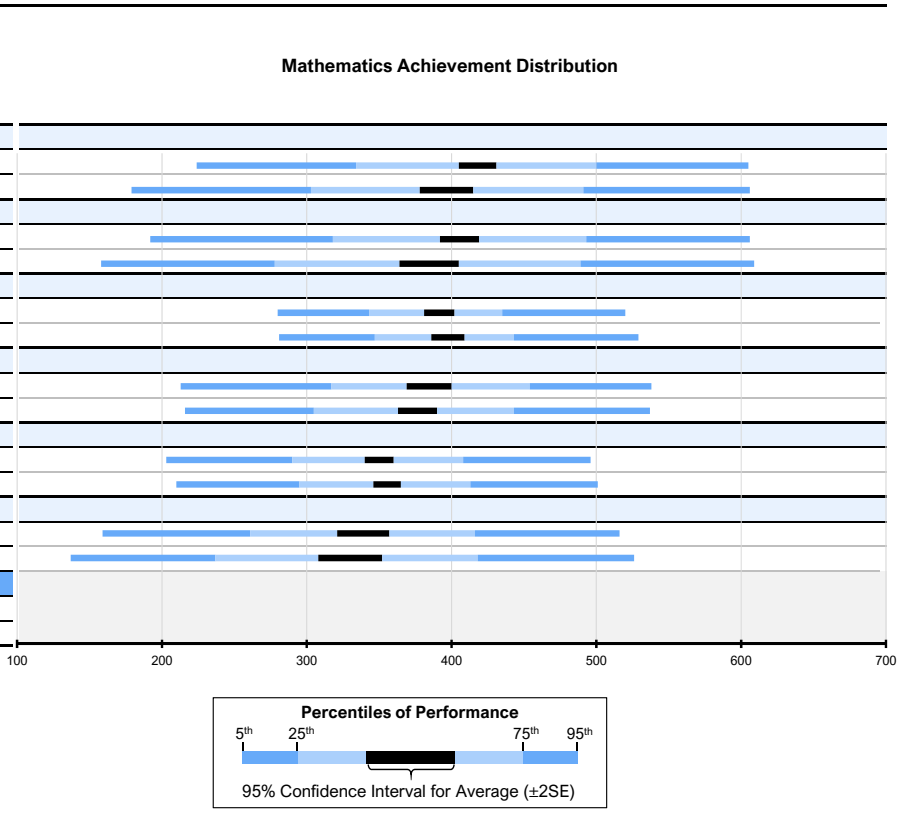
Mathematics Scale Score Distributions for Girls and Boys

Exhibit 1.1.2 illustrates the differences in achievement distributions between girls and boys across the six participating countries. While the 95% confidence intervals for boys and girls overlap in all countries, the upper boundaries of the confidence interval for girls are more than ten points above those for boys in Egypt and the Palestinian National Authority and the lower boundaries for the confidence interval associated with girls are 25 points above the lower confidence bound for boys in these countries. There is more overlap in Burkina Faso,

Nigeria, Pakistan, and Senegal since the difference between the lower and upper confidence interval boundaries for boys and girls is smaller. A similar picture emerges from the range of percentages: the fifth percentiles differ more between girls and boys for Egypt and the Palestinian National Authority than the differences between percentiles estimated for the other countries. Note that the percentages of girls and boys attending school differs across countries. This, in part, may lead to differences in results.

Exhibit 1.1.2: Mathemathics Achievement Scale Score Distributions for Girls and Boys

	Percent of Students	95% Confidence Interval for Average Mathematics Scale Score
[1] **Palestinian Nat'l Auth.**		
Girls	47 (1.7)	404 - 430
Boys	53 (1.7)	379 - 416
Egypt		
Girls	48 (2.2)	392 - 419
Boys	52 (2.2)	364 - 405
Burkina Faso		
Girls	56 (0.8)	381 - 402
Boys	44 (0.8)	386 - 409
† **Pakistan**		
Girls	46 (2.6)	369 - 400
Boys	54 (2.6)	362 - 389
Senegal		
Girls	57 (1.5)	341 - 360
Boys	43 (1.5)	346 - 365
[2] **Nigeria**		
Girls	49 (1.8)	320 - 356
Boys	51 (1.8)	308 - 352
LaNA 2023 Average		
Girls	51 (0.8)	376 - 387
Boys	49 (0.8)	367 - 380



LaNA mathematics scale scores are presented on the TIMSS trend scale established in 1995 with a centerpoint of 500 located at the mean of the combined achievement distribution. The units of the scale were chosen so that 100 scale score points corresponded to the standard deviation of the distribution. The LaNA link is based on a limited number of easy TIMSS trend blocks. See Appendix A for more information.( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.
See Appendix B.2 for population coverage notes 1 and 2. See Appendix B.6 for sampling guidelines and the sampling the participation note †.

International Benchmarks of Mathematics Achievement

To describe the types of tasks that students at different points on the mathematics achievement scale can do, LaNA uses the defined points on the TIMSS achievement scale known as International Benchmarks: Low International Benchmark (400 scale score points), Intermediate International Benchmark (475), High International Benchmark (550), and Advanced International Benchmark (625).

The LaNA 2023 Linking Study extends reporting on the TIMSS scale by establishing and describing the new Basic International Benchmark (325). New item development in LaNA

focused on foundational mathematics skills to extend the TIMSS 2019 Grade 4 Mathematics achievement test and provide a better measure of achievement for students with scale scores below the TIMSS Low International Benchmark, a region of the achievement scale previously uncategorized in TIMSS. The new LaNA Basic Benchmark summarizes the content knowledge and cognitive processes that characterize students' proficiency at that level based on the items in the LaNA Linking Study. Appendix E of this report contains further details on the analysis and procedures resulting in these classifications.

Mathematics

Exhibit 1.1.3a: Description of the LaNA 2023 Basic International Benchmark of Mathematics Achievement

	LaNA 2023 **Basic** International Benchmark
	Summary *Students show emerging understanding of calculations and can work with simple whole numbers.* Students can add and subtract quantities when presented in mathematical contexts; they have some understanding of multiplication; and they can compare whole numbers. Students can identify shapes and have some knowledge about basic properties of common shapes. They can read straightforward data representations.
325	Students can perform basic arithmetic. They can add two numbers up to three digits, or three numbers up to two digits, including when set in a simple context; and they can subtract a single-digit number or a double-digit number from a double-digit number. Students can multiply a number up to three digits by a single-digit number; and they can divide a 2-digit number by one of its single-digit factors. Students can order 3-digit numbers, relate representations of whole numbers in words and digits, and find a missing number in an arithmetic sequence. They can identify the missing number in a number sentence involving single-digit subtraction or multiplication. Students can use simple line and geometric properties to visually distinguish between two-dimensional shapes such as a square, triangle, or circle. They can calculate the perimeter of a triangle given its side lengths, determine the area of a rectangle on a unit grid, and visually compare angles in two-dimensional shapes and volumes of cylinders. Students can identify quantities represented in bar graphs and pictograms.

Exhibit 1.1.3a includes the full description of the Basic International Benchmark that summarizes the mathematics tasks students reaching this benchmark will likely solve successfully. A longer description of the benchmark is provided below the horizontal line, and a condensed version is provided in a summary above. Students at this new benchmark level can do straightforward arithmetic and show some emergent understanding in the mathematics content domains: number, measurement and geometry, and data. This benchmark description captures mathematics knowledge and skills that were among the most straightforward across LaNA items: calculations or simple applications of mathematical knowledge. While all three content domains are represented in the Basic International Benchmark description, not all cognitive processes can be expected to be reflected in the descriptions of the lower benchmarks. The cognitive domains have an association with item level difficulty such that, while all TIMSS cognitive domains are represented within the entire LaNA item pool, it is expected that they cannot be included in the new Basic International Benchmark.

Exhibit 1.1.3b summarizes what students attaining each TIMSS 2019 International Benchmark can likely do. The benchmark summaries offer an overview of the mathematics understanding demonstrated by students who performed at or above each International Benchmark on the mathematics achievement scale.

Reviewing the International Benchmarks allows for a deeper understanding of the development of mathematics skills. Overall, the benchmark descriptions outline a progression in mathematics achievement. Across the benchmarks for mathematics, the skills described range from demonstrating foundational mathematics knowledge at the Basic International Benchmark to applying and justifying their mathematical understanding across problem contexts at the Advanced International Benchmark. As written, the benchmarks and their descriptions generalize to a broad range of mathematics problems represented in LaNA and TIMSS.

Mathematics

Exhibit 1.1.3b: Summary of TIMSS 2019 International Benchmarks of Mathematics Achievement

TIMSS 2019 **Advanced** International Benchmark	
625	*Students can apply their understanding and knowledge in a variety of relatively complex situations and explain their reasoning.* Students can solve a variety of multistep word problems involving whole numbers and show an understanding of fractions and decimals. They can apply knowledge of two- and three-dimensional shapes in a variety of situations. Students can interpret and represent data to solve multistep problems.
TIMSS 2019 **High** International Benchmark	
550	*Students apply conceptual understanding to solve problems.* They can apply conceptual understanding of whole numbers to solve two-step word problems. They show understanding of the number line, multiples, factors, and rounding numbers, and operations with fractions and decimals. Students can solve simple measurement problems. They demonstrate understanding of geometric properties of shapes and angles. Students can interpret and use data in tables and a variety of graphs to solve problems.
TIMSS 2019 **Intermediate** International Benchmark	
475	*Students can apply basic mathematical knowledge in simple situations.* They can compute with three- and four-digit whole numbers in a variety of situations. They have some understanding of decimals and fractions. Students can identify and draw shapes with simple properties. They can read, label, and interpret information in graphs and tables.
TIMSS 2019 **Low** International Benchmark	
400	*Students have some basic mathematical knowledge.* They can add, subtract, multiply, and divide one- and two-digit whole numbers. They can solve simple word problems. They have some knowledge of simple fractions and common geometric shapes. Students can read and complete simple bar graphs and tables.

Exhibit 1.1.4a provides graphical representations alongside the percentages of students in each country reaching the LaNA 2023 Basic International Benchmark for mathematics and each of the International Benchmarks established in TIMSS 2019. The percentages do not add up to 100% by design. The benchmark percentages indicate cumulative proportions:

The lower the benchmark, the higher the proportion of students reaching that benchmark. Additionally, if students reached a higher benchmark, they also reached each of the lower benchmarks. The exhibit also provides the standard error (in parentheses) associated with each estimated percentage. As a reference point, the exhibit reports the median percentage of students reaching each international benchmark across participating countries. By definition, half of the countries have percent values above the median, and half are below the median.

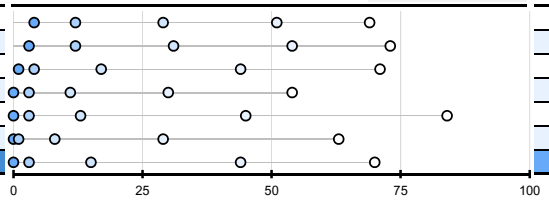
Across the LaNA 2023 Linking Study countries, the median percentages of students reaching each International Benchmark were 0 percent at Advanced, 3 percent at High, 15 percent at Intermediate, 44 percent at Low, and 70 percent at Basic. Students who reached the Advanced Benchmark also surpassed the High, Intermediate, Low, and Basic International Benchmarks. Students who reach a particular international benchmark can do what is described in Exhibits 1.1.3a and 1.1.3b for that benchmark.

Exhibit 1.1.4b is given for comparison only and presents the data provided in Exhibit 1.8 from the *TIMSS 2019 International Results*. Note that the LaNA 2023 Linking Study and TIMSS 2019 were different studies based on different student samples and countries. The exhibit has been updated to include the percentages of TIMSS 2019 students who would reach the Basic International Benchmark.

Mathematics

Exhibit 1.1.4a: Percentages of Students Reaching International Benchmarks of Mathematics Achievement

Country	Percentages of Students Reaching International Benchmarks (Advanced, High, Intermediate, Low, Basic)	Advanced Benchmark (625)	High Benchmark (550)	Intermediate Benchmark (475)	Low Benchmark (400)	Basic Benchmark (325)
Egypt		4 (0.6)	12 (1.3)	29 (2.1)	51 (2.3)	69 (2.2)
[1] Palestinian Nat'l Auth.		3 (1.0)	12 (1.3)	31 (1.9)	54 (2.0)	73 (1.7)
† Pakistan		1 ~	4 (0.8)	17 (2.0)	44 (2.7)	71 (2.2)
[2] Nigeria		0 ~	3 (0.7)	11 (2.0)	30 (3.1)	54 (3.7)
Burkina Faso		0 ~	3 (1.1)	13 (2.4)	45 (2.8)	84 (1.6)
Senegal		0 ~	1 ~	8 (1.3)	29 (2.2)	63 (2.0)
LaNA 2023 Median		**0**	**3**	**15**	**44**	**70**



( ) Standard errors appear in parentheses. Because of rounding some results may appear inconsistent.
A tilde (~) indicates result not reported because estimation is not reliable.
See Exhibits 1.1.3a and 1.1.3b for International Benchmark information. The Advanced, High, Intermediate, and Low Benchmark summaries come from TIMSS 2019. The Basic Benchmark description comes from the LaNA 2023 Linking Study.
See Appendix B.2 for population coverage notes 1 and 2. See Appendix B.6 for sampling guidelines and the sampling the participation note †.

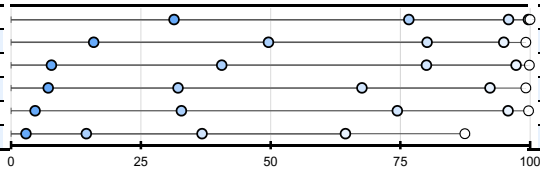
Exhibit 1.1.4b: Percentages of Students Reaching International Benchmarks of Mathematics Achievement from TIMSS 2019 (Including Basic Benchmark)

Country	Percentages of Students Reaching International Benchmarks (Advanced, High, Intermediate, Low, Basic)	Advanced Benchmark (625)	High Benchmark (550)	Intermediate Benchmark (475)	Low Benchmark (400)	Basic Benchmark (325)
[3] Singapore		54 (2.2)	84 (1.5)	96 (0.7)	99 ~	100 ~
† Hong Kong SAR		38 (1.9)	78 (1.6)	96 (0.7)	100 ~	100 ~
Korea, Rep. of		37 (1.4)	77 (1.2)	95 (0.5)	99 ~	100 ~
Chinese Taipei		37 (1.3)	78 (1.1)	96 (0.5)	100 ~	100 ~
Japan		33 (1.3)	74 (0.9)	95 (0.4)	99 ~	100 ~
† Northern Ireland		26 (1.4)	60 (1.4)	85 (1.1)	96 (0.6)	99 ~
[2] England		21 (1.4)	53 (1.5)	83 (1.2)	96 (0.5)	100 ~
[2] Russian Federation		20 (1.6)	61 (1.9)	91 (1.0)	99 ~	100 ~
Ireland		15 (1.0)	52 (1.4)	84 (1.0)	97 (0.5)	100 ~
[3] Türkiye (5)		15 (1.3)	43 (1.8)	70 (1.7)	88 (1.3)	97 (0.6)
[2] † United States		14 (0.8)	46 (1.3)	77 (1.1)	93 (0.6)	99 ~
[2] Lithuania		13 (1.1)	48 (1.6)	81 (1.1)	96 (0.6)	100 ~
† Norway (5)		13 (0.9)	48 (1.3)	82 (1.2)	97 (0.6)	100 ~
Cyprus		12 (0.9)	42 (1.6)	77 (1.3)	95 (0.6)	99 ~
[2] Latvia		11 (0.9)	50 (1.7)	85 (1.2)	98 ~	100 ~
Finland		11 (0.8)	42 (1.3)	78 (1.2)	95 (0.6)	99 ~
Czech Republic		10 (1.0)	42 (1.5)	78 (1.3)	96 (0.6)	100 ~
Australia		10 (0.9)	36 (1.2)	70 (1.3)	90 (1.0)	98 ~
Austria		9 (0.7)	45 (1.4)	84 (1.1)	98 ~	100 ~
Hungary		9 (0.8)	39 (1.4)	74 (1.3)	93 (0.8)	99 ~
[2] Portugal		9 (0.7)	39 (1.6)	74 (1.2)	95 (0.7)	100 ~
† Denmark		8 (0.9)	37 (1.3)	75 (1.0)	95 (0.5)	100 ~
† Belgium (Flemish)		8 (0.5)	40 (1.2)	80 (1.2)	97 (0.4)	100 ~
Bulgaria		8 (0.6)	37 (1.7)	71 (1.9)	90 (1.5)	98 ~
Poland		8 (0.8)	36 (1.4)	73 (1.4)	93 (0.6)	99 ~
Azerbaijan		8 (0.6)	36 (1.3)	72 (1.5)	92 (0.8)	97 (0.3)
Sweden		8 (0.8)	36 (1.7)	74 (1.4)	94 (0.7)	99 ~
≡ Netherlands		7 (0.9)	44 (1.7)	84 (1.1)	98 ~	100 ~
[2] Serbia		7 (0.7)	32 (1.4)	68 (1.5)	89 (1.1)	98 ~
United Arab Emirates		7 (0.3)	26 (0.6)	53 (0.8)	78 (0.7)	94 (0.4)
[1 2] Canada		6 (0.6)	32 (1.0)	69 (0.9)	92 (0.6)	99 ~
[2] New Zealand		6 (0.5)	25 (1.2)	56 (1.3)	83 (0.9)	96 (0.5)
Germany		6 (0.6)	36 (1.5)	75 (1.2)	96 (0.6)	100 ~
Albania		5 (0.6)	26 (1.4)	62 (1.8)	86 (1.3)	97 (0.6)
[2] Slovak Republic		5 (0.7)	31 (1.7)	71 (1.7)	91 (1.2)	98 ~
Malta		5 (0.5)	32 (0.9)	69 (0.8)	91 (0.6)	99 ~
North Macedonia		5 (0.8)	21 (1.8)	52 (2.4)	78 (1.7)	92 (1.1)
[2] Kazakhstan		5 (0.6)	29 (1.5)	71 (1.4)	95 (0.6)	100 ~
Bahrain		4 (0.4)	21 (1.0)	54 (1.2)	81 (1.0)	96 (0.6)
Italy		4 (0.5)	30 (1.5)	73 (1.3)	95 (0.5)	100 ~
Croatia		4 (0.6)	28 (1.3)	70 (1.5)	95 (0.7)	99 ~
Spain		4 (0.4)	27 (0.9)	65 (1.3)	91 (1.0)	99 ~
France		3 (0.5)	21 (1.2)	57 (1.6)	85 (1.2)	97 (0.5)
Oman		3 (0.8)	12 (1.3)	33 (1.5)	62 (1.3)	85 (0.9)
[1] Georgia		3 (0.4)	20 (1.4)	56 (2.0)	84 (1.4)	97 (0.7)
Armenia		3 (0.5)	23 (1.4)	64 (1.6)	92 (0.7)	99 ~
Qatar		2 ~	14 (1.2)	40 (1.6)	70 (1.4)	91 (1.0)
Iran, Islamic Rep. of		2 ~	13 (1.0)	39 (1.6)	68 (1.5)	89 (1.2)
Montenegro		1 ~	11 (0.7)	43 (0.9)	76 (0.9)	92 (0.5)
Morocco		1 ~	6 (1.1)	18 (1.4)	43 (1.7)	71 (1.5)
South Africa (5)		1 ~	5 (0.5)	16 (1.1)	37 (1.5)	66 (1.4)
[2] Saudi Arabia		1 ~	6 (0.6)	23 (1.2)	51 (1.4)	77 (1.2)
Kuwait		1 ~	6 (0.9)	21 (1.6)	47 (1.8)	70 (1.6)
[2] Kosovo		1 ~	8 (0.8)	37 (1.5)	73 (1.4)	92 (0.8)
Chile		1 ~	7 (0.6)	33 (1.4)	70 (1.5)	94 (0.8)
[1] Bosnia & Herzegovina		1 ~	9 (0.7)	40 (1.5)	76 (1.1)	94 (0.5)
[2 ψ] Pakistan		0 ~	1 ~	8 (1.5)	27 (4.7)	52 (5.5)
[2 ψ] Philippines		0 ~	1 ~	6 (0.8)	19 (1.8)	40 (2.5)
TIMSS 2019 Median		**7**	**34**	**71**	**92**	**99**

Mathematics

Exhibit 1.1.4b: Percentages of Students Reaching International Benchmarks of Mathematics Achievement from TIMSS 2019 (Including Basic Benchmark) (Continued)

Benchmarking Participants						
Moscow City, Russian Federation	31 (1.5)	77 (1.4)	96 (0.5)	100 ~	100 ~
[2] Dubai, UAE	16 (0.9)	50 (0.9)	80 (0.8)	95 (0.5)	99 ~
Quebec, Canada	8 (0.8)	41 (1.4)	80 (1.3)	97 (0.5)	100 ~
[2] Ontario, Canada	7 (1.0)	32 (1.8)	68 (1.6)	92 (0.9)	99 ~
Madrid, Spain	5 (0.5)	33 (1.2)	74 (1.5)	96 (0.6)	100 ~
Abu Dhabi, UAE	3 (0.2)	15 (0.6)	37 (1.0)	64 (1.1)	87 (0.8)



( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.
A tilde (~) indicates result not reported because estimation is not reliable.
See Appendix B.2 in the *TIMSS 2019 International Results in Mathematics and Science* for target population coverage notes 1, 2, and 3. See Appendix B.5 in the *TIMSS 2019 International Results in Mathematics and Science* for sampling guidelines and sampling participation notes †, ‡, and ≡.
Ψ Reservations about reliability because the percentage of students with achievement too low for estimation exceeds 15% but does not exceed 25%.

Reading Achievement

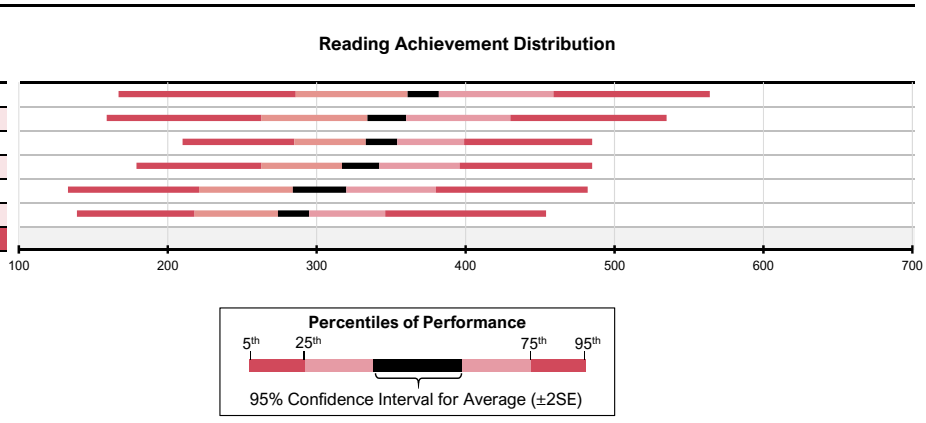
Reading Scale Score Distributions

Exhibit 1.2.1 shows the 95% confidence intervals for average reading achievement at the fifth and sixth grades for the six participating countries. Additionally, it includes a visualization of the achievement distribution, showing the percentiles of the distribution and the 95% confidence interval for the average (LaNA 2023 Average) of the participating countries.

Reading

Exhibit 1.2.1: Reading Scale Score Distributions

Country	95% Confidence Interval for Average Reading Scale Score
[1] Palestinian Nat'l Auth.	360 - 381
Egypt	335 - 361
Burkina Faso	334 - 355
[†] Pakistan	318 - 343
[2] Nigeria	283 - 320
Senegal	274 - 295
LaNA 2023 Average	**325 - 335**



LaNA reading scale scores are presented on the PIRLS trend scale established in 2001 with a centerpoint of 500 located at the mean of the combined achievement distribution. The units of the scale were chosen so that 100 scale score points corresponded to the standard deviation of the distribution. The LaNA link is based on a limited number of easy PIRLS trend passages. See Appendix A for more information.
See Appendix B.2 for population coverage notes 1 and 2. See Appendix B.6 for sampling guidelines and the sampling the participation note †.

The six countries are presented sorted by the location of their confidence intervals. The confidence interval upper boundaries of all countries are below 400 (381 for the Palestinian National Authority, 361 for Egypt, 355 for Burkina Faso, 343 for Pakistan, 320 for Nigeria, and 295 for Senegal, respectively). There is an overlap between the confidence intervals for Egypt,

Burkina Faso, and Pakistan, as well as among the confidence intervals for Nigeria and Senegal. There is only a small overlap between the confidence intervals of the Palestinian National Authority and Egypt (by 1 score point), and between the confidence intervals of Pakistan and Nigeria (by 2 score points).

While comparing the confidence intervals of the average across countries is informative, the LaNA 2023 Linking Study data also provides information about the observed range of student achievement. The achievement distribution on the right-hand side of Exhibit 1.2.1 shows the range of student achievement observed in each country. Similar to the results in mathematics (see Exhibit 1.1.1), the Palestinian National Authority and Egypt have the widest observed distributions ranging from below 200 to above 500 for the fifth and ninety-fifth percentiles. The twenty-fifth and seventy-fifth percentile scores for these two countries range from below 300 to above 400 scale score points. Somewhat more narrow achievement ranges are observed in the other four countries; the top five percent of students (those above the 95th percentile) in all four countries reach between 450 and 500 on the reading scale, while the lower end of the achievement range is somewhat more staggered.

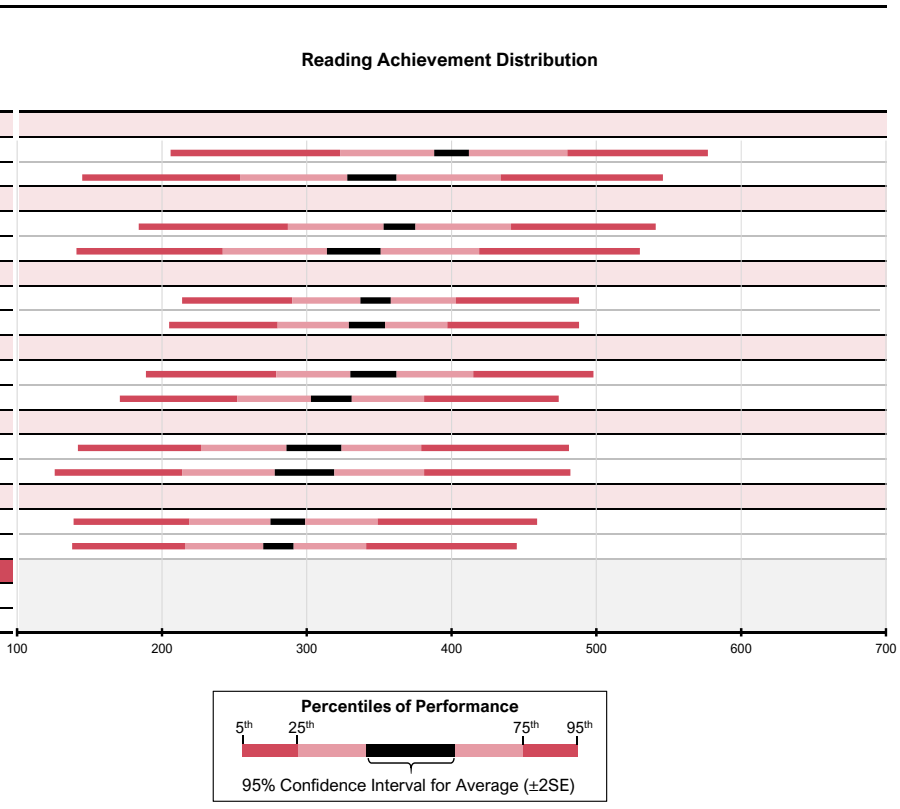
Later sections of the repoort further enrich and contextualize the results by presenting the associations between achievement in reading and various contextual factors, providing a deeper understanding of how achievement is related to other factors in each participating country.

Reading Scale Score Distributions for Girls and Boys

Exhibit 1.2.2 illustrates the differences in achievement distributions between girls and boys across the six participating countries. The confidence intervals for boys and girls overlap in most countries except the Palestinian National Authority, where the lower boundary of the confidence interval for girls is 25 points above the upper boundary for boys, and Egypt, with a difference of 2 points between the lower CI boundary for girls and the upper CI boundary for boys. In Pakistan, the 95% confidence intervals for boys and girls overlap only by 2 points. In all countries, the confidence intervals for girls are located higher than for boys, with the largest differences being shown for the Palestinian National Authority (50 points difference for the upper boundary), Pakistan (30 points difference), and Egypt (24 points difference), and Pakistan. The difference in the upper confidence interval boundaries between girls and boys in other countries is below 10 points. As with the mathematics achievement results, the percentages of girls and boys attending school differs across countries. This, in part, may lead to differences in results.

Exhibit 1.2.2: Reading Achievement Scale Score Distributions for Girls and Boys

	Percent of Students	95% Confidence Interval for Average Reading Scale Score
[1] **Palestinian Nat'l Auth.**		
Girls	47 (1.7)	387 - 412
Boys	53 (1.7)	328 - 362
Egypt		
Girls	48 (2.2)	353 - 375
Boys	52 (2.2)	314 - 351
Burkina Faso		
Girls	56 (0.8)	337 - 358
Boys	44 (0.8)	328 - 353
† **Pakistan**		
Girls	46 (2.6)	330 - 362
Boys	54 (2.6)	303 - 332
[2] **Nigeria**		
Girls	49 (1.8)	286 - 325
Boys	51 (1.8)	277 - 318
Senegal		
Girls	57 (1.5)	276 - 299
Boys	43 (1.5)	270 - 291
LaNA 2023 Average		
Girls	51 (0.8)	336 - 347
Boys	49 (0.8)	313 - 326



LaNA reading scale scores are presented on the PIRLS trend scale established in 2001 with a centerpoint of 500 located at the mean of the combined achievement distribution. The units of the scale were chosen so that 100 scale score points corresponded to the standard deviation of the distribution. The LaNA link is based on a limited number of easy PIRLS trend passages. See Appendix A for more information.
( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.
See Appendix B.2 for population coverage notes 1 and 2. See Appendix B.6 for sampling guidelines and the sampling the participation note †.

International Benchmarks of Reading Achievement

To describe the types of tasks that students at different points on the reading achievement scale can do, LaNA uses the defined points on the PIRLS achievement scale known as International Benchmarks: Low International Benchmark (400 scale score points), Intermediate International Benchmark (475), High International Benchmark (550), and Advanced International Benchmark (625).

The LaNA 2023 Linking Study extends reporting on the PIRLS scale by establishing and describing the new Basic International Benchmark (325). The addition of this new benchmark provides an opportunity to understand the emerging reading skills of students with scale scores below the PIRLS Low International Benchmark—performance previously uncategorized in PIRLS.

The International Benchmark descriptions were developed separately for the two overall reading purposes defined in the PIRLS assessment framework: Literary and Informational. Texts that characterize each purpose are integral to the benchmark descriptions because key

features of a text (overall length, structure, sentence complexity, and vocabulary) are associated with reading comprehension. The LaNA literary texts are complete short stories or episodes with supportive illustrations. The texts include contemporary and traditional stories with one or two main characters, a simple plot, and an explicit overall theme or message. The literary texts are relatively easy and accessible, approximately 500 words long, with a clear linear structure, explicit meanings, and simply described characters. The language features everyday vocabulary and straightforward sentence structures.

The LaNA informational texts include a variety of expository texts about the natural world. These texts also present information in diagrams, illustrations, or tables. Texts are structured in a number of ways, including by logic, argument, chronology, and topic. Several include organizational features such as subheadings or text boxes. The informational texts are approximately 500 words in length with a clear structure and explicit meanings, and straightforward sentences with everyday vocabulary.

There are a few key differences between the reading comprehension questions in the LaNA reading assessment compared to the questions in PIRLS 2021.[1] The questions in LaNA include a different proportion of items across the comprehension processes than in PIRLS. Due to the need to ask easier questions, LaNA places greater emphasis on straightforward retrieval of information than PIRLS and less on straightforward inferencing, interpreting, and integrating ideas and information and on evaluating and critiquing content and textual elements.

Exhibit 1.2.3a includes the full description of the Basic International Benchmark of reading achievement. It summarizes the proficiencies in reading comprehension that students reaching this benchmark demonstrate. Students at this new benchmark level can retrieve details from a well-defined section within a short literary or informational text and make simple inferences about characters or actions when reading literary texts.

Reading

Exhibit 1.2.3a: Description of the LaNA 2023 Basic International Benchmark of Reading Achievement

○ LaNA 2023 **Basic** International Benchmark		
325	**Literary**	**Informational**
	When reading LaNA literary texts, students can: • Retrieve and recognize explicitly stated details and information from a limited section of the text. • Make simple straightforward inferences to recognize a character's action or feeling.	When reading LaNA informational texts, students can: • Retrieve and recognize explicitly stated details from a confined section of the text. • Make simple straightforward inferences about an action or a process.

1 For more detailed information about the digital PIRLS texts see the *PIRLS 2021 International Results*. For more details about the paper-based LaNA reading texts see Appendix A in this report.

Exhibit 1.2.3b: Summary of PIRLS 2021 International Benchmarks of Reading Achievement

	Literary	Informational
PIRLS 2021 Advanced International Benchmark		
625	When reading predominantly difficult literary texts, students can: • Interpret and integrate story events and character actions to describe reasons, motivations, feelings, and character development. • Evaluate the intended effect of the author's language, style, and composition choices.	When reading predominantly difficult informational texts or online tasks, students can: • Make inferences about complex information across different web pages and parts of text to recognize the relevant information in a list and use evidence in the text to support ideas. • Interpret and integrate multiple pieces of different information across text and web pages to present an overview of ideas in the text and provide comparisons and explanations. • Evaluate textual, visual, and interactive elements to explain their purpose, and identify the writer's point of view and provide supporting evidence.
PIRLS 2021 High International Benchmark		
550	When reading medium and difficult literary texts, students can: • Locate and identify significant actions and details embedded across the text. • Make inferences about relationships between intentions, actions, events, and feelings. • Interpret and integrate story events to give reasons for character actions and feelings. • Recognize the meaning of some figurative language (e.g., metaphor, imagery).	When reading informational texts or online tasks of medium or high difficulty, students can: • Locate and identify relevant information in texts with a variety of features, such as diagrams and illustrations. • Make inferences to provide comparisons, descriptions, explanations, predictions, and choose a relevant website. • Interpret and integrate textual and visual information across texts and web pages to connect ideas, sequence events, identify characteristics, and provide explanations. • Evaluate the content to take and justify a position; describe how illustrations, diagrams, photographs, and maps convey and support content; and recognize the contribution of word choice in conveying the writer's point of view.
PIRLS 2021 Intermediate International Benchmark		
475	When reading literary texts of easy or medium difficulty, students can: • Locate, recognize, and reproduce explicitly stated actions, events, and feelings. • Make straightforward inferences about events and characters' actions. • Interpret reasons for characters' feelings or actions and identify supporting evidence.	When reading informational texts or online tasks of easy or medium difficulty, students can: • Locate, recognize, and reproduce explicitly stated information across texts. • Make straightforward inferences to provide comparisons, descriptions, and explanations. • Interpret and integrate to provide information about central ideas and reasons for actions, events, and outcomes.
PIRLS 2021 Low International Benchmark		
400	When reading predominantly easy literary texts, students can: • Locate, retrieve, and reproduce explicitly stated information, actions, or ideas. • Make simple, straightforward inferences about characters' actions.	When reading predominantly easy informational texts, students can: • Locate, retrieve, and reproduce explicitly stated information. • Make simple, straightforward inferences to provide a reason for an outcome.

Exhibit 1.2.3b summarizes what students attaining each of the PIRLS 2021 International Benchmarks can likely do. The benchmark summaries give an overview of the reading skills and strategies demonstrated by students who performed at or above each benchmark on the reading achievement scale. There are key differences between the digital texts used to assess

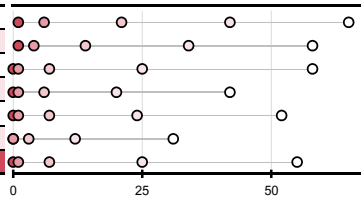
reading skills in PIRLS 2021 and the paper-based texts used in LaNA. These differences should be considered when comparing LaNA students reaching the PIRLS International Benchmarks with PIRLS 2021 students reaching these benchmarks.

Overall, the benchmark descriptions outline a progression in reading ability from more straightforward comprehension skills involving shorter texts with simpler text structure to more diverse skills involving texts of longer lengths. At the Basic International Benchmark, students are likely to be able to retrieve explicitly stated details and make straightforward inferences when reading predominantly easy literary or informational texts. Students can read more challenging texts at the subsequent international benchmarks, making more complex inferences and demonstrating greater proficiency in interpreting and integrating information. At the Advanced International Benchmark, students can interpret and evaluate predominantly difficult literary texts and make inferences, and integrate complex information, using evidence from the text to support ideas, within predominantly difficult informational texts.

Exhibit 1.2.4a provides graphical representations alongside the percentages of students in each country reaching the LaNA 2023 Basic International Benchmark for reading and the International Benchmarks established in PIRLS 2021. The percentages do not add up to 100% by design. The benchmark percentages indicate cumulative proportions: The lower the benchmark, the higher the proportion of students reaching that benchmark. Additionally, if students reached a higher benchmark, they also reached each of the lower benchmarks. The exhibit also provides the standard error (in parentheses) associated with each estimated percentage. As a reference point, the exhibit reports the median percentage of students reaching each international benchmark across participating countries. By definition, half of the countries have percent values above the median, and half are below the median.

Reading

Exhibit 1.2.4a: Percentages of Students Reaching International Benchmarks of Reading Achievement

Country	Percentages of Students Reaching International Benchmarks	Advanced Benchmark (625)	High Benchmark (550)	Intermediate Benchmark (475)	Low Benchmark (400)	Basic Benchmark (325)
[1] Palestinian Nat'l Auth.		1 ~	6 (0.9)	21 (1.5)	42 (1.7)	65 (1.8)
Egypt		1 ~	4 (0.6)	14 (1.3)	34 (2.1)	58 (2.4)
Burkina Faso		0 ~	1 ~	7 (1.4)	25 (2.3)	58 (2.4)
[2] Nigeria		0 ~	1 ~	6 (1.3)	20 (2.6)	42 (3.8)
† Pakistan		0 ~	1 ~	7 (1.3)	24 (2.5)	52 (3.0)
Senegal		0 ~	0 ~	3 (0.8)	12 (1.7)	31 (2.3)
LaNA 2023 Median		**0**	**1**	**7**	**25**	**55**

( ) Standard errors appear in parentheses. Because of rounding some results may appear inconsistent.
A tilde (~) indicates result not reported because estimation is not reliable.
See Exhibits 1.2.3a and 1.2.3b for International Benchmark information. The Advanced, High, Intermediate, and Low Benchmark summaries come from PIRLS 2021. The Basic Benchmark description comes from the LaNA 2023 Linking Study.
See Appendix B.2 for population coverage notes 1 and 2. See Appendix B.6 for sampling guidelines and the sampling the participation note †.

Across the LaNA 2023 Linking Study countries, the median percentages of students reaching each International Benchmark were 0 at Advanced, 1 percent at High, 7 percent at Intermediate, 25 percent at Low, and 55 at Basic. Students who reached the Advanced Benchmark also surpassed the High, Intermediate, Low, and Basic International Benchmarks. Students who reach a particular international benchmark can do what is described in Exhibits 1.2.3a and 1.2.3b for that benchmark.

Exhibit 1.2.4b is given for comparison only and presents the data provided in Exhibit 4.1 from the *PIRLS 2021 International Results*. Note that the LaNA 2023 Linking Study and PIRLS 2021 were different studies based on different student samples and countries. The exhibit has been updated to include the percentages of PIRLS 2021 students who would reach the Basic International Benchmark.

Exhibit 1.2.4b: Percentages of Students Reaching International Benchmarks of Reading Achievement From PIRLS 2021 (Including Basic Benchmark)

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 due to the COVID-19 Pandemic***

Country	Percentages of Students Reaching International Benchmarks (Advanced, High, Intermediate, Low, Basic)	Advanced Benchmark (625)	High Benchmark (550)	Intermediate Benchmark (475)	Low Benchmark (400)	Basic Benchmark (325)
[3] Singapore		35 (1.4)	71 (1.6)	90 (0.9)	97 (0.5)	99 ~
Ireland		27 (1.3)	67 (1.5)	91 (0.7)	98 ~	100 ~
[2 †] Northern Ireland		23 (1.2)	61 (1.5)	87 (1.0)	97 (0.5)	99 ~
[2 †] Hong Kong SAR		21 (1.4)	68 (1.8)	92 (1.0)	98 ~	100 ~
Russian Federation		21 (1.3)	63 (2.0)	89 (1.4)	98 ~	100 ~
England ⋈		18 (1.2)	57 (1.3)	86 (0.9)	97 (0.4)	100 ~
[2 ≡] United States		18 (2.1)	52 (3.2)	81 (2.9)	95 (1.5)	98 ~
Bulgaria		16 (0.8)	49 (1.4)	78 (1.4)	93 (0.9)	98 ~
[†] Croatia		15 (1.0)	56 (1.5)	88 (1.0)	98 ~	100 ~
[2] Sweden		15 (0.9)	50 (1.2)	81 (0.9)	95 (0.6)	99 ~
Finland		14 (1.0)	53 (1.4)	84 (1.0)	96 (0.5)	100 ~
Lithuania		14 (1.1)	54 (1.4)	86 (0.8)	97 (0.4)	100 ~
Australia ⋈		14 (0.7)	48 (1.3)	80 (1.0)	94 (0.5)	99 ~
Poland		14 (0.8)	52 (1.5)	85 (1.0)	97 (0.5)	100 ~
Hungary		13 (0.9)	49 (1.6)	79 (1.5)	94 (1.0)	99 ~
United Arab Emirates		12 (0.4)	34 (0.6)	58 (0.7)	75 (0.7)	87 (0.4)
[†] New Zealand		11 (0.8)	41 (1.1)	71 (1.0)	90 (0.6)	98 ~
Czech Republic		11 (0.8)	47 (1.3)	82 (1.0)	96 (0.5)	100 ~
[2 †] Denmark		11 (0.8)	48 (1.3)	81 (0.9)	96 (0.6)	100 ~
Norway (5)		11 (0.6)	47 (1.3)	81 (1.0)	96 (0.6)	100 ~
Chinese Taipei		10 (0.6)	50 (1.4)	85 (1.0)	97 (0.4)	100 ~
Macao SAR		9 (0.7)	45 (0.7)	82 (0.6)	96 (0.4)	99 ~
Latvia		8 (0.7)	40 (1.4)	78 (1.4)	94 (0.7)	99 ~
Germany		8 (0.7)	39 (1.2)	75 (1.1)	94 (0.5)	99 ~
[2] Italy		8 (0.6)	44 (1.5)	83 (1.0)	97 (0.3)	100 ~
[3] Israel ⋈		8 (0.7)	35 (1.1)	67 (1.1)	88 (0.8)	97 (0.4)
[†] Slovak Republic		8 (0.8)	42 (1.7)	79 (1.2)	94 (0.8)	98 ~
Malta		8 (0.6)	36 (1.3)	70 (1.4)	90 (0.9)	98 ~
Austria		7 (0.7)	41 (1.6)	80 (1.2)	96 (0.4)	100 ~
[2] Albania		7 (0.7)	33 (1.6)	69 (1.7)	92 (0.9)	99 ~
[≡] Netherlands		6 (0.7)	37 (1.4)	79 (1.3)	96 (0.7)	100 ~
Cyprus		6 (0.6)	32 (1.4)	69 (1.5)	92 (0.7)	99 ~
Qatar		6 (0.8)	27 (1.6)	57 (1.7)	80 (1.2)	93 (0.6)
[2] Portugal		6 (0.6)	36 (1.4)	75 (1.0)	94 (0.6)	99 ~
Spain		6 (0.6)	35 (1.3)	76 (1.2)	95 (0.6)	100 ~
[2] Türkiye		5 (0.5)	29 (1.3)	62 (1.7)	86 (1.2)	96 (0.5)
Bahrain		5 (0.7)	20 (1.1)	47 (1.2)	71 (0.9)	88 (0.6)
Slovenia		5 (0.5)	35 (1.1)	75 (1.1)	94 (0.5)	99 ~
[3] Serbia		5 (0.5)	33 (1.6)	73 (1.7)	93 (0.8)	99 ~
France		5 (0.6)	32 (1.5)	72 (1.4)	94 (0.7)	99 ~
[1] Georgia		4 (0.5)	25 (1.1)	61 (1.5)	87 (1.0)	97 (0.5)
Kazakhstan		4 (0.5)	28 (1.3)	67 (1.4)	91 (0.8)	99 ~
Belgium (Flemish)		3 (0.5)	29 (1.5)	71 (1.4)	94 (0.5)	100 ~
Oman		3 (0.4)	13 (1.0)	35 (1.4)	62 (1.4)	83 (1.0)
[2] Belgium (French)		3 (0.5)	23 (1.1)	62 (1.6)	89 (0.9)	99 ~
[3] Montenegro		2 ~	21 (0.8)	59 (1.2)	87 (0.7)	97 (0.3)
[2 †] Brazil ⋈		2 ~	13 (1.0)	37 (1.6)	61 (1.9)	77 (1.9)
[3] Saudi Arabia		2 ~	12 (0.9)	41 (1.8)	71 (1.6)	90 (1.0)
North Macedonia		1 ~	11 (1.1)	38 (2.5)	70 (2.2)	89 (1.3)
Azerbaijan		1 ~	11 (0.9)	37 (1.7)	67 (1.5)	89 (0.9)
[2 ψ] Egypt		1 ~	5 (0.7)	19 (1.4)	45 (2.0)	69 (1.9)
Jordan		1 ~	5 (0.8)	22 (1.6)	47 (2.0)	69 (1.9)
[ж] South Africa ⋈		1 ~	3 (0.5)	9 (0.9)	19 (1.2)	37 (1.4)
Morocco		1 ~	5 (1.0)	17 (1.4)	41 (1.6)	67 (1.7)
Iran, Islamic Rep. of ⋈		1 ~	7 (0.6)	29 (1.5)	59 (2.0)	81 (1.8)
Uzbekistan		0 ~	7 (0.7)	34 (1.3)	70 (1.4)	90 (1.0)
[2] Kosovo		0 ~	5 (0.7)	27 (1.3)	62 (1.5)	87 (1.1)
PIRLS 2021 Median		**7**	**36**	**75**	**94**	**99**

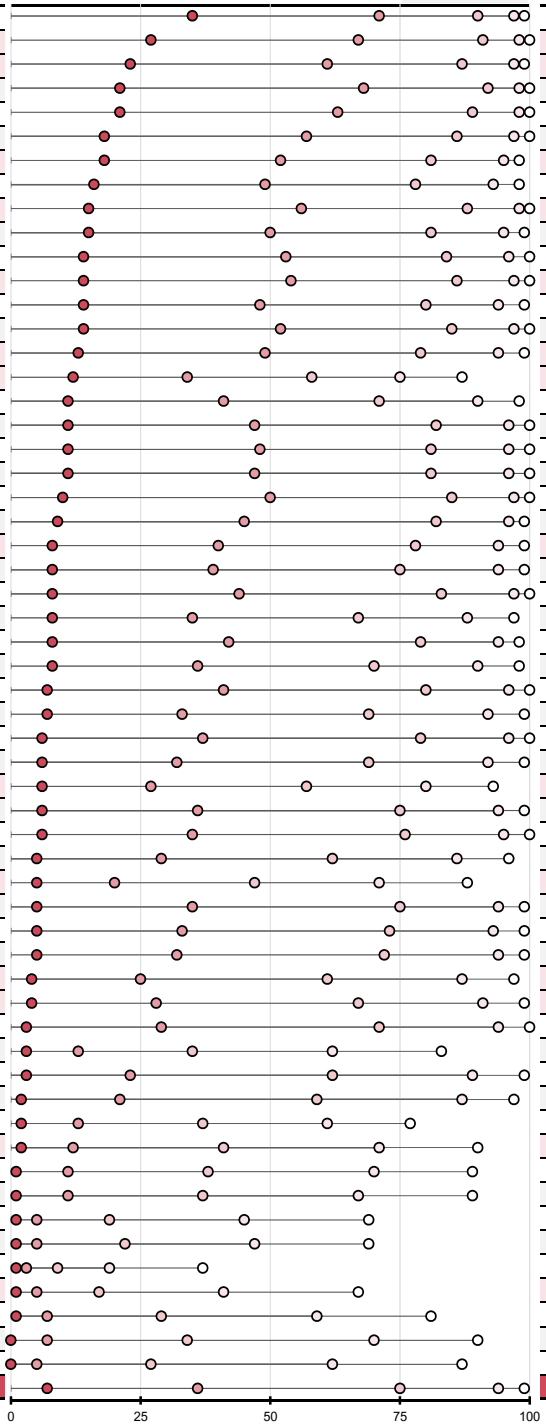


Exhibit 1.2.4b: Percentages of Students Reaching International Benchmarks of Reading Achievement From PIRLS 2021 (Including Basic Benchmark) (Continued)

Benchmarking Participants					
Moscow City, Russian Federation	35 (1.6)	79 (1.1)	96 (0.4)	100 ~	100 ~
[2] Dubai, UAE	24 (0.8)	56 (0.7)	80 (0.6)	92 (0.4)	97 (0.3)
[3] ≡ Alberta, Canada	12 (1.1)	47 (1.9)	80 (1.5)	95 (0.8)	99 ~
[2] British Columbia, Canada	12 (1.1)	45 (1.9)	79 (1.5)	94 (0.8)	99 ~
≡ Quebec, Canada	12 (1.3)	53 (1.9)	88 (1.0)	99 ~	100 ~
Abu Dhabi, UAE	9 (0.6)	26 (1.0)	45 (1.2)	61 (1.2)	75 (1.0)
[2] Newfoundland & Labrador, Canada	8 (1.0)	40 (1.8)	74 (1.6)	93 (0.7)	99 ~
South Africa (6) ⋈	3 (0.5)	11 (1.0)	25 (1.3)	44 (1.5)	65 (1.5)

( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.
A tilde (~) indicates result not reported because estimation is not reliable.
See Appendix A.1 in the *PIRLS 2021 International Results in Reading* for information on students assessed.
See Appendix A.2 in the *PIRLS 2021 International Results in Reading* for population coverage notes 1, 2, and 3. See Appendix A.5 in the *PIRLS 2021 International Results in Reading* for sampling guidelines and sampling participation notes †, ‡, and ≡.
Ψ Reservations about reliability because the percentage of students with achievement too low for estimation exceeds 15% but does not exceed 25%.
Ж Reservations about reliability because the percentage of students with achievement too low for estimation exceeds 25%.

Contexts for Student Learning

The LaNA 2023 Linking Study continues the tradition of TIMSS and PIRLS reporting on contextual factors related to students' learning experiences and how these factors relate to achievement. The items in the LaNA 2023 Linking Study context questionnaires are similar to those developed from the Context Questionnaire Frameworks of TIMSS and PIRLS and measure personal attitudes and environmental factors that are known to relate to learning outcomes. In some cases, multiple items are combined to form a scale such that responses to those items are combined into one overall scale score. To assist in interpretation, cut scores are provided for each scale to establish regions that can be described based on the item responses represented by a score on that scale. More details on creating these scales can be found in Appendix D of this report.

Selected results from the Student Questionnaire and School Questionnaire are presented in the subsequent exhibits. Each set of exhibits begins with a description of the item(s) or scale, followed by tables relating responses or response categories to the 95% confidence intervals (CIs) of average mathematics and reading achievement in each country and, on average, across countries. Short descriptions guide how to interpret the differences in the confidence intervals across item responses or scale categories. The similarity of achievement confidence intervals across item responses or scale categories can suggest relationships with achievement. Readers are encouraged to consult the introduction of this report for information about the interpretation of 95% CIs.

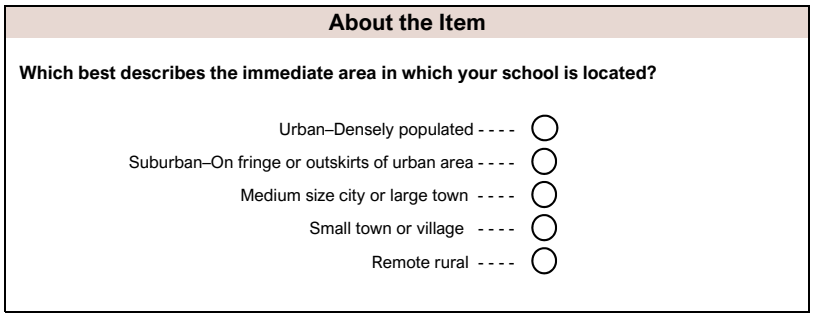
School Level Contexts

School Geographic Location

School location as reported by principals shows some relationship with average reading and mathematics achievement. School principals were asked to respond to an item describing their school's location (Exhibit 2.1.1), with five response categories ranging from "Urban—Densely populated" to "Remote rural." Across all countries, the largest percentages of students attended schools in densely populated urban areas or in small towns or villages.

Mathematics & Reading

Exhibit 2.1.1: School Geographic Location – Principals' Reports



Across the countries, on average, 23 percent of students were reported to attend a school in an "urban—densely populated" area, 11 percent of students were reported to attend a school in a "suburban—on fringe or outskirts of urban area," 19 percent of students were reported to attend a school in a "medium size city or large town," 34 percent of students were reported to attend a school in a "small town or village," and 12 percent of students were reported to attend a school in a "remote rural" area. In mathematics and reading, students who attended schools in densely populated urban areas tended to demonstrate higher average achievement than those in smaller towns or villages. In mathematics, the 95% confidence interval (CI) calculated across countries for students in "remote rural" areas is 335–372, while the corresponding interval for students in "urban—densely populated" areas ranges from 382–408 (see Exhibit 2.1.2). This lack of overlap indicates that there is likely a difference in average mathematics achievement between students attending schools in these locations. A similar pattern is observed in reading (see Exhibit 2.1.3). The 95% CI for students in "remote rural" areas is 285–320, while the corresponding CI for the "urban—densely populated" areas is 336–360, suggesting there is also likely a difference in average reading achievement between students attending schools in these regions.

Mathematics

Exhibit 2.1.2: School Geographic Location – Principals' Reports

Country	Urban–Densely Populated		Suburban–On Fringe or Outskirts of Urban Area		Medium Size City or Large Town		Small Town or Village		Remote Rural	
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement
Burkina Faso	28 (6.0)	392 - 433	13 (3.7)	365 - 408	13 (4.5)	366 - 410	28 (5.7)	361 - 391	18 (4.7)	365 - 435
Egypt	30 (3.5)	391 - 437	5 (2.2)	409 - 580	18 (4.1)	382 - 445	37 (4.2)	335 - 387	10 (3.2)	338 - 433
Nigeria	26 (5.8)	319 - 417	12 (5.3)	234 - 390	16 (5.8)	256 - 369	37 (7.7)	306 - 365	9 (3.4)	250 - 370
Pakistan	21 (3.7)	379 - 431	3 (1.3)	332 - 420	20 (3.9)	355 - 407	42 (4.3)	362 - 393	14 (2.8)	322 - 373
Palestinian Nat'l Auth.	15 (4.5)	362 - 436	11 (3.5)	382 - 453	36 (5.0)	400 - 437	36 (4.7)	377 - 418	2 ~	~
Senegal	20 (3.4)	348 - 398	22 (3.5)	374 - 409	13 (3.7)	327 - 385	24 (3.9)	308 - 336	21 (3.7)	303 - 346
LaNA 2023 Average	**23 (1.9)**	**382 - 408**	**11 (1.4)**	**374 - 418**	**19 (1.9)**	**365 - 392**	**34 (2.1)**	**353 - 370**	**12 (1.4)**	**335 - 372**

Reading

Exhibit 2.1.3: School Geographic Location – Principals' Reports

Country	Urban–Densely Populated		Suburban–On Fringe or Outskirts of Urban Area		Medium Size City or Large Town		Small Town or Village		Remote Rural	
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement
Burkina Faso	28 (6.0)	352 - 393	13 (3.7)	314 - 361	13 (4.5)	320 - 361	28 (5.7)	303 - 337	18 (4.7)	311 - 378
Egypt	30 (3.5)	352 - 390	5 (2.2)	347 - 499	18 (4.1)	332 - 383	37 (4.2)	295 - 337	10 (3.2)	298 - 387
Nigeria	26 (5.8)	285 - 366	12 (5.3)	203 - 371	16 (5.8)	228 - 345	37 (7.7)	273 - 336	9 (3.4)	223 - 335
Pakistan	21 (3.7)	323 - 371	3 (1.3)	284 - 392	20 (3.9)	307 - 353	42 (4.3)	313 - 351	14 (2.8)	272 - 325
Palestinian Nat'l Auth.	15 (4.5)	329 - 394	11 (3.5)	345 - 412	36 (5.0)	370 - 403	36 (4.7)	336 - 378	2 ~	~
Senegal	20 (3.4)	283 - 341	22 (3.5)	309 - 349	13 (3.7)	258 - 320	24 (3.9)	237 - 261	21 (3.7)	227 - 270
LaNA 2023 Average	**23 (1.9)**	**336 - 360**	**11 (1.4)**	**327 - 371**	**19 (1.9)**	**319 - 345**	**34 (2.1)**	**305 - 322**	**12 (1.4)**	**285 - 320**

( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.
A tilde (~) indicates result not reported because estimation is not reliable.

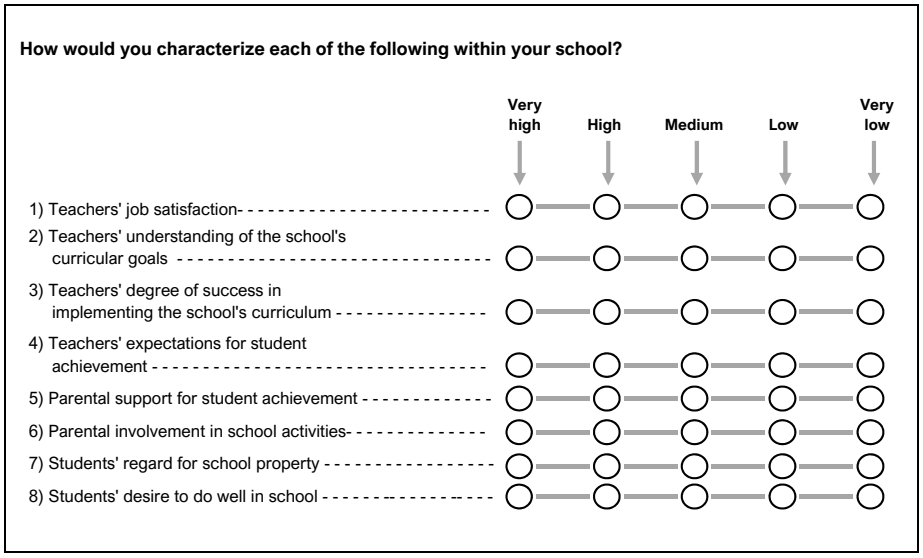
School Climate

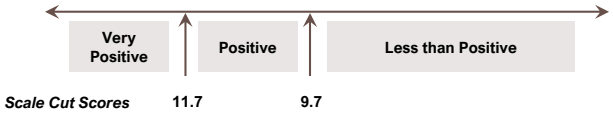
School climate as reported by principals has a modest positive relationship with average reading and mathematics achievement. Items in the *School Climate* scale (Exhibit 2.1.4) asked school principals about their perceptions of how different school community members (e.g., teachers, parents, students) exhibit dispositions and behaviors indicative of a positive school social environment. Based on these responses, students were classified as attending schools with "very positive," "positive," or "less than positive" school climate.

Exhibit 2.1.4: School Climate Scale Description – Principals' Reports

About the Scale

Students were scored according to their principals' responses characterizing eight statements on the *School Climate* scale. Cut scores divide the scale into three categories. Students in schools with a **Very Positive** school climate had a score at or above the cut score corresponding to their principal characterizing one statement as "very high" and the other seven as "high," on average. Students in schools with a **Less than Positive** school climate had a score at or below the cut score corresponding to their principal characterizing one statement as "low," two statements as "medium," and the other five as "high," on average. All other students were in schools with a **Positive** school climate.





Across the countries, on average, 17 percent of students attended schools where the principal reported a "very positive" school climate. Roughly half (51%) attended schools with a "less than positive" school climate, and 33 percent attended schools with a "positive school climate." Across the countries, average mathematics and reading achievement have a positive relation with school climate. The three groups can be characterized by, on average, increasing mathematics achievement when looking at students from schools with "less than positive," "positive," and "very positive" school climate. The 95% confidence interval (CI) calculated across countries for students in the "less than positive" is 361–377, while the corresponding range for the "very positive" group is 379–405 (see Exhibit 2.1.5). This provides evidence that attending a school with a "very positive" school climate is likely associated with, on average, higher mathematics achievement than attending a school with a less positive school climate. The 95% CI for average mathematics achievement among students in schools with "positive" school climate (367–390) overlaps with the CI for students in schools with "very positive" and "less than positive" school climates, which indicates that average mathematics achievement

for this group is likely located in between the two others. In reading, there is less difference between the 95% CIs of achievement associated with the three scale categories (328–357 for students in schools with "very positive" school climate, 321–343 for students in schools with "positive" school climate, and 316–331 for students in schools with "less than positive" school climate, see Exhibit 2.1.6). These findings support the statement that there is a general pattern of increasing achievement with more positive reports of school climate.

Mathematics

Exhibit 2.1.5: School Climate – Principals' Reports

Country	Very Positive School Climate		Positive School Climate		Less Than Positive School Climate		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	8 (2.6)	358 - 424	24 (4.8)	370 - 436	68 (5.5)	379 - 401	9.4 (0.14)
Egypt	34 (4.4)	395 - 444	33 (4.0)	366 - 426	32 (4.1)	342 - 391	10.8 (0.19)
Nigeria	16 (4.5)	320 - 393	34 (6.3)	275 - 363	51 (6.8)	316 - 360	10.0 (0.17)
Pakistan	24 (3.9)	358 - 400	30 (3.9)	371 - 411	46 (4.4)	355 - 388	10.3 (0.21)
Palestinian Nat'l Auth.	10 (2.9)	386 - 478	38 (5.1)	406 - 431	52 (5.7)	372 - 414	9.7 (0.19)
Senegal	8 (2.4)	351 - 399	37 (4.5)	328 - 362	55 (4.7)	341 - 368	9.7 (0.22)
LaNA 2023 Average	**17 (1.5)**	**379 - 405**	**33 (2.0)**	**367 - 390**	**51 (2.2)**	**361 - 377**	

Reading

Exhibit 2.1.6: School Climate – Principals' Reports

Country	Very Positive School Climate		Positive School Climate		Less Than Positive School Climate		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	8 (2.6)	282 - 396	24 (4.8)	323 - 385	68 (5.5)	329 - 352	9.4 (0.14)
Egypt	34 (4.4)	347 - 388	33 (4.0)	323 - 373	32 (4.1)	305 - 348	10.8 (0.19)
Nigeria	16 (4.5)	287 - 345	34 (6.3)	254 - 338	51 (6.8)	278 - 325	10.0 (0.17)
Pakistan	24 (3.9)	308 - 347	30 (3.9)	316 - 354	46 (4.4)	310 - 346	10.3 (0.21)
Palestinian Nat'l Auth.	10 (2.9)	359 - 435	38 (5.1)	370 - 396	52 (5.7)	336 - 377	9.7 (0.19)
Senegal	8 (2.4)	271 - 339	37 (4.5)	258 - 296	55 (4.7)	272 - 303	9.7 (0.22)
LaNA 2023 Average	**17 (1.5)**	**328 - 357**	**33 (2.0)**	**321 - 343**	**51 (2.2)**	**316 - 331**	

This context questionnaire scale was established based on the combined response distribution of countries that participated in the LaNA 2023 Linking Study. To provide a point of reference for country comparisons, the scale centerpoint of 10 was located at the mean of the combined distribution. The units of the scale were chosen so that 2 scale score points corresponded to the standard deviation of the distribution.
( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.

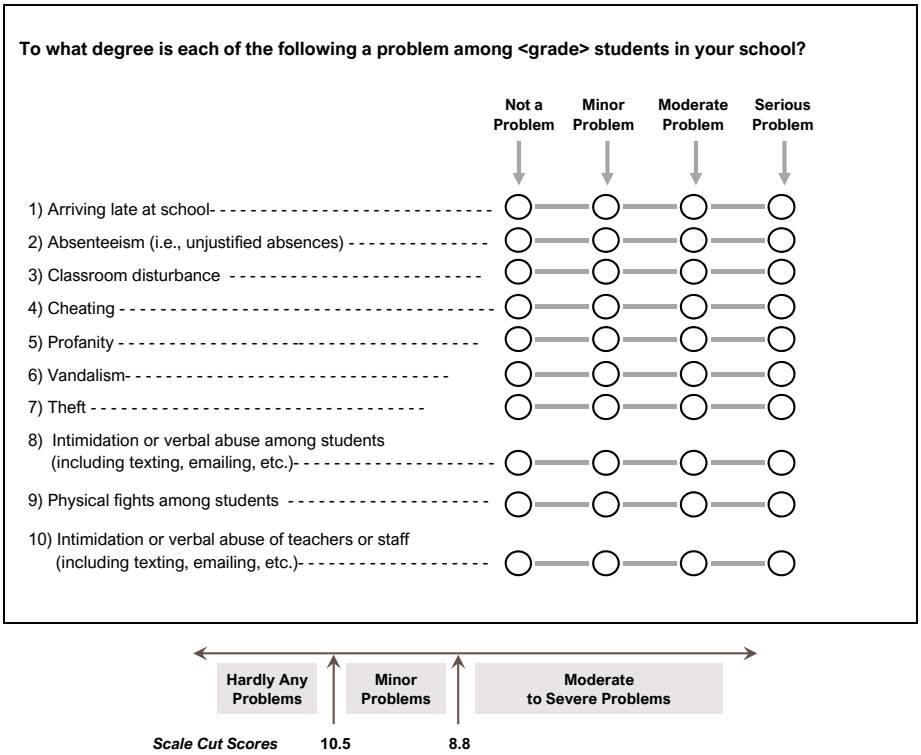
School Discipline

Similar to school climate, principals' reports of school discipline have a modest positive relationship with average reading and mathematics achievement. The *School Discipline* scale (Exhibit 2.1.7) items asked school principals about their perceptions of the extent to which 10 disciplinary issues were a problem in their school. Based on these responses, students were classified as attending schools with "hardly any problems," "minor problems," or "moderate to severe problems."

Exhibit 2.1.7: School Discipline Scale Description – Principals' Reports

About the Scale

Students were scored according to their principals' responses characterizing the severity of ten problems on the *School Discipline* scale. Cut scores divide the scale into three categories. Students in schools with **Hardly Any Problems** had a score at or above the cut score corresponding to their principals characterizing four problems as "not a problem," five problems as a "minor problem," and one problem as a "moderate problem." Students in schools with **Moderate to Severe Problems** had a score at or below the cut score corresponding to their principals characterizing three problems as a "severe problem," four problems as a "moderate problem," and three problems as a "minor problem." All other students were in schools with **Minor Problems**.



Across the countries, 45 percent of students attended schools where the principal reported "hardly any problems," and 31 percent attended schools with "minor problems" on average. Twenty-four percent of students were in schools with "moderate to severe problems" on average, though percentages by country range from 7 to 59 percent. On average, a positive relationship can be observed across countries between attending a school with fewer discipline issues and average achievement in mathematics and reading. While the 95% confidence intervals (CIs) of achievement among the three categories show some level of overlap, both lower and upper boundaries of the intervals generally increase from the group with "moderate to severe problems" to the CIs for "minor problems" and "hardly any problems" (Exhibit 2.1.8). A similar observation can be made in reading (Exhibit 2.1.9). The group consisting of students in schools with "hardly any problems" has 95% CI upper and lower boundaries located above those for the other two groups, indicating that more students with lower average achievement can be found in schools with higher levels of disciplinary problems.

Mathematics

Exhibit 2.1.8: School Discipline – Principals' Reports

Country	Hardly Any Problems		Minor Problems		Moderate to Severe Problems		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	18 (3.8)	358 - 429	22 (4.4)	369 - 424	59 (5.8)	381 - 403	8.3 (0.24)
Egypt	39 (5.1)	382 - 432	35 (5.3)	351 - 405	27 (3.8)	371 - 422	9.9 (0.18)
Nigeria	35 (7.1)	333 - 387	54 (7.9)	300 - 356	11 (4.2)	235 - 336	10.0 (0.15)
Pakistan	74 (3.3)	370 - 398	19 (2.6)	336 - 374	7 (2.4)	325 - 455	11.2 (0.11)
Palestinian Nat'l Auth. r	53 (6.6)	408 - 431	31 (6.2)	386 - 438	16 (4.8)	355 - 438	10.5 (0.22)
Senegal	51 (5.3)	340 - 362	23 (3.9)	320 - 359	26 (4.4)	339 - 395	10.2 (0.21)
LaNA 2023 Average	**45 (2.2)**	**377 - 395**	**31 (2.2)**	**358 - 378**	**24 (1.8)**	**355 - 388**	

Reading

Exhibit 2.1.9: School Discipline – Principals' Reports

Country	Hardly Any Problems		Minor Problems		Moderate to Severe Problems		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	18 (3.8)	315 - 385	22 (4.4)	315 - 368	59 (5.8)	330 - 354	8.3 (0.24)
Egypt	39 (5.1)	335 - 379	35 (5.3)	312 - 354	27 (3.8)	327 - 373	9.9 (0.18)
Nigeria	35 (7.1)	306 - 364	54 (7.9)	266 - 315	11 (4.2)	196 - 312	10.0 (0.15)
Pakistan	74 (3.3)	321 - 352	19 (2.6)	285 - 326	7 (2.4)	269 - 380	11.2 (0.11)
Palestinian Nat'l Auth. r	53 (6.6)	374 - 397	31 (6.2)	347 - 400	16 (4.8)	323 - 399	10.5 (0.22)
Senegal	51 (5.3)	271 - 295	23 (3.9)	252 - 290	26 (4.4)	268 - 333	10.2 (0.21)
LaNA 2023 Average	**45 (2.2)**	**332 - 350**	**31 (2.2)**	**310 - 329**	**24 (1.8)**	**306 - 338**	

This context questionnaire scale was established based on the combined response distribution of countries that participated in the LaNA 2023 Linking Study. To provide a point of reference for country comparisons, the scale centerpoint of 10 was located at the mean of the combined distribution. The units of the scale were chosen so that 2 scale score points corresponded to the standard deviation of the distribution.
An "r" indicates data are available for at least 70% but less than 85% of the students.
( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.

Instructional Resource Shortages

Students in schools where principals report increased impact of instructional resource shortages tend to have lower average reading and mathematics achievement. Items in the *Instructional Resource Shortages* scale (Exhibit 2.1.10) asked school principals to report on how much instruction was limited by shortages in instructional materials, school supplies, and qualified teachers. Based on these responses, students were classified as attending schools "not affected," "somewhat affected," or "affected a lot" by resource shortages.

Across the countries, 15 percent of students attended schools that principals reported were "not affected" by resource shortages, and 60 percent attended schools "somewhat affected," according to principal reports, on average. Twenty-five percent of students were in schools categorized as "affected a lot" by resource shortages. In both subjects, the data suggest a relationship between attending a school less affected by resource shortages and average

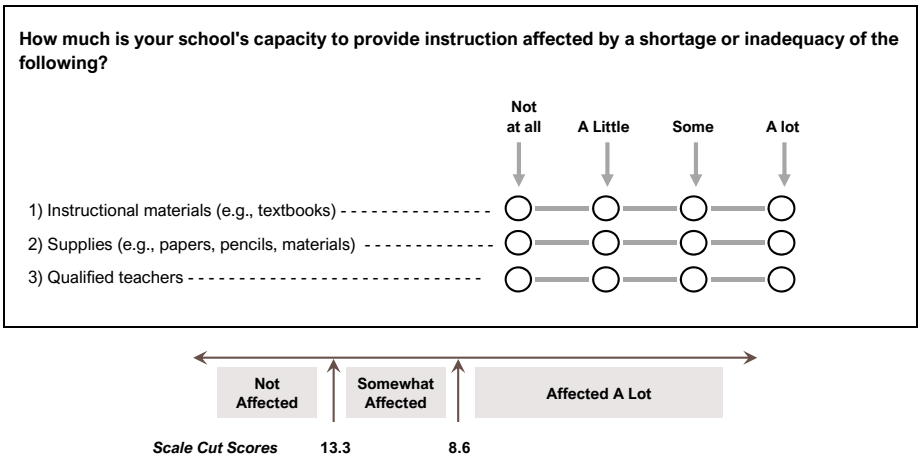
achievement across the countries. In mathematics, the 95% confidence interval (CI) for students in schools "not affected" by resource shortages is 383–411, while the CI for students in schools that were "somewhat affected" is 367–382 (Exhibit 2.1.11). The lack of overlap between these CIs provides evidence of a likely difference in average mathematics achievement between students in schools that were "not affected" by resource shortages and students in schools that were "somewhat affected" by resource shortages. A similar pattern is evident in reading (Exhibit 2.1.12). The 95% CI for average reading achievement is 336–364 for students in schools "not affected" by resource shortages and 319–334 for students in schools "somewhat affected" by resource shortages, also pointing to a likely difference in average reading achievement between students classified in these regions. The 95% CI for average reading achievement for students in schools "affected a lot" by resource shortages is 315–338.

Mathematics & Reading

Exhibit 2.1.10: Instructional Resource Shortages Scale Description – Principals' Reports

About the Scale

Students were scored according to their principals' reponses regarding three resources on the *Instruction Affected by Resource Shortages* scale. Cut scores divide the scale into three categories. Students in schools **Not Affected** by resource shortages had a score corresponding to their principal reporting that shortages affected instruction "not at all" for all three resources. Students in schools **Affected A Lot** by resource shortages had a score at or below the cut score corresponding to their principals reporting that shortage of one resource affected instruction "a lot," shortage of one resource affected instruction "some," and shortage of one resource affected instruction "a little." All other students attended schools **Somewhat Affected** by resource shortages.



Mathematics

Exhibit 2.1.11: Instructional Resource Shortages – Principals' Reports

Country	Not Affected		Somewhat Affected		Affected a Lot		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	7 (2.9)	383 - 422	62 (5.8)	375 - 405	32 (5.8)	376 - 416	9.5 (0.18)
Egypt	21 (4.2)	379 - 436	58 (5.0)	369 - 417	21 (4.5)	348 - 410	10.4 (0.23)
Nigeria	2 ~	~	75 (6.2)	312 - 357	23 (6.2)	310 - 376	9.4 (0.19)
Pakistan	13 (3.1)	359 - 432	42 (4.1)	352 - 390	44 (4.5)	369 - 396	9.2 (0.24)
Palestinian Nat'l Auth.	35 (5.6)	400 - 435	45 (6.3)	389 - 423	19 (5.1)	350 - 421	11.1 (0.27)
Senegal	13 (3.0)	318 - 407	77 (3.3)	341 - 362	10 (2.8)	325 - 368	10.4 (0.15)
LaNA 2023 Average	**15 (1.5)**	**383 - 411**	**60 (2.1)**	**367 - 382**	**25 (2.0)**	**361 - 383**	

Reading

Exhibit 2.1.12: Instructional Resource Shortages – Principals' Reports

Country	Not Affected		Somewhat Affected		Affected a Lot		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	7 (2.9)	336 - 386	62 (5.8)	322 - 352	32 (5.8)	330 - 368	9.5 (0.18)
Egypt	21 (4.2)	329 - 382	58 (5.0)	327 - 367	21 (4.5)	311 - 361	10.4 (0.23)
Nigeria	2 ~	~	75 (6.2)	278 - 323	23 (6.2)	274 - 357	9.4 (0.19)
Pakistan	13 (3.1)	314 - 368	42 (4.1)	304 - 346	44 (4.5)	316 - 345	9.2 (0.24)
Palestinian Nat'l Auth.	35 (5.6)	366 - 403	45 (6.3)	349 - 385	19 (5.1)	319 - 385	11.1 (0.27)
Senegal	13 (3.0)	257 - 359	77 (3.3)	271 - 293	10 (2.8)	256 - 296	10.4 (0.15)
LaNA 2023 Average	**15 (1.5)**	**336 - 364**	**60 (2.1)**	**319 - 334**	**25 (2.0)**	**315 - 338**	

This context questionnaire scale was established based on the combined response distribution of countries that participated in the LaNA 2023 Linking Study. To provide a point of reference for country comparisons, the scale centerpoint of 10 was located at the mean of the combined distribution. The units of the scale were chosen so that 2 scale score points corresponded to the standard deviation of the distribution.
( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.
A tilde (~) indicates result not reported because estimation is not reliable.

Highlights: School Contexts

On average, across the countries participating in the LaNA 2023 Linking Study:

- Students in schools where principals reported a positive school climate tend to have higher average mathematics and reading achievement than students in schools where principals reported a less positive school climate.
- Students in schools where principals reported fewer discipline problems tend to have higher average mathematics and reading achievement than students in schools where principals reported more discipline problems.

Student Attributes

Language Spoken at Home

Speaking the language of the test at home (at least some of the time) has a positive relationship with average reading and mathematics achievement. This relationship appears stronger for reading than for mathematics. Students reported how frequently they speak the language of the test at home: "always," "almost always," "sometimes," or "never" (Exhibit 2.2.1). The majority of students across countries, on average, reported speaking the language of the test at home at least "sometimes," and less than 15 percent of students across all countries, on average, reported that they "never" speak the language of the test at home.

Mathematics & Reading

Exhibit 2.2.1: Students Speak the Language of the Test at Home

About the Item

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

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

Across the countries, 39 percent of students reported "always" speaking the language of the test at home, 12 percent reported "almost always," 36 percent reported "sometimes," and 13 percent of students reported "never" speaking the language of the test at home, on average. In both mathematics and reading, students who reported "never" speaking the language of the test at home had the lowest average achievement. These students had a 95% confidence interval (CI) of 345–364 in mathematics and 287–307 in reading. Students who reported that they "sometimes" speak the language of the test at home were observed to have the highest average achievement in both subjects (95% CI of 388–403 for mathematics and 341–358 for reading; see Exhibits 2.2.2 and 2.2.3). The 95% CI for reading achievement of students who reported "never" speaking the language of the test at home does not overlap with the CIs of any other group; in mathematics the 95% CI for these students has upper and lower boundaries below the other groups. In summary, the data suggest that, on average, across all countries, no exposure to the language of the test at home is associated with lower average achievement. This pattern is not uniform across participating countries and is more evident in some countries than others. In addition, in five out of six countries, "sometimes" speaking the language of the test at home is associated with higher average achievement than other groups. However, the magnitude of the differences varies across countries.

Mathematics

Exhibit 2.2.2: Students Speak the Language of the Test at Home

Country	Always		Almost Always		Sometimes		Never	
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement
Burkina Faso	9 (1.0)	363 - 391	17 (1.5)	376 - 405	52 (2.2)	394 - 418	22 (2.2)	365 - 394
Egypt	76 (2.3)	384 - 413	13 (1.5)	380 - 452	8 (1.0)	376 - 427	3 (0.7)	378 - 427
Nigeria	28 (3.0)	314 - 370	15 (1.9)	273 - 336	47 (3.3)	336 - 372	10 (1.9)	265 - 331
Pakistan	30 (2.0)	353 - 393	12 (1.0)	361 - 399	40 (1.8)	384 - 410	18 (1.8)	349 - 377
Palestinian Nat'l Auth.	86 (1.7)	409 - 430	6 (0.8)	359 - 429	6 (0.8)	423 - 480	2 ~	~
Senegal	6 (0.7)	308 - 356	11 (0.9)	344 - 371	61 (2.0)	354 - 374	23 (2.1)	319 - 344
LaNA 2023 Average	**39 (0.8)**	**366 - 382**	**12 (0.5)**	**363 - 385**	**36 (0.8)**	**388 - 403**	**13 (0.7)**	**345 - 364**

Reading

Exhibit 2.2.3: Students Speak the Language of the Test at Home

Country	Always		Almost Always		Sometimes		Never	
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement
Burkina Faso	9 (1.0)	312 - 347	17 (1.5)	332 - 362	52 (2.2)	345 - 371	22 (2.2)	305 - 336
Egypt	76 (2.3)	340 - 366	13 (1.5)	325 - 380	8 (1.0)	341 - 391	3 (0.7)	313 - 363
Nigeria	28 (3.0)	286 - 338	15 (1.9)	238 - 290	47 (3.3)	304 - 344	10 (1.9)	224 - 295
Pakistan	30 (2.0)	319 - 354	12 (1.0)	317 - 350	40 (1.8)	325 - 357	18 (1.8)	290 - 322
Palestinian Nat'l Auth.	86 (1.7)	374 - 395	6 (0.8)	316 - 384	6 (0.8)	379 - 442	2 ~	~
Senegal	6 (0.7)	229 - 291	11 (0.9)	273 - 302	61 (2.0)	287 - 308	23 (2.1)	248 - 275
LaNA 2023 Average	**39 (0.8)**	**321 - 338**	**12 (0.5)**	**313 - 332**	**36 (0.8)**	**341 - 358**	**13 (0.7)**	**287 - 307**

( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.
A tilde (~) indicates result not reported because estimation is not reliable.

Home Resources and Digital Devices

Access to resources and digital devices at home has a positive relationship with average mathematics and reading achievement. Students were asked about their access to running water, electricity, a refrigerator, a mobile phone, and a computer or tablet at home (Exhibit 2.2.4). In two countries, only about half of the students reported having access to running water, electricity, and a refrigerator. More than two thirds of the students reported having access to a mobile phone in five countries. In three countries, less than half of the students reported having access to a computer or tablet.

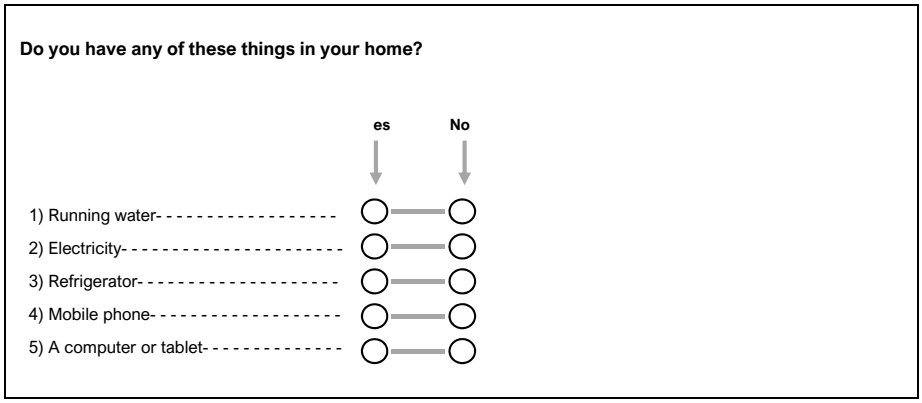
Across the countries, 74 percent of students reported having running water, 79 percent reported having electricity, 69 percent reported having a refrigerator, 80 percent reported having a mobile phone, and 48 percent reported having a computer or tablet at home, on average. In mathematics and reading, the data suggest that students with access to resources and digital devices at home have higher average achievement than their peers without access

to these resources. In terms of home resources, the 95% confidence intervals (CIs) of average achievement in both mathematics (Exhibit 2.2.5) and reading (Exhibit 2.2.6) are located higher for students who reported having access to running water, electricity, and a refrigerator than for students who reported not having access to these resources. Similarly, for digital devices, the 95% CIs of average achievement in both mathematics (Exhibit 2.2.7) and reading (Exhibit 2.2.8) are higher for students who reported having access to a mobile phone and a computer or tablet than for students who reported not having access to these devices.

Mathematics & Reading

Exhibit 2.2.4: Students' Home Resources and Digital Devices

About the Items



In mathematics and reading, the 95% CIs for the averages across countries for students reporting having access to resources do not overlap with the corresponding CIs for students who report not having access to resources. This suggests that there are likely achievement differences between these groups. It can be noted that digital devices and other material goods are likely proxy variables for constructs related to home resources known to be related to academic achievement, such as cultural capital and socioeconomic status.

Mathematics

Exhibit 2.2.5: Student Home Resources

Country	Running Water			Electricity			Refrigerator		
	Percent of Students Responding "Yes"	95% Confidence Interval for Average Achievement		Percent of Students Responding "Yes"	95% Confidence Interval for Average Achievement		Percent of Students Responding "Yes"	95% Confidence Interval for Average Achievement	
		Students Responding "Yes"	Students Responding "No"		Students Responding "Yes"	Students Responding "No"		Students Responding "Yes"	Students Responding "No"
Burkina Faso	45 (3.5)	390 - 421	376 - 397	47 (3.7)	390 - 421	375 - 399	29 (3.0)	388 - 428	381 - 401
Egypt	87 (1.6)	386 - 416	360 - 402	96 (0.7)	387 - 417	276 - 328	95 (0.7)	388 - 417	282 - 328
Nigeria	52 (2.8)	313 - 363	323 - 356	64 (3.5)	338 - 376	285 - 329	46 (3.3)	331 - 373	307 - 350
Pakistan	96 (0.5)	371 - 394	324 - 373	90 (1.0)	374 - 398	313 - 360	77 (1.5)	372 - 397	355 - 387
Palestinian Nat'l Auth.	80 (3.7)	402 - 427	385 - 426	97 (0.6)	405 - 428	262 - 325	97 (0.6)	404 - 428	263 - 332
Senegal	81 (2.8)	348 - 369	314 - 349	81 (2.6)	350 - 370	311 - 337	71 (2.1)	352 - 374	319 - 340
LaNA 2023 Average	**74 (1.1)**	**377 - 390**	**358 - 373**	**79 (1.0)**	**382 - 394**	**316 - 334**	**69 (0.9)**	**381 - 394**	**329 - 346**

Reading

Exhibit 2.2.6: Student Home Resources

Country	Running Water			Electricity			Refrigerator		
	Percent of Students Responding "Yes"	95% Confidence Interval for Average Achievement		Percent of Students Responding "Yes"	95% Confidence Interval for Average Achievement		Percent of Students Responding "Yes"	95% Confidence Interval for Average Achievement	
		Students Responding "Yes"	Students Responding "No"		Students Responding "Yes"	Students Responding "No"		Students Responding "Yes"	Students Responding "No"
Burkina Faso	45 (3.5)	348 - 378	321 - 340	47 (3.7)	350 - 380	318 - 342	29 (3.0)	349 - 389	327 - 346
Egypt	87 (1.6)	341 - 366	313 - 353	96 (0.7)	341 - 366	252 - 295	95 (0.7)	342 - 366	252 - 288
Nigeria	52 (2.8)	283 - 329	286 - 323	64 (3.5)	305 - 344	251 - 295	46 (3.3)	305 - 346	268 - 311
Pakistan	96 (0.5)	321 - 347	276 - 318	90 (1.0)	324 - 350	267 - 311	77 (1.5)	322 - 350	303 - 335
Palestinian Nat'l Auth.	80 (3.7)	366 - 389	346 - 393	97 (0.6)	369 - 391	221 - 273	97 (0.6)	369 - 391	226 - 280
Senegal	81 (2.8)	279 - 302	243 - 282	81 (2.6)	282 - 306	237 - 260	71 (2.1)	284 - 310	249 - 267
LaNA 2023 Average	**74 (1.1)**	**331 - 343**	**308 - 324**	**79 (1.0)**	**336 - 348**	**269 - 285**	**69 (0.9)**	**337 - 350**	**280 - 295**

( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.

Mathematics

Exhibit 2.2.7: Students' Home Digital Devices

Country	Mobile Phone			Computer or Tablet		
	Percent of Students Responding "Yes"	95% Confidence Interval for Average Achievement		Percent of Students Responding "Yes"	95% Confidence Interval for Average Achievement	
		Students Responding "Yes"	Students Responding "No"		Students Responding "Yes"	Students Responding "No"
Burkina Faso	55 (3.8)	388 - 415	375 - 400	24 (2.5)	392 - 433	381 - 399
Egypt	91 (0.9)	390 - 419	312 - 362	69 (1.5)	395 - 423	355 - 394
Nigeria	73 (2.8)	331 - 368	288 - 346	26 (2.0)	330 - 377	318 - 355
Pakistan	90 (0.9)	374 - 397	321 - 366	34 (1.9)	369 - 405	368 - 390
Palestinian Nat'l Auth.	89 (1.2)	404 - 428	361 - 407	74 (1.3)	412 - 437	366 - 393
Senegal	85 (1.9)	346 - 366	327 - 351	60 (2.2)	352 - 375	330 - 349
LaNA 2023 Average	**80 (0.9)**	**380 - 391**	**343 - 360**	**48 (0.8)**	**385 - 399**	**361 - 372**

Reading

Exhibit 2.2.8: Students' Home Digital Devices

Country	Mobile Phone			Computer or Tablet		
	Percent of Students Responding "Yes"	95% Confidence Interval for Average Achievement		Percent of Students Responding "Yes"	95% Confidence Interval for Average Achievement	
		Students Responding "Yes"	Students Responding "No"		Students Responding "Yes"	Students Responding "No"
Burkina Faso	55 (3.8)	340 - 366	323 - 349	24 (2.5)	348 - 393	330 - 348
Egypt	91 (0.9)	343 - 367	284 - 324	69 (1.5)	346 - 370	316 - 350
Nigeria	73 (2.8)	294 - 329	265 - 323	26 (2.0)	297 - 343	284 - 321
Pakistan	90 (0.9)	323 - 349	278 - 324	34 (1.9)	320 - 354	318 - 343
Palestinian Nat'l Auth.	89 (1.2)	369 - 391	324 - 364	74 (1.3)	374 - 397	336 - 362
Senegal	85 (1.9)	278 - 301	251 - 276	60 (2.2)	284 - 311	260 - 278
LaNA 2023 Average	**80 (0.9)**	**332 - 343**	**299 - 315**	**48 (0.8)**	**338 - 352**	**315 - 326**

( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.

Student Absences

More frequent absences are associated with lower average reading and mathematics achievement. Students were asked to report how often they are absent from school: "never or almost never," "once every two months," "once a month," "once every two weeks," or "once a week" (Exhibit 2.2.9). In three countries, more than 60 percent of students reported "never or almost never" being absent from school. In contrast, less than half of the students do so in the other three countries.

Across the countries, a little more than half of the students (55%) reported "never or almost never" being absent, on average. Seven percent on average reported being absent "once every two months," 9 percent on average reported being absent "once a month," 7 percent on average reported being absent "once every two weeks," and 22 percent on average reported being absent "once a week." In both mathematics and reading, the data suggest students who are absent "never or almost never" or "once every two months" have higher average achievement than their peers with more frequent absences. In mathematics, the 95%

confidence intervals (CIs) of average achievement for students who reported being absent more frequently ("once a week," "once every two weeks," or "once a month") are located lower than those for students who are only absent "once every two months" or "never or almost never." The upper boundaries of the 95% CIs of average mathematics achievement for students who reported being absent "once a week," "once every two weeks," and "once a month" are 362, 360, and 372, not reaching the lower boundaries of 376 for students who are absent "once every two months" and 386 for students who are absent "never or almost never." This means there is no overlap between the CIs of the three groups with frequent absences and the two with infrequent absences, indicating a likely difference in achievement (Exhibit 2.2.10). A similar pattern—a lack of overlap in the 95% CIs—is observed in reading (Exhibit 2.2.11). Students who reported being absent "once a week," "once every two weeks," or "once a month" had 95% CI upper boundaries of 318, 313, and 317, respectively, compared to lower boundaries 323 and 340 for students who reported being absent "once every two months" or "never or almost never," indicating that higher average achievement can be expected for students who are absent less frequently.

Mathematics & Reading

Exhibit 2.2.9: Frequency of Student Absences

About the Item

About how often are you absent from school?

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

Mathematics

Exhibit 2.2.10: Frequency of Student Absences

Country	Never or Almost Never		Once Every Two Months		Once a Month		Once Every Two Weeks		Once a Week	
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement
Burkina Faso	68 (2.4)	392 - 417	7 (0.9)	368 - 392	5 (0.7)	366 - 399	4 (0.6)	337 - 376	16 (1.3)	362 - 380
Egypt	39 (2.0)	390 - 425	4 (0.6)	401 - 457	7 (0.8)	354 - 404	10 (0.9)	372 - 420	40 (2.1)	375 - 411
Nigeria	41 (2.7)	336 - 373	10 (1.9)	337 - 386	11 (1.2)	286 - 352	11 (1.7)	261 - 351	27 (2.1)	306 - 349
Pakistan	41 (1.8)	381 - 410	8 (1.1)	359 - 417	17 (1.3)	363 - 394	8 (0.6)	348 - 380	26 (1.3)	348 - 376
Palestinian Nat'l Auth.	69 (1.6)	421 - 444	9 (0.9)	387 - 438	7 (0.5)	340 - 386	3 (0.3)	303 - 368	12 (1.0)	323 - 363
Senegal	73 (1.9)	348 - 368	7 (1.1)	326 - 367	6 (0.7)	336 - 368	4 (0.5)	317 - 354	9 (1.0)	320 - 352
LaNA 2023 Average	**55 (0.8)**	**386 - 398**	**7 (0.5)**	**376 - 396**	**9 (0.4)**	**353 - 372**	**7 (0.4)**	**338 - 360**	**22 (0.6)**	**348 - 362**

Reading

Exhibit 2.2.11: Frequency of Student Absences

Country	Never or Almost Never		Once Every Two Months		Once a Month		Once Every Two Weeks		Once a Week	
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement
Burkina Faso	68 (2.4)	342 - 367	7 (0.9)	315 - 341	5 (0.7)	304 - 340	4 (0.6)	298 - 337	16 (1.3)	314 - 336
Egypt	39 (2.0)	347 - 376	4 (0.6)	347 - 400	7 (0.8)	292 - 335	10 (0.9)	324 - 367	40 (2.1)	334 - 362
Nigeria	41 (2.7)	305 - 344	10 (1.9)	289 - 350	11 (1.2)	250 - 308	11 (1.7)	235 - 309	27 (2.1)	274 - 318
Pakistan	41 (1.8)	331 - 361	8 (1.1)	309 - 358	17 (1.3)	316 - 349	8 (0.6)	291 - 324	26 (1.3)	301 - 329
Palestinian Nat'l Auth.	69 (1.6)	387 - 408	9 (0.9)	344 - 396	7 (0.5)	301 - 345	3 (0.3)	274 - 343	12 (1.0)	286 - 322
Senegal	73 (1.9)	279 - 302	7 (1.1)	249 - 295	6 (0.7)	263 - 301	4 (0.5)	242 - 283	9 (1.0)	261 - 295
LaNA 2023 Average	**55 (0.8)**	**340 - 352**	**7 (0.5)**	**323 - 343**	**9 (0.4)**	**300 - 317**	**7 (0.4)**	**292 - 313**	**22 (0.6)**	**304 - 318**

( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.

Arriving at School Feeling Tired

Students who reported that they are often tired upon arriving at school appear to have lower average reading and mathematics achievement than their peers who arrive tired less often. Students were asked to report on how frequently they feel tired when they arrive at school: "every day," "almost every day," "sometimes," or "never" (Exhibit 2.2.12).

Mathematics & Reading

Exhibit 2.2.12: Students Report Arriving at School Feeling Tired

About the Item

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

I feel tired

Every day - - - - ◯
Almost every day - - - - ◯
Sometimes - - - - ◯
Never - - - - ◯

Across the countries, 31 percent of students reported "never" arriving at school feeling tired, 44 percent reported "sometimes" arriving tired, nine percent reported arriving tired "almost every day," and 16 percent reported arriving tired "every day," on average. In both mathematics and reading, the 95% confidence intervals (CIs) for average achievement of students who reported that they "never" or "sometimes" arrive at school feeling tired do not overlap with the CIs for students who reported that they arrive at school feeling tired "almost every day" or "every day." The lower boundaries of the 95% CIs for average mathematics achievement for students who reported "never" or "sometimes" feeling tired upon school arrival are 376 and 390, respectively. The corresponding upper boundaries for students who reported arriving at school tired "almost every day" or "every day" are 373 and 363 (Exhibit 2.2.13). A similar result is seen in reading (Exhibit 2.2.14). The lower boundaries of the 95% CIs for students who reported arriving at school feeling tired "never" or "sometimes," are 330 and 342, respectively and the corresponding upper boundaries for students who reported arriving at school tired "almost every day" or "every day" are 323 and 314, respectively. This lack of overlap in CIs between the groups who report feeling tired either more or less often when arriving at school provides evidence of a likely difference in average mathematics and reading achievement.

Mathematics

Exhibit 2.2.13: Students Report Arriving at School Feeling Tired

Country	Never		Sometimes		Almost Every Day		Every Day	
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement
Burkina Faso	45 (3.1)	382 - 410	36 (3.0)	391 - 422	9 (0.9)	367 - 394	10 (1.3)	350 - 375
Egypt	19 (2.0)	385 - 445	53 (2.1)	392 - 424	11 (1.1)	372 - 410	17 (1.2)	352 - 395
Nigeria	26 (2.1)	312 - 360	38 (2.4)	356 - 384	13 (1.6)	273 - 340	23 (2.7)	296 - 344
Pakistan	33 (2.3)	372 - 403	46 (2.0)	376 - 402	6 (0.8)	337 - 400	15 (1.3)	335 - 373
Palestinian Nat'l Auth.	25 (1.8)	415 - 439	40 (1.9)	418 - 447	10 (0.9)	383 - 419	24 (1.4)	364 - 395
Senegal	38 (2.6)	330 - 355	48 (2.3)	356 - 378	7 (0.7)	313 - 355	7 (0.9)	323 - 360
LaNA 2023 Average	**31 (1.0)**	**376 - 392**	**44 (0.9)**	**390 - 401**	**9 (0.4)**	**354 - 373**	**16 (0.6)**	**347 - 363**

Reading

Exhibit 2.2.14: Students Report Arriving at School Feeling Tired

Country	Never		Sometimes		Almost Every Day		Every Day	
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement
Burkina Faso	45 (3.1)	331 - 358	36 (3.0)	346 - 377	9 (0.9)	312 - 342	10 (1.3)	292 - 321
Egypt	19 (2.0)	348 - 400	53 (2.1)	345 - 370	11 (1.1)	317 - 355	17 (1.2)	310 - 345
Nigeria	26 (2.1)	273 - 326	38 (2.4)	325 - 354	13 (1.6)	253 - 308	23 (2.7)	259 - 302
Pakistan	33 (2.3)	324 - 354	46 (2.0)	322 - 354	6 (0.8)	278 - 326	15 (1.3)	293 - 330
Palestinian Nat'l Auth.	25 (1.8)	384 - 409	40 (1.9)	375 - 406	10 (0.9)	353 - 392	24 (1.4)	332 - 359
Senegal	38 (2.6)	258 - 286	48 (2.3)	289 - 313	7 (0.7)	250 - 291	7 (0.9)	251 - 288
LaNA 2023 Average	**31 (1.0)**	**330 - 345**	**44 (0.9)**	**342 - 354**	**9 (0.4)**	**306 - 323**	**16 (0.6)**	**300 - 314**

( ) Standard errors appear in parentheses. Because of rounding some results may appear inconsistent.

Arriving at School Feeling Hungry

Students who reported frequently feeling hungry upon arrival at school have lower average achievement in mathematics and reading compared to students who arrive hungry less often. Students were asked to report on how frequently they feel hungry when they arrive at school: "every day," "almost every day," "sometimes," or "never" (Exhibit 2.2.15).

Mathematics & Reading

Exhibit 2.2.15: Students Report Arriving at School Feeling Hungry

About the Item

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

I feel hungry

- Every day - - - - ◯
- Almost every day - - - - ◯
- Sometimes - - - - ◯
- Never - - - - ◯

Across the countries, 23 percent of students reported "never" arriving at school feeling hungry, 39 percent reported "sometimes" arriving hungry, 12 percent reported arriving hungry "almost every day," and 26 percent reported arriving hungry "every day" on average. In both mathematics and reading, the 95% confidence intervals (CIs) of average achievement for students who reported "never" or "sometimes" arriving at school hungry are located above those who reported arriving at school hungry "almost every day" or "every day," on average. In mathematics (Exhibit 2.2.16), the upper boundaries of the 95% CIs of average mathematics achievement for students who reported arriving at school hungry "almost every day" and "every day" are 381 and 380, respectively, while the lower boundaries for both students who reported "never" or "sometimes" arriving hungry is 383. A similar pattern is observed in reading (Exhibit 2.2.17). Students who reported arriving at school feeling hungry "almost every day" or "every day" had 95% CI upper boundaries of 331 and 334, respectively, while the lower boundary for both students who reported "never" or "sometimes" arriving at school feeling hungry is 335.

This lack of overlap in the 95% CIs in average mathematics and reading achievement provides evidence of a likely achievement difference between students who report rarely being hungry when arriving at school and those who report more frequently being hungry when they arrive at school.

Mathematics

Exhibit 2.2.16: Students Report Arriving at School Feeling Hungry

Country	Never		Sometimes		Almost Every Day		Every Day	
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement
Burkina Faso	30 (2.8)	377 - 416	43 (2.9)	392 - 416	12 (1.1)	372 - 394	15 (2.4)	364 - 397
Egypt	17 (1.9)	393 - 456	34 (1.9)	384 - 421	15 (1.3)	387 - 428	33 (1.9)	382 - 410
Nigeria r	20 (2.4)	319 - 367	40 (2.7)	337 - 379	15 (1.6)	303 - 360	25 (2.8)	306 - 350
Pakistan	19 (1.8)	379 - 422	28 (1.6)	374 - 399	9 (0.9)	340 - 381	44 (2.0)	366 - 394
Palestinian Nat'l Auth.	28 (2.2)	428 - 453	36 (1.9)	413 - 442	11 (1.0)	375 - 424	24 (1.6)	379 - 413
Senegal	22 (1.6)	335 - 355	50 (1.6)	346 - 370	12 (1.0)	338 - 367	16 (1.2)	337 - 373
LaNA 2023 Average	**23 (0.9)**	**383 - 400**	**39 (0.9)**	**383 - 396**	**12 (0.5)**	**364 - 381**	**26 (0.8)**	**366 - 380**

Reading

Exhibit 2.2.17: Students Report Arriving at School Feeling Hungry

Country	Never		Sometimes		Almost Every Day		Every Day	
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement
Burkina Faso	30 (2.8)	319 - 360	43 (2.9)	343 - 369	12 (1.1)	327 - 352	15 (2.4)	322 - 351
Egypt	17 (1.9)	355 - 407	34 (1.9)	339 - 368	15 (1.3)	340 - 377	33 (1.9)	333 - 358
Nigeria r	20 (2.4)	284 - 332	40 (2.7)	304 - 348	15 (1.6)	271 - 319	25 (2.8)	270 - 311
Pakistan	19 (1.8)	333 - 372	28 (1.6)	316 - 345	9 (0.9)	285 - 318	44 (2.0)	321 - 350
Palestinian Nat'l Auth.	28 (2.2)	391 - 414	36 (1.9)	374 - 406	11 (1.0)	339 - 385	24 (1.6)	350 - 382
Senegal	22 (1.6)	259 - 289	50 (1.6)	278 - 303	12 (1.0)	264 - 297	16 (1.2)	273 - 312
LaNA 2023 Average	**23 (0.9)**	**335 - 351**	**39 (0.9)**	**335 - 348**	**12 (0.5)**	**315 - 331**	**26 (0.8)**	**321 - 334**

An "r" indicates data are available for at least 70% but less than 85% of the students.
( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.

Students' Sense of School Belonging

Students' reports of their sense of school belonging have a positive relationship with average mathematics and reading achievement. The *Students' Sense of School Belonging* scale items (Exhibit 2.2.18) asked students to report the extent to which they agreed with three statements about their school, including that they like being in school. Based on their responses to these items, students were classified as having a "high sense of school belonging," "some sense of school belonging," or "little sense of school belonging."

Across the countries, 69 percent of students were classified as having "high sense of school belonging," and 25 percent were classified as having "some sense of school belonging," while only 7 percent of students were classified as having "little sense of school belonging," on average. A positive relationship can be observed between reports of higher school belonging and average achievement across countries for mathematics and reading. In mathematics, students reporting a "high sense of school belonging" have a 95% confidence interval (CI) higher than that of students reporting "some sense of school belonging" (384–396 vs. 368–382), which

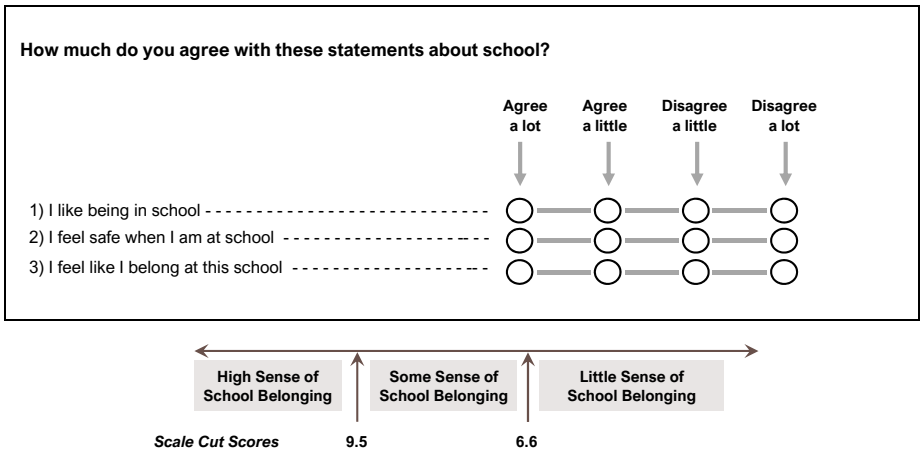
in turn is located above the CI for students with "little sense of school belonging" (339–366). These differences suggest a positive progression in average mathematics achievement with higher levels of school belonging (Exhibit 2.2.19). In reading, there was a similarly consistent, positive relationship between the CIs for average achievement and school belonging across countries. Students categorized as having "high sense of school belonging" had an associated 95% CI ranging from 338 to 349, students reporting "some sense of school belonging" had a CI ranging from 319 to 333, and students with "little sense of school belonging" had a CI ranging from 287 to 313 (Exhibit 2.2.20).

Mathematics & Reading

Exhibit 2.2.18: Students' Sense of School Belonging Scale Description

About the Scale

Students were scored according to their responses to three statements on the *Students' Sense of School Belonging* scale. Cut scores divide the scale into three categories. Students with a **High Sense of School Belonging** had a score at or above the cut score corresponding to "agreeing a lot" with two of the three statements and "agreeing a little" with the other statement, on average. Students with **Little Sense of School Belonging** had a score at or below the cut score corresponding to "disagreeing a lot" with one statement, "disagreeing a little" with one statement, and "agreeing a lot" with one statement, on average. All other students had **Some Sense of School Belonging**.



In mathematics and reading, the 95% CIs for average achievement across countries for these groups of students classified as having "high sense of school belonging," "some sense of school belonging," or "little sense of school belonging" do not overlap. This provides evidence of likely differences in average achievement between these groups for both mathematics and reading.

Mathematics

Exhibit 2.2.19: Students' Sense of School Belonging

Country	High Sense of School Belonging		Some Sense of School Belonging		Little Sense of School Belonging		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	58 (3.1)	388 - 416	33 (2.2)	380 - 401	9 (1.2)	355 - 395	9.5 (0.15)
Egypt	72 (2.1)	397 - 429	22 (1.5)	376 - 408	6 (1.0)	323 - 401	10.1 (0.10)
Nigeria r	70 (2.6)	336 - 374	25 (2.4)	310 - 361	5 (0.5)	272 - 351	10.0 (0.11)
Pakistan	83 (1.2)	376 - 401	15 (1.1)	350 - 379	2 ~	~	10.7 (0.06)
Palestinian Nat'l Auth.	60 (2.1)	409 - 436	28 (1.7)	403 - 434	12 (1.2)	370 - 419	9.6 (0.10)
Senegal	70 (2.7)	349 - 367	25 (2.3)	334 - 364	6 (0.8)	303 - 340	10.1 (0.11)
LaNA 2023 Average	**69 (1.0)**	**384 - 396**	**25 (0.8)**	**368 - 382**	**7 (0.4)**	**339 - 366**	

Reading

Exhibit 2.2.20: Students' Sense of School Belonging

Country	High Sense of School Belonging		Some Sense of School Belonging		Little Sense of School Belonging		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	58 (3.1)	339 - 368	33 (2.2)	331 - 353	9 (1.2)	300 - 339	9.5 (0.15)
Egypt	72 (2.1)	351 - 377	22 (1.5)	335 - 362	6 (1.0)	264 - 324	10.1 (0.10)
Nigeria r	70 (2.6)	301 - 340	25 (2.4)	274 - 321	5 (0.5)	238 - 330	10.0 (0.11)
Pakistan	83 (1.2)	327 - 353	15 (1.1)	295 - 326	2 ~	~	10.7 (0.06)
Palestinian Nat'l Auth.	60 (2.1)	380 - 402	28 (1.7)	360 - 393	12 (1.2)	330 - 369	9.6 (0.10)
Senegal	70 (2.7)	280 - 300	25 (2.3)	264 - 298	6 (0.8)	234 - 273	10.1 (0.11)
LaNA 2023 Average	**69 (1.0)**	**338 - 349**	**25 (0.8)**	**319 - 333**	**7 (0.4)**	**287 - 313**	

This context questionnaire scale was established based on the combined response distribution of countries that participated in the LaNA 2023 Linking Study. To provide a point of reference for country comparisons, the scale centerpoint of 10 was located at the mean of the combined distribution. The units of the scale were chosen so that 2 scale score points corresponded to the standard deviation of the distribution.
( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.
An "r" indicates data are available for at least 70% but less than 85% of the students. A tilde (~) indicates result not reported because estimation is not reliable.

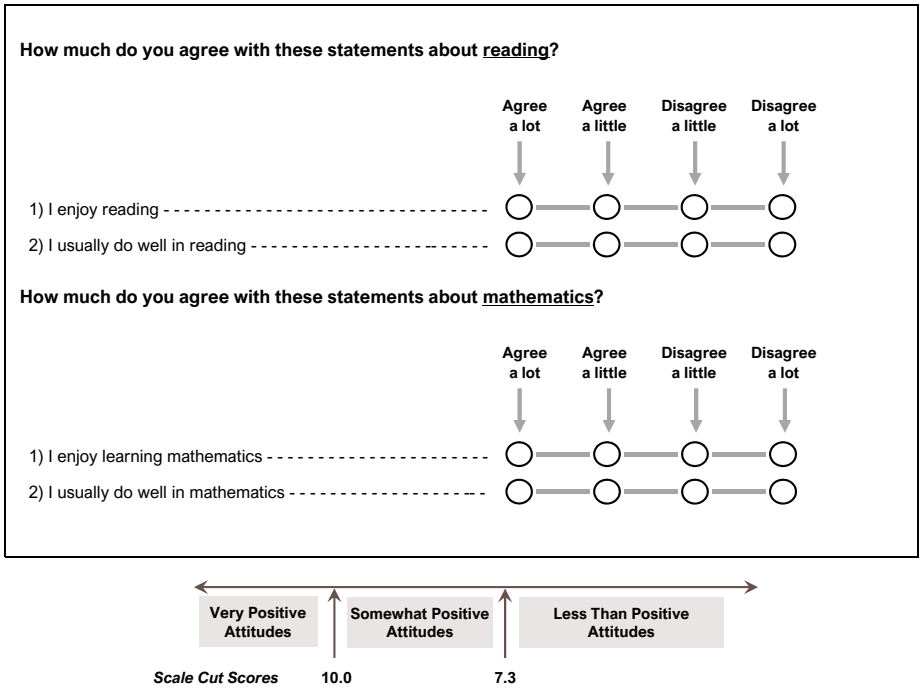
Attitudes Towards Reading and Mathematics

Students who reported positive attitudes toward reading and mathematics had higher average achievement in reading and mathematics than their peers with less positive attitudes. Items in the *Students' Attitudes Towards Reading and Mathematics* scale (Exhibit 2.2.21) asked students to report the extent of their agreement with four statements about enjoying and doing well in mathematics and reading. Based on these responses, students were classified as having "very positive attitudes," "somewhat positive attitudes," or "less than positive attitudes." Within each country, the majority of students had either "very positive attitudes" or "somewhat positive attitudes" towards mathematics and reading, and no more than 13 percent of students had "less than positive attitudes."

Exhibit 2.2.21: Students' Attitudes Toward Reading and Mathematics Scale Description

About the Scale

Students were scored according to their responses to four statements on the *Students' Attitudes Toward Reading and Mathematics* scale. Cut scores divide the scale into three categories. Students with **Very Positive Attitudes** toward reading and mathematics had a score at or above the cut score corresponding to "agreeing a lot" with three of the statements and "agreeing a little" with one statement, on average. Students with **Less Than Positive Attitudes** toward reading and mathematics had a score at or below the cut score corresponding to "disagreeing a little" with one statement and "agreeing a little" with the other three statements, on average. All other students had **Somewhat Positive Attitudes** toward reading and mathematics.



Across the countries, 56 percent of students reported "very positive attitudes," and 34 percent reported "somewhat positive attitudes," while only 10 percent of students reported "less than positive attitudes" on average. In both mathematics and reading, there is a positive relationship between more positive attitudes and average achievement among the countries. In terms of mathematics achievement, students classified as having "very positive attitudes" are characterized by a 95% confidence interval (CI) located above the interval of students having "somewhat positive attitudes" (388–399 vs. 364–377). Both CIs associated with more positive attitudes are located above the 95% CI for students with "less than positive attitudes" (327–345). These differences suggest a positive progression in average mathematics achievement with increasingly positive attitudes toward mathematics and reading (Exhibit 2.2.22). A similarly consistent, positive relationship existed between the CIs for average reading achievement and more positive attitudes towards the two subjects across countries. Students categorized as having "very positive attitudes" had an associated 95% CI ranging from 339 to 351, students

reporting “somewhat positive attitudes” had a CI ranging from 319 to 332, and students with “less than positive attitudes” had a CI ranging from 281 to 299 (Exhibit 2.2.23).

In both mathematics and reading, the 95% CIs for average achievement across countries for the three groups of students (describing “very positive attitudes,” “somewhat positive attitudes,” and “less than positive attitudes”) do not overlap. This provides evidence of likely differences in average mathematics and reading achievement between students in these groups.

Mathematics

Exhibit 2.2.22: Students’ Attitudes Toward Reading and Mathematics

Country	Very Positive Attitudes		Somewhat Positive Attitudes		Less Than Positive Attitudes		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	43 (2.3)	390 - 418	45 (1.8)	383 - 405	12 (1.7)	349 - 377	9.4 (0.11)
Egypt	56 (2.4)	397 - 424	35 (2.0)	378 - 416	9 (1.1)	310 - 363	10.0 (0.10)
Nigeria	49 (3.0)	342 - 382	38 (2.6)	303 - 348	12 (1.4)	255 - 315	9.8 (0.11)
Pakistan	71 (1.5)	378 - 403	24 (1.3)	350 - 380	5 (0.5)	309 - 346	10.8 (0.07)
Palestinian Nat'l Auth.	52 (1.6)	420 - 445	35 (1.2)	384 - 410	13 (1.1)	371 - 424	9.8 (0.07)
Senegal	64 (2.1)	352 - 371	29 (1.6)	334 - 359	7 (1.1)	290 - 323	10.2 (0.09)
LaNA 2023 Average	**56 (0.9)**	**388 - 399**	**34 (0.7)**	**364 - 377**	**10 (0.5)**	**327 - 345**	

Reading

Exhibit 2.2.23: Students' Attitudes Toward Reading and Mathematics

Country	Very Positive Attitudes		Somewhat Positive Attitudes		Less Than Positive Attitudes		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	43 (2.3)	340 - 367	45 (1.8)	335 - 358	12 (1.7)	287 - 327	9.4 (0.11)
Egypt	56 (2.4)	351 - 374	35 (2.0)	334 - 365	9 (1.1)	267 - 315	10.0 (0.10)
Nigeria	49 (3.0)	305 - 350	38 (2.6)	273 - 312	12 (1.4)	226 - 279	9.8 (0.11)
Pakistan	71 (1.5)	326 - 354	24 (1.3)	303 - 330	5 (0.5)	269 - 316	10.8 (0.07)
Palestinian Nat'l Auth.	52 (1.6)	383 - 408	35 (1.2)	351 - 376	13 (1.1)	336 - 376	9.8 (0.07)
Senegal	64 (2.1)	280 - 301	29 (1.6)	270 - 299	7 (1.1)	225 - 260	10.2 (0.09)
LaNA 2023 Average	**56 (0.9)**	**339 - 351**	**34 (0.7)**	**319 - 332**	**10 (0.5)**	**281 - 299**	

This context questionnaire scale was established based on the combined response distribution of countries that participated in the LaNA 2023 Linking Study. To provide a point of reference for country comparisons, the scale centerpoint of 10 was located at the mean of the combined distribution. The units of the scale were chosen so that 2 scale score points corresponded to the standard deviation of the distribution.
( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.

Students’ Perceptions of Instruction

Students’ reports of their perceptions of mathematics and reading instruction have a positive relationship with average achievement in reading and mathematics. The *Students’ Perceptions of Instruction* scale items (Exhibit 2.2.24) asked students to report the extent to which they agree with four statements about their mathematics and science instruction. Based on these responses, students were classified as having “very positive perceptions,” “somewhat positive

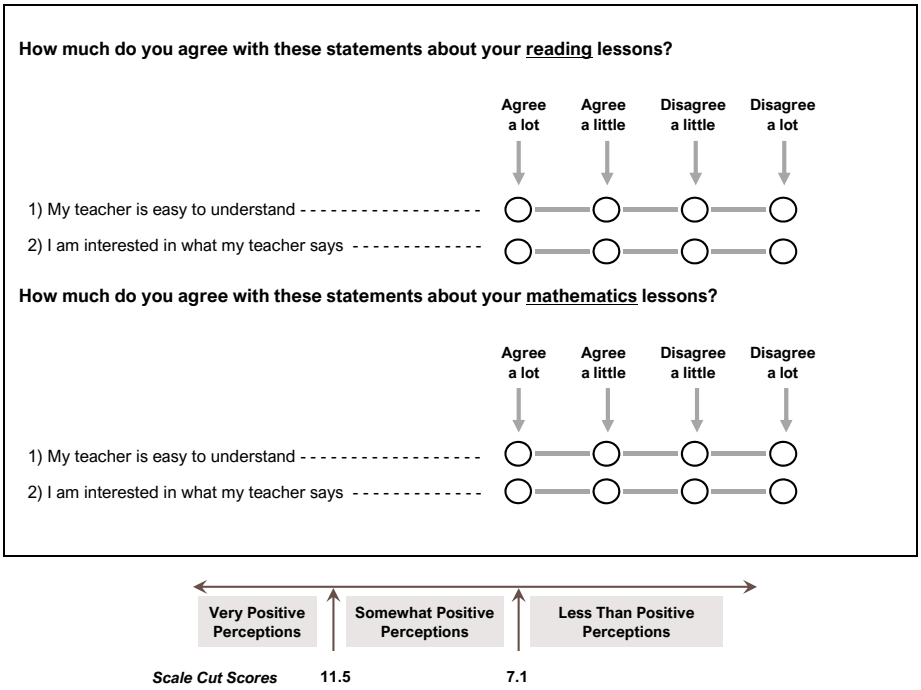
perceptions," or "less than positive perceptions." In all countries, the majority of students participating in LaNA had either "very positive perceptions" or "somewhat positive perceptions" of their mathematics and reading instruction.

Mathematics & Reading

Exhibit 2.2.24: Students' Perceptions of Instruction Scale Description

About the Scale

Students were score according to their responses to four statements on the *Students' Perceptions of Instruction* scale. Cut scores divide the scale into three categories. Students with **Very Positive Perceptions** of their instruction had a cut score corresponding to "agreeing a lot" with all four statements. Students with **Less Than Positive Perceptions** of instruction had a score at or below the cut score corresponding to "agreeing a little" with all four statements. All other students had **Somewhat Positive Perceptions** of instruction.



Across the countries, 56 percent of students reported "very positive perceptions" of their mathematics and reading instruction, 30 percent of students reported "somewhat positive perceptions" and 14 percent of students reported "less than positive perceptions" of their instruction, on average. In mathematics and reading, there is a positive relationship between more positive perceptions of instruction and average achievement across the countries. Students classified as having "very positive perceptions" of instruction have a 95% confidence interval (CI) for average mathematics achievement that does not overlap with the CI for students classified as having "somewhat positive perceptions" (391–403 vs. 369–381). The 95% CI for students with "less than positive perceptions" is located below both of these (328–348). The lack of overlap between these CIs suggest that there are likely differences in average

mathematics achievement between these three groups and a positive progression in average mathematics achievement with increasingly positive perceptions of instruction (Exhibit 2.2.25). A similar pattern is observed between the CIs for average reading achievement and more positive perceptions of instruction across countries. Students categorized as having "very positive perceptions" had a CI for average reading achievement ranging from 343 to 355, students with "somewhat positive perceptions" had a CI ranging from 322 to 334, and students with "less than positive perceptions" had a CI ranging from 281 to 298. As with mathematics, this lack of overlap in CIs suggests likely differences in average reading achievement between these three groups and a positive progression in average reading achievement with increasingly positive perceptions of instruction (Exhibit 2.2.26).

Mathematics

Exhibit 2.2.25: Students' Perceptions of Instruction

Country	Very Positive Perceptions		Somewhat Positive Perceptions		Less Than Positive Perceptions		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	51 (2.6)	395 - 422	34 (2.5)	376 - 407	15 (2.0)	341 - 368	9.9 (0.11)
Egypt	49 (2.7)	403 - 433	33 (1.7)	381 - 412	17 (2.2)	327 - 397	9.8 (0.13)
Nigeria	47 (2.7)	350 - 385	35 (2.4)	315 - 357	18 (1.8)	262 - 320	9.7 (0.11)
Pakistan	66 (1.9)	377 - 404	23 (1.3)	361 - 386	12 (1.1)	340 - 374	10.3 (0.08)
Palestinian Nat'l Auth.	58 (1.4)	421 - 443	30 (1.1)	399 - 427	12 (0.9)	314 - 369	10.1 (0.05)
Senegal	63 (2.4)	354 - 376	24 (1.4)	328 - 352	13 (1.9)	309 - 333	10.2 (0.11)
LaNA 2023 Average	**56 (1.0)**	**391 - 403**	**30 (0.7)**	**369 - 381**	**14 (0.7)**	**328 - 348**	

Reading

Exhibit 2.2.26: Students' Perceptions of Instruction

Country	Very Positive Perceptions		Somewhat Positive Perceptions		Less Than Positive Perceptions		Average Scale Score
	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	Percent of Students	95% Confidence Interval for Average Achievement	
Burkina Faso	51 (2.6)	347 - 374	34 (2.5)	325 - 356	15 (2.0)	289 - 319	9.9 (0.11)
Egypt	49 (2.7)	356 - 381	33 (1.7)	338 - 365	17 (2.2)	285 - 336	9.8 (0.13)
Nigeria	47 (2.7)	311 - 353	35 (2.4)	283 - 320	18 (1.8)	232 - 283	9.7 (0.11)
Pakistan	66 (1.9)	328 - 356	23 (1.3)	309 - 336	12 (1.1)	285 - 327	10.3 (0.08)
Palestinian Nat'l Auth.	58 (1.4)	386 - 406	30 (1.1)	366 - 393	12 (0.9)	277 - 328	10.1 (0.05)
Senegal	63 (2.4)	284 - 309	24 (1.4)	260 - 284	13 (1.9)	242 - 270	10.2 (0.11)
LaNA 2023 Average	**56 (1.0)**	**343 - 355**	**30 (0.7)**	**322 - 334**	**14 (0.7)**	**281 - 298**	

This context questionnaire scale was established based on the combined response distribution of countries that participated in the LaNA 2023 Linking Study. To provide a point of reference for country comparisons, the scale centerpoint of 10 was located at the mean of the combined distribution. The units of the scale were chosen so that 2 scale score points corresponded to the standard deviation of the distribution.
( ) Standard errors appear in parentheses. All percentages are rounded. Because of rounding some results may appear inconsistent.

Highlights: Student Attributes

On average, across the countries participating in the LaNA 2023 Linking Study:

- Students who reported speaking the language of the test at home at least some of the time had higher average achievement in both mathematics and reading than students who reported never speaking the language of the test at home.
- Students who reported more frequent absences from school had lower average achievement in both mathematics and reading than students who reported less frequent absences from school.
- Both students' attitudes towards mathematics and reading and their perceptions of instruction have positive relationships with average mathematics and reading achievement. Students with more positive attitudes and more positive perceptions of instruction have higher average achievement in both subjects.

Appendix A: Instrument Development and the LaNA 2023 Linking Study Design

Charlotte E. A. Aldrich and Lale Khorramdel

Overview

The content of the LaNA Linking Study can be traced back to innovations introduced in the TIMSS 2015 cycle, which launched TIMSS Numeracy as part of its assessment framework (Mullis & Martin, 2013). The special inclusion of foundational mathematics skills into TIMSS development aimed to assess "the fundamental mathematical knowledge, procedures, and problem-solving strategies that are prerequisites for success on TIMSS Mathematics—Fourth Grade" (Mullis, p. 7). PIRLS 2016 also extended the scope of the PIRLS assessment with PIRLS Literacy, which included passages and items targeting the emerging reading comprehension skills required to interact with texts (Mullis & Martin, 2015). Considering how to best extend the TIMSS and PIRLS scales to estimate student achievement at the lower end of the proficiency scales has been a longstanding priority for both assessments. Even still, there are educational contexts in which the TIMSS and PIRLS assessments are too difficult to measure the skills of some student populations accurately. LaNA is designed as a less demanding internationally valid assessment to measure emerging skills in mathematics and reading comprehension at the end of primary school. Currently available in paper-and-pencil format, LaNA's items were developed based on the TIMSS and PIRLS frameworks.

LaNA represents multiple cycles of development, design, administration, and modification to most appropriately match student ability and TIMSS and PIRLS content for more reliable measurement of basic mathematics and reading proficiency in student populations not yet ready to take TIMSS or PIRLS. Three pilot phases, each phase with its own version of LaNA instruments, took place between 2016 and 2020. The results of the pilots informed the final design and content of the LaNA 2023 Linking Study that was administered in 2023. All pilot versions share a basis in the TIMSS and PIRLS Assessment Frameworks, although the characteristics of assessment materials vary across them. Each pilot implementation led to adjustments to item content to improve the coherence between item difficulty and student ability.

The LaNA 2023 Linking Study followed a similar approach to the development of the Context Questionnaires administered as part of the study. Each cycle of TIMSS and PIRLS also includes Context Questionnaires based on updates from the previous cycle's questionnaire frameworks.

This appendix provides an overview of the LaNA project's history and objectives, details the alignment with the TIMSS and PIRLS frameworks, and describes each of the pilot phases, including modifications made to improve the assessment after each phase. It also summarizes the final LaNA 2023 Linking Study design and implementation, including block and booklet composition, as well as key administration dates and activities.

Content Development for LaNA

Objectives and Alignment with TIMSS and PIRLS

Item development for LaNA is based on the TIMSS and PIRLS frameworks for mathematics and reading. The TIMSS and PIRLS assessment frameworks are updated each cycle (every four years for TIMSS; every five years for PIRLS) to support new development, but revisions to the content and cognitive requirements are carefully orchestrated to ensure trend reporting can be maintained.[1] The earlier versions of LaNA correspond to the frameworks for the TIMSS or PIRLS cycle nearest their development. For LaNA, this means that while development began as early as 2015, items developed throughout its history reflect content targeting lower difficulty levels and included in the TIMSS 2019 and PIRLS 2021 frameworks.

In both TIMSS and PIRLS, achievement items are written to meet the requirements outlined in the assessment frameworks. LaNA is a logical extension of past efforts to expand the range of item difficulty in TIMSS and PIRLS, particularly at the lower end of the scales. The TIMSS 2019 Framework established Less Difficult TIMSS as a continuation of what started with TIMSS Numeracy from TIMSS 2015. PIRLS 2021 did not include the PIRLS Literacy categorization for passages but instead classified blocks of passages and items as easy, medium, or difficult. The target material for LaNA was either directly adapted from materials written to be included in the TIMSS 2019 or PIRLS 2021 assessments or was written to represent the foundational skills, knowledge, or processes covered in the content and cognitive domains described within each of these frameworks. The final LaNA materials are adapted from TIMSS 2019 Less Difficult and PIRLS 2021 achievement test content.

Administration of the LaNA 2023 Linking Study also included questionnaires completed by the sampled students (Student Questionnaire) and the principals of the sampled schools (School Questionnaire). The final questionnaires for the LaNA 2023 Linking Study were adapted from items included in either PIRLS 2021 or TIMSS 2019; many of the items in the student and school questionnaires appear in both assessments. Development for those assessments was based on the questionnaire frameworks established for those cycles. The questionnaires collect data on attitudes and environments associated with learning and achievement while keeping the response burden at a minimum. Adapting the TIMSS and PIRLS questionnaire content for LaNA included removing some items from scales to reduce the response burden further and adjust response options for the LaNA countries. Items were also added to the

1 See TIMSS 2019 Assessment Frameworks and PIRLS 2021 Assessment Frameworks for more information.

questionnaires to ask students about household resources (not included in TIMSS and PIRLS questionnaires) and ask principals about the number of unforeseen school closures. Further comparison of the questionnaire contents can be made by referring to the published versions of all questionnaires.

Piloting the LaNA Achievement Instruments

The first iteration of LaNA was conducted as a small pilot study in Haiti in 2016. The design, content, and size of LaNA have developed over the course of iterative pilots since then. Three pilots were conducted across five countries (Haiti, Pakistan, Serbia, North Macedonia, and Niergia) to inform the LaNA 2023 Linking Study content. The main purpose of the pilots was to find the appropriate content and associated difficulty for the LaNA target population. Following the results of each of the pilot studies, the instruments for both mathematics and reading were adjusted to better suit the audience of students for whom TIMSS and PIRLS are too difficult. Changes to the mathematics items and reading passages were made based on performance characteristics such as nonresponse, item difficulty, and item discrimination. Mathematics items were newly developed, and the pilots were essential to determining what content domains and item contexts were appropriate. For the reading items, PIRLS passages were simplified, or different passages were chosen altogether.

Summary of Pilot Versions

The final LaNA instruments are the product of three pilot studies beginning in 2016 in Haiti, referred to as Version 1. Version 2 was implemented in Punjab, Pakistan in 2017, and the Version 3 pilot involved four participating countries between 2019 and 2020. The participants and timeframes of these versions are summarized in Exhibit A.1.

Exhibit A.1: Overview of Pilot LaNA Versions 1–3

Pilot	Participants	Design Overview
Version 1	Haiti—2016 (674 Grade 4 students)	4 Booklets (60 mathematics items; 3 reading passages with 32 items) 1 hour total testing time
Version 2	Punjab, Pakistan—2017 (383 Grade 4 students)	4 booklets (80 mathematics items; 4 reading passages with 43 items) One common block of number items (12 items); one common block of vocabulary items (25 items) 40 minutes for each of two parts

Exhibit A.1: Overview of Pilot LaNA Versions 1–3 (Continued)

Pilot	Participants	Design Overview
Version 3	Serbia—March 2019 (1,131 Grade 3 students)	4 LaNA booklets (80 mathematics items; 4 reading passages (43 items)
	North Macedonia—March 2019 (1,222 Grade 3 students)	One common block of number items (14 items); one common block of vocabulary items (25 items)
	Nigeria—October 2019 (1,308 Grade 4 students)	One Linking Booklet with TIMSS mathematics material (27 items) and PIRLS reading material (one literary passage with 13 items)
	Haiti—March 2020 (1,190 students; 652 at Grade 4, 535 at Grade 6)	LaNA Booklets were administered on Day 1; Linking Booklet was administered on Day 2

Version 1

Version 1 was organized and administered in Haiti. Students were assessed in Grade 4 with a total sample of 674 students. Three reading and three mathematics blocks were developed for the achievement instruments; each mathematics block comprised 20 items developed specifically for LaNA, and each reading passage came from PIRLS but was shortened to 10 items. Each booklet contained one mathematics block and two reading passages to mimic the format of the regional assessments in Haiti. All items were developed to be multiple-choice to simplify scoring and reduce the cognitive load of the assessment. Students had one hour to complete their assigned booklets. Results of the pilot indicated that more than half of the mathematics items were too difficult for students; revisions had to be made to the reading passages.

Exhibit A.2: Blocks for LaNA Pilot Version 1

Mathematics		
Block	**Contents**	**Items**
N1	Mathematics Items	20
N2	Mathematics Items	20
N3	Mathematics Items	20
	Total Mathematics Items	60
Reading		
Block	**Contents**	**Items**
L1	Literary Passage 1	10
L2	Literary Passage 2	11
L3	Informational Passage	11
	Total Reading Items	32

Version 2

Following the results of the initial pilot, substantial improvements were made to the assessment design and item blocks. A new feature introduced at this stage was common blocks of number and vocabulary items shared across booklets. Comprising 12 number and 25 vocabulary items, these LaNA-specific blocks were included in each of the four booklets in the rotational design of Version 2. Exhibit A.3 illustrates the item blocks used in Version 2 in the second pilot study and the composition of the four booklets included in this administration. All mathematics items were developed specifically for LaNA, including some items from the initial Version 1 pilot. At the same time, the reading passages were adapted from existing PIRLS 2016 texts but shortened and simplified for LaNA. Three of the four passages included in this version were also included in the Version 1 pilot; the fourth passage was a PIRLS passage adapted for administration in LaNA. The second pilot study based on these revised instruments was administered to 383 Grade 4 students in Punjab, Pakistan in 2017. Results indicated that mathematics items continued to be too difficult for students while the reading items were largely appropriate.

Exhibit A.3: Blocks and Booklets for LaNA Pilot Version 2

LaNA Version 2—Blocks

Mathematics		
Block	**Contents**	**Items**
N0	Common Number Items	12
N1	Mathematics Items	20
N2	Mathematics Items	20
N3	Mathematics Items	20
N4	Mathematics Items	20
	Total Mathematics Items	92
Reading		
Block	**Contents**	**Items**
L0	Common Vocabulary Items	25
L1	Literary Passage 1	10
L2	Literary Passage 2	11
L3	Informational Passage 1	11
L4	Informational Passage 2	11
	Total Reading Items	68

Exhibit A.3: Blocks and Booklets for LaNA Pilot Version 2 (Continued)

LaNA Version 2—Booklets

Booklet	Part 1		Part 2	
	Common	Unique	Common	Unique
1	L0	L1	N0	N1
2	N0	N2	L0	L2
3	L0	L3	N0	N3
4	N0	N4	L0	L4

Version 3

The final pilot before the LaNA 2023 Linking Study was the most extensive and formalized; four countries participated, with at least 1,000 students sampled within each country. Countries tested students in different grades based on consultation with the TIMSS & PIRLS International Study Center and IEA to determine the age group best-suited for the material. The assessment design for this pilot included unique LaNA booklets as well as a linking booklet containing assessment blocks from TIMSS 2015, TIMSS Numeracy 2015, and PIRLS 2016. The linking booklet was administered to compare the difficulty of the unique LaNA items to easier TIMSS and PIRLS items. The linking item blocks included both multiple-choice and constructed-response items. Thus, national teams participated in scoring training conducted by the TIMSS & PIRLS International Study Center. The pilot administration took place over two days, with students taking one LaNA booklet on the first day and one linking booklet on the second day. The block and booklet design are summarized in Exhibits A.4 and A.5.

Exhibit A.4: LaNA Pilot Version 3—Item Blocks

Mathematics		
Block	**Contents**	**Items**
N0	Common Number Items	14
N1	Mathematics Items	20
N2	Mathematics Items	20
N3	Mathematics Items	20
N4	Mathematics Items	20
NL1	LaNA Linking Block—TIMSS Numeracy 2015—N01	13
NL2	LaNA Linking Block—TIMSS 2015/TIMSS Numeracy 2015—M01/N02	14
	Total Mathematics Items	121

Exhibit A.4: LaNA Pilot Version 3—Item Blocks (Continued)

Reading		
L0	Common Vocabulary Items	25
L1	Literary Passage 1	10
L2	Literary Passage 2	11
L3	Informational Passage 1	11
L4	Informational Passage 2	11
LL1	Linking Block—PIRLS 2016/ PIRLS Literacy 2016 Literary Passage	13
	Total Reading Items	81

Exhibit A.5: LaNA Pilot Version 3—Booklets

Booklet	Part 1		Part 2	
	Common	Unique	Common	Unique
1	L0	L1	N0	N1
2	N0	N2	L0	L2
3	L0	L3	N0	N3
4	N0	N4	L0	L4
Linking	NL1	NL2	LL1	

Results of the third pilot study showed that the items from the common N0 and L0 blocks were too easy for the targeted LaNA population and, therefore, did not provide sufficient information about students' reading and mathematics proficiencies. Moreover, results showed that in line with expectations, the unique LaNA items were less difficult than the TIMSS and PIRLS linking items. This finding indicated that the development of easier assessment items for LaNA was successful.

LaNA 2023 Linking Study Design

Based on the results of the third pilot study (Version 3), item blocks N0 and L0 were removed and substituted with an additional (common) reading passage for the LaNA 2023 Linking Study. In the next step, these final LaNA instruments need to be linked to the TIMSS and PIRLS scales to establish the two new basic LaNA benchmarks (one for reading and one for mathematics), which will extend the TIMSS and PIRLS scales. To establish a stable link, the LaNA 2023 Linking Study was conducted based on a new linking booklet design developed to strengthen the link to the TIMSS and PIRLS assessments. The number of linking booklets was increased from one to four booklets including four times as many TIMSS and PIRLS linking items compared to the third pilot study.

Altogether, the LaNA 2023 Linking Study data collection comprised four LaNA booklets and four linking booklets. LaNA booklets comprise simplified items written based on the TIMSS and PIRLS assessment frameworks, and adapted from earlier cycles. The mathematics portion contains 80 items in total and the reading portion contains 54 items in total. The linking booklets comprise TIMSS 2019 and PIRLS 2021 item blocks. The mathematics portion contains 52 items in total and the reading portion contains a total of 66 items. In both types of booklets, LaNA and linking booklets, each booklet consists of two parts and each part has a duration of 40 minutes, amounting to 80 min in total for each booklet with a short break between the two parts.

The LaNA 2023 Linking Study block and booklet design are summarized in Exhibits A.6 and A.7, respectively. In addition to four unique LaNA booklets (1–4), there are four linking booklets (booklets 5–8). Hence, the LaNA 2023 Linking Study consists of eight booklets in total.

Each of the unique LaNA booklets contains two regular LaNA mathematics blocks, one common reading passage, and one unique LaNA regular passage. All LaNA mathematics blocks appear two times and at different positions. All LaNA reading passages appear one time; the updated common reading passage (L0_Common) appears in all four LaNA booklets and is always presented at the first position in either part 1 or part 2. Each linking booklet contains two linking mathematics blocks and one linking reading passage. All linking mathematics blocks appear two times and at different positions, while all linking reading passages appear one time.

Exhibit A.6: LaNA Version 4 Booklets and Linking Booklets for the LaNA 2023 Linking Study

Blocks	Block Label	Block Source	Items
LaNA Math Blocks	N1	LaNA Version 3—Block N1	20
	N2	LaNA Version 3—Block N2	20
	N3	LaNA Version 3—Block N3	20
	N4	LaNA Version 3—Block N4	20
Math Linking Blocks	NL1	T19 Less Difficult—Block MN01	13
	NL2	T19 Less Difficult—Block MN03	13
	NL3	T19 Less Difficult—Block MN05	13
	NL4	T19 Less Difficult—Block MN07	13
LaNA Reading Passages	L0_Common	Informational Passage	11
	L1	LaNA Version 3—Passage L1	10
	L2	LaNA Version 3—Passage L2	11
	L3	LaNA Version 3—Passage L3	11
	L4	LaNA Version 3—Passage L4	11
Reading Linking Passages	LL1	P21—Literary Easy Passage	18
	LL2	P21—Literary Easy Passage	17
	LL3	P21—Informational Easy Passage	16
	LL4	P21—Informational Easy Passage	15

Exhibit A.7: Booklet Design for the LaNA 2023 Linking Study

LaNA Booklets	Part 1	Part 2
Booklet 1	N1, N2	L0_Common, L1
Booklet 2	L0_Common, L2	N2, N3
Booklet 3	N3, N4	L0_Common, L3
Booklet 4	L0_Common, L4	N4, N1
Key: L—Reading; N—Numeracy		
Linking Booklets	**Part 1**	**Part 2**
Booklet 5	NL1, NL2	LL1
Booklet 6	LL2	NL2, NL3
Booklet 7	NL3, NL4	LL3
Booklet 8	LL4	NL4, NL1
Key: LL—Reading Linking; NL—Mathematics Linking		

The LaNA 2023 Linking Study consisted of a smaller practice administration before main data collection and was administered in six countries: Burkina Faso, Egypt, Nigeria, Pakistan, the Palestinian National Authority, and Senegal. The practice administration had the goal to prepare countries for the main data collection. It utilized two booklets (one LaNA and one linking booklet) administered to about 300 students in each country. The main data collection was based on the full booklet design (including all eight booklets) and aimed for a sample of 4,500 students from at least 100 schools in each country.

In both the practice administration and the main data collection, half of the students received a LaNA booklet, and half received a linking booklet; these booklet types were randomly assigned to each sampled class. Hence, the LaNA 2023 Linking Study is based on an equivalent sample design. All countries administered the practice administration and the main data collection in the same year (2023) and at different grades. Egypt and the Palestinian National Authority administered the practice administration at the end of Grade 4 and the main data collection at the beginning of Grade 5. Pakistan completed both the practice administration and main data collection in Grade 5. All other countries collected data for the practice administration at the end of Grade 5, and the main data collection at the beginning of Grade 6. In all countries, different students were sampled for the practice administration and the main data collection.

Activities and Sample of the LaNA 2023 Linking Study Administration

Development Schedule

In April 2021, the TIMSS & PIRLS International Study Center prepared a revised LaNA assessment design and revised LaNA instruments and began planning the administration of the LaNA 2023 Linking Study. Following further preparations and country recruitment (conducted by IEA), project activities began in 2023 and are summarized in Exhibit A.8.

Exhibit A.8: Overview of Key Project Milestones

2023	
April	TIMSS & PIRLS International Study Center and IEA share practice administration booklets and questionnaire materials with countries Selection of samples for main data collection
May	1st National Research Coordinator (NRC) meeting Practice administration preparations Scoring training for human-scored items (practice administration) Data entry workshop
May–July	Practice administration is conducted
July–August	Translation and layout verification of LaNA booklets and questionnaires
July	2nd NRC meeting Review of practice administration results
September	3rd NRC meeting Scoring training for human-scored items (main data collection)
November–December	Main data collection
2024	
February	All LaNA 2023 Linking Study data received from countries; start of data processing and data cleaning
April	4th NRC meeting (data collection results, overview of planned psychometric analysis, and reporting overview)
April–May	Sampling adjudication and weighting
June–August	Psychometric analysis
September	Scale Anchoring activities and development of descriptions for the new basic reading and mathematics benchmarks
November	Database and IDB Analyzer training for countries

Exhibit A.8: Overview of Key Project Milestones (Continued)

2025	
January	5th NRC meeting (review of draft International Report and benchmark descriptions)
February	LaNA 2023 Linking Study Release Event Release of the International Report
April	Release of International Database

Practice Administration

A small practice administration was orchestrated to support the implementation of the full-scale LaNA 2023 Linking Study (i.e., the main data collection). The sample for the Practice Administration was a convenience sample of about 10 schools chosen by each country. The sample was designed to at least cover different demographic and socioeconomic groups in the target population, such as different geographic regions, urban and rural areas, and public and private schools. Further information is detailed in Appendix B of this report.

One LaNA booklet and one linking booklet were administered to about 300 students per country to provide the national center staff experience in the procedures that would be required to successfully administer the full LaNA assessment during the main data collection. This included translation and layout verification, scoring training for human-scored items at the first National Research Coordinator (NRC) meeting, and training for the administration of instruments. After the administration, some basic analyses were conducted and reviewed with countries at the second NRC meeting. Unlike a formal field test, results from the practice administration did not indicate any changes to the materials included in the main data collection. The two booklets were also included in the data collection administration conducted at the end of 2023.

Main Data Collection

The LaNA instruments were not revised between the practice administration and the main data collection. The administration took place in the six participating countries from October to November 2023. All eight booklets were administered to a full sample of students in each country, as described further in Appendix B. The sample for the main data collection was selected after a series of consultations with the NRCs and was finalized by the sampling experts using a two-stage cluster sample design. The target sample was 4,500 students from at least 100 schools per country. Each student saw items in both mathematics and reading. All training provided in the practice administration was provided again to prepare countries for the main data collection. For the scoring training, additional training materials were prepared for items not included in the practice administration scoring training based on the linking items included in the main data collection. Exhibit A.9 summarizes the number of schools sampled per country, the resulting student sample, and the target grade for the LaNA 2023 Linking Study; detailed sampling information is included in Appendix B.

Exhibit A.9: Overview of LaNA 2023 Linking Study

Country	Number of Schools	Number of Students	Grade Tested
Burkina Faso	100	4,916	6
Egypt	104	5,424	5
Nigeria	100	4,232	6
Pakistan	384	8,744	5
Palestinian National Authority	102	2,744	5
Senegal	122	4,636	6

Scoring Constructed-Response Items

Ensuring reliable and valid scoring of constructed-response items is crucial for accurate assessment results. The TIMSS & PIRLS International Study Center provided explicit scoring guidelines for each individual item, along with comprehensive training on their use. Additionally, the Survey Operations Procedures units outlined an efficient procedure for organizing and implementing the scoring process.

International Scoring Training

Two international scoring training sessions were held: one for the practice administration and one for the main data collection. During these sessions, NRCs or their appointed representatives were trained to score each of the constructed-response items based on the scoring guides developed by the TIMSS & PIRLS International Study Center.

To support human scoring of student responses in the practice administration and for the main data collection, the TIMSS & PIRLS International Study Center provided training to ensure that the scoring guides were applied consistently across countries. Training materials consisted of student responses from previous cycles of TIMSS and PIRLS. TIMSS & PIRLS International Study Center experts and consultants followed a consistent procedure to introduce the item and scoring guide criteria for each item subjected to training. Training items were chosen based on the complexity of the scoring guide or as examples of principles of scoring that would need to be readily applied. For each item, 8–12 scored student responses were prepared as exemplars (example responses), followed by 8–12 unscored student responses for trainees to review and assign scores independently during the training (practice responses). The example responses covered a range of response types and demonstrated how to apply the scoring guides to actual student responses. Afterward, participants used the scoring guides and example responses to score the set of practice responses. Each response and its score were reviewed as a group, and any discrepancies in scores were discussed and corrected.

Following international scoring training, national centers trained their scoring staff in applying the scoring guides for constructed-response items that were administered as part

of the LaNA 2023 Linking Study. NRCs were also encouraged to create additional example papers and practice materials using student responses collected in their respective countries for further training.

Documenting Scoring Reliability

Ensuring reliable scoring of constructed-response items is essential for obtaining high-quality data. To achieve this, it is important to document the reliability of the scoring process. A high degree of agreement among scorers indicates that they have consistently applied the scoring guidelines. For this purpose, in each country, two independent scorers evaluate a random sample of the student responses to constructed-response items in about half of the linking booklets. The level of agreement between their assigned scores serves as a measure of the scoring process's reliability. Importantly, scorers work independently and are unaware of each other's scores. Further information is provided in detail in Appendix C: Survey Operations and International Database.

Citation

Aldrich, C. E. A. & Khorramdel, L. (2025). Appendix A: Instrument development and the LaNA 2023 linking study design. In M. von Davier, L. Khorramdel, K. Reynolds, C. E. A. Aldrich, A. Bookbinder, A. Kennedy, & E. Gonzalez, *LaNA 2023 Linking Study Results* (pp. A.1-A.13). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

References

Mullis, I.V.S. (2013). Introduction. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMSS 2015 Assessment Frameworks* (pp. 3–9). Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2015/frameworks.html

Mullis, I.V.S. & Martin, M.O. (Eds.) (2013). *TIMSS 2015 Assessment Frameworks. Boston College*, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2015/frameworks.html

Mullis, I. V. S., & Martin, M. O. (Eds.). (2015). *PIRLS 2016 Assessment Framework* (2nd ed.). Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/pirls2016/framework.html

Appendix B: Sample Implementation

Duygu Savaşcı-Smith, Rodrigo Leyton, and Sabine Tieck

Overview

Ensuring a representative sample of schools and students was a key component of the LaNA 2023 assessment, with clear roles outlined for each organization involved. The IEA Sampling Unit held primary responsibility for overseeing all sampling-related aspects of the study, including the development and approval of national sampling plans to ensure they met the requirements of the assessment. These plans were collaboratively developed by the National Research Coordinators (NRCs) in each participating country, who were tasked with implementing the approved sampling strategy within their own contexts. NRCs were further responsible for supplying detailed documentation on their national sampling plans, school sampling frames, and any exclusions or adjustments specific to their country. Once this documentation was submitted, the IEA Sampling Unit selected the school samples for each country and trained NRCs on using the Within-School Sampling and Student Tracking (W3ST) Macro to implement within-school sampling effectively. For any queries regarding sampling, NRCs were directed to consult the sampling experts at the IEA Sampling Unit, ensuring the integrity and consistency of sampling practices across all participating countries.

The international sample design framework for the main data collection of LaNA was a stratified two-stage cluster sample design. The sample for the Practice Administration was a convenience sample chosen by each country. Although this was not a probability sample, the sample was designed to at least cover different demographic and socioeconomic groups in the target population, such as different geographic regions, urban and rural areas, and public and private schools. Each sample was reviewed and verified by the IEA Hamburg Sampling Unit. For the Main Data Collection, the sampling design was implemented in two stages, similar to the approach used in TIMSS 2019 (LaRoche & Foy, 2020). In the first stage, schools were sorted by strata, and selected with probabilities proportional to their size, using the number of students in target grade as the measure of size. The number of schools allocated to a stratum was proportional to the number of students expected in that stratum. Replacement schools were assigned to originally sampled schools to be used in cases where originally selected schools refused to participate. The second sampling stage consisted of selecting one classroom from each selected school. Classrooms were selected with equal probability within schools using the W3ST Macro. All students in a sampled class were asked to participate in all countries except one. In Burkina Faso, student subsampling was applied to large classes. Specifically, if a sampled class exceeded a predetermined threshold, a subset of students was selected to participate in the assessment, with random selection as the guiding principle.

After the data were collected and processed, sampling consultants computed the sampling weights. The student sampling weight in LaNA is a combination of weighting components on three levels: school, class, and student. At each level, the weighting component consisted of a basic weight that is the inverse of the probability of selection, together with an adjustment factor for nonparticipation. The overall sampling weight for each student is the product of these weighting components and their corresponding adjustments.

The documentation for the sampling and weighting process for each LaNA participant was completed by the sampling consultants at the IEA Sampling Unit, including detailed information on coverage and exclusion levels, stratification variables, sampling, participation rates, and variance estimates. The TIMSS & PIRLS International Study Center and the LaNA 2023 Sampling Referee used this information to evaluate the quality of the implementation of each sample.

This chapter summarizes the major characteristics of the national samples for LaNA 2023. The Supplement "National Characteristics of Participating Countries" below provides more detailed descriptions of the sample design for each country, including details of population coverage and exclusions, stratification variables, and schools' sampling allocations.

Target Population

The international target population for the LaNA 2023 assessment is defined as the grade representing either five or six years of formal schooling, counting from the first year of primary or elementary schooling. The TIMSS & PIRLS International Study Center, in collaboration with the participating countries, determined the most appropriate grade level for the assessment. The Practice Administration was used to confirm that the achievement difficulties are well targeted for the students in the grade level selected within each participating country.

Egypt, Palestine, and Pakistan administered the test at the beginning of fifth grade, while Burkina Faso, Nigeria, and Senegal administered LaNA at the beginning of sixth grade. Exhibit B.1 presents the grades identified as the target grades by each country. It also presents the number of years of formal schooling at the target grade and the average age of the participating students in the target grades at the time of testing.

Mathematics & Reading

Exhibit B.1: National Grade Definition

Country	Country's Name for Grade Tested	Years of Formal Schooling	Average Age at the Time of Testing
Burkina Faso	Cours Moyens 2	6	12.4
Egypt	Grade 5	5	10.7
Nigeria	Primary 6	6	11.8
Pakistan	5th Class	5	11.1
Palestinian Nat'l Auth.	Grade 5	5	10.4
Senegal	Cours Moyens 2	6	12.5

National Coverage and Exclusions of the LaNA 2023 National Samples

Exhibit B.2 summarizes population coverage and exclusions for the LaNA 2023.

Mathematics & Reading

Exhibit B.2: Coverage of Target Population

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Burkina Faso	100%		5.2%	0.0%	5.2%
Egypt	100%		0.4%	0.2%	0.6%
[2] Nigeria	100%		10.2%	0.0%	10.2%
Pakistan	100%		0.0%	0.1%	0.1%
[1] Palestinian Nat'l Auth.	60%	Schools located in West Bank	2.0%	0.1%	2.1%
Senegal	100%		5.5%	0.0%	5.5%

Coverage

National coverage of the international target population was generally comprehensive, with the exception of Palestine, where it was restricted to schools located in the West Bank, resulting in a coverage of only 60.2%. This exception is annotated in the LaNA 2023 Linking Study Results.

School-Level and Student-Level Exclusions

Within the national target population, it was possible to exclude certain types of schools and students. For most countries, school-level exclusions consisted of schools for students with disabilities and very small or remote schools. Occasionally, schools were excluded for other reasons, as documented in the Supplement "National Characteristics of Participating Countries" below. School-level exclusions were implemented before selecting schools.

Student-level or within-sample exclusions were carried out in participating schools and generally consisted of students with disabilities or students who could not be assessed in the language of the test. For all but one participant, the overall percentage of excluded students combining school and within-school exclusions was 5% or less after rounding. Nigeria, had exclusions of 10% of the target population. Results for participants with an exclusion rate of more than 5% were annotated in the LaNA 2023 Linking Study Results. Note that all LaNA 2023 participants had zero, or close to zero, within-sample exclusions.

Target Population Size

For the Practice Administration, a convenience sample of about 10 schools was selected, with the aim of obtaining approximately 150 responses per achievement item. This required a total of around 300 students for testing.

For the Main Data Collection, each participating country was required to sample at least 100 schools and approximately 4,500 students in the target grade to allow for sample loss due to anticipated nonparticipation. The objective was to achieve at least 4,000 tested students per country.

Exhibit B.3 shows the number of schools and students in each participating country after school-level exclusions, the actual sample sizes, average class sizes, and an estimate of the student population size at the target grade.

The target population figures were derived from the sampling frame used to select the LaNA 2023 samples. These figures do not account for the portion of the population excluded within sampled schools, nor were they adjusted for changes in the population between the date when the information in the sampling frame was collected and the date of the LaNA 2023 data collection (a one-year interval). Nevertheless, a comparison of the two estimates of population size can be seen as a check on the sample selection procedure. In nearly all cases, the population size estimated from the sample closely matched the population size from the sampling frame. However, there were exceptions. In Senegal, the population was underestimated by approximately 18%. This difference may reflect that the sampling frame for Senegal was based on students in Grade 5 in 2022, while the assessment was conducted with Grade 6 students in 2023, during which some students left the school system. In Nigeria, the estimated population was 17% higher than the defined target population, likely due to the use of an outdated frame based on the 2018 National Population Assessment.

Mathematics & Reading

Exhibit B.3: LaNA 2023 Linking Study Target Population and Sample Sizes

Country	Target Population		Sample		
	Schools	Students	Total Number of Schools that Participated	Students	Student Population Size Estimated from Sample
Burkina Faso	9,349	413,302	100	4,916	454,969
Egypt	18,741	2,323,661	104	5,424	2,197,349
Nigeria	53,675	3,145,857	98	4,232	3,692,377
Pakistan	186,258	3,569,227	360	8,744	3,885,405
Palestinian Nat'l Auth.	1,420	68,171	96	2,744	68,517
Senegal	7,946	328,597	116	4,636	268,786

Stratification

LaNA 2023 NRCs consulted with the sampling consultants from the IEA Sampling Unit to identify stratification variables to be used during sample selection. Exhibit B.4 provides the list of explicit and implicit stratification variables used by the participating countries.

Mathematics & Reading

Exhibit B.4: LaNA 2023 Linking Study Stratification Variables

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Burkina Faso	School Type (2)	2	Urbanization (2)
Egypt	School Type (3) Region (3)	5	School type (4)
Nigeria	School Type (2) State size (3)	4	Urbanization (2) State (37)
Pakistan	Region (6) School Type (2)	12	None
Palestinian Nat'l Auth.	School Type (3) Gender (3)	9	None
Senegal	Region (16)	16	School type (2) Urbanization (2)

Meeting LaNA 2023 Standards for Sampling Participation

The goal for participation, set at 100% for selected students and schools, was communicated to LaNA 2023 participants. Guidelines for annotating achievement data for participants securing less than full participation were modeled on IEA's previous TIMSS and PIRLS assessment cycles. As summarized below in Exhibit B.5, countries were assigned to one of two categories based on their sampling participation. Countries in Category 1 were considered to have met all LaNA 2023 sampling requirements and acceptable participation rates. Countries in Category 2 met the participation requirements only after including replacement schools. As in other studies, one of the main goals for quality data in LaNA 2023 was to have as many countries as possible achieve Category 1 status.

Mathematics & Reading

Exhibit B.5: LaNA 2023 Linking Study Categories of Sampling Participation

Category 1	Acceptable sampling participation rate **without** the use of replacement schools. In order to be placed in this category, a country had to have: An **unweighted** school response rate without replacement of at least 85% (after rounding to nearest whole percent) AND an **unweighted** student response rate (after rounding) of at least 85% OR A **weighted** school response rate without replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR The product of the (unrounded) **weighted** school response rate without replacement and the (unrounded) **weighted** student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category would appear in the tables and figures in international reports without annotation, and will be ordered by achievement as appropriate.
Category 2	Acceptable sampling participation rate **only when replacement schools are included**. A country would be placed in this category 2 if: It failed to meet the requirements for Category 1 but had a weighted school response rate **without** replacement of at least 50% (after rounding to the nearest percent) AND HAD EITHER A **weighted** school response rate **with** replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR The product of the (unrounded) **weighted** school response rate with replacement and the (unrounded) **weighted** student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category would be annotated with a "†" in the tables and figures in international reports, and ordered by achievement as appropriate.

Participation Rates of the LaNA 2023 National Samples

Exhibits B.6 and B.7 present the school, classroom, student, and overall weighted and unweighted participation rates for each of the participants in the LaNA 2023 assessment. Almost all participants had excellent participation rates and belonged in Category 1. Pakistan achieved the minimum acceptable participation rate only after including replacement schools; therefore, their results are annotated with a dagger (†) and placed in Category 2.

Mathematics & Reading

Exhibit B.6: Participation Rates (Weighted)

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Burkina Faso	89%	100%	100%	99%	88%	99%
Egypt	100%	100%	100%	96%	96%	96%
Nigeria	97%	98%	100%	99%	96%	98%
† Pakistan	77%	95%	100%	92%	71%	87%
Palestinian Nat'l Auth.	94%	95%	100%	94%	89%	89%
Senegal	100%	100%	100%	96%	96%	96%

LaNA guidelines for sampling participation: The minimum acceptable participation rates were 85 percent of schools, 95 percent of classes, and 85 percent of students, or a combined rate (the product of school, class, and student participation) of 75 percent. Participants not meeting these guidelines were annotated as follows:
† Met guidelines for sample participation rates only after replacement schools were included.

Mathematics & Reading

Exhibit B.7: Participation Rates (Unweighted)

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Burkina Faso	94%	100%	100%	99%	93%	99%
Egypt	100%	100%	100%	96%	96%	96%
Nigeria	95%	98%	99%	99%	93%	96%
† Pakistan	71%	95%	100%	89%	63%	85%
Palestinian Nat'l Auth.	93%	94%	100%	94%	88%	89%
Senegal	100%	100%	100%	96%	96%	96%

LaNA guidelines for sampling participation: The minimum acceptable participation rates were 85 percent of schools, 95 percent of classes, and 85 percent of students, or a combined rate (the product of school, class, and student participation) of 75 percent. Participants not meeting these guidelines were annotated as follows:
† Met guidelines for sample participation rates only after replacement schools were included.

Achieved Sample Sizes of the LaNA 2023 National Samples

Exhibits B.8 through B.9 show the achieved sample sizes in terms of schools and students for each of the participants in the LaNA 2023 assessment.

Mathematics & Reading

Exhibit B.8: School Sample Sizes

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that participated
Burkina Faso	100	100	94	6	100
Egypt	104	104	104	0	104
Nigeria	100	100	95	3	98
Pakistan	384	380	269	92	361
Palestinian Nat'l Auth.	102	102	95	1	96
Senegal	122	117	117	0	117

Mathematics & Reading

Exhibit B.9: Student Sample Sizes

Country	Within-School Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Students Assessed
Burkina Faso	1	4,992	9	0	4,983	67	4,916
Egypt	1	5,743	79	13	5,651	227	5,424
Nigeria	1	4,272	5	0	4,267	35	4,232
Pakistan	1	9,926	130	10	9,786	1,042	8,744
Palestinian Nat'l Auth.	1	2,938	17	7	2,914	170	2,744
Senegal	1	4,988	177	0	4,811	175	4,636

Citation

Savaşcı-Smith, D., Leyton, R., & Tieck, S. (2025). Appendix B: Sample implementation. In M. von Davier, L. Khorramdel, K. Reynolds, C. E. A. Aldrich, A. Bookbinder, A. Kennedy, & E. Gonzalez, *LaNA 2023 Linking Study Results* (pp. B.1-B.9). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

References

LaRoche, S., & Foy, P. (2020). Sample implementation in TIMSS 2019. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report* (pp. 9.1–9.261). TIMSS & PIRLS International Study Center, Boston College. https://timssandpirls.bc.edu/timss2019/methods/chapter-9.html

Supplement B1: National Characteristics of Participating Countries

Burkina Faso

Sixth Grade

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of private Muslim schools
- No within-school exclusions

Sample design

- Explicit stratification by school type (public, private)
- Implicit stratification by urbanization (rural, urban)
- Sampled one class per school, student subsampling of 50 students in large classes (measure of size > 99)

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public	82	0	78	4	0	0	0
Private	18	0	16	2	0	0	0
Total	**100**	**0**	**94**	**6**	**0**	**0**	**0**

Egypt

Fifth Grade

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of very small schools (measure of size < 20)
- Within-school exclusions consisted of students with functional disabilities

Sample design

- Explicit stratification by school type (governmental, governmental language, private) and region (capital, north, south) within governmental stratum
- Implicit stratification by school type (private with fees, private funded without fees, private language, international governmental) within private stratum
- Sampled one class per school

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Governmental School—Capital	18	0	18	0	0	0	0
Governmental School—North	40	0	40	0	0	0	0
Governmental School—South	28	0	28	0	0	0	0
Governmental Language Schools—All	8	0	8	0	0	0	0
Private Schools—All	10	0	10	0	0	0	0
Total	**104**	**0**	**104**	**0**	**0**	**0**	**0**

Nigeria

Sixth Grade

Coverage and Exclusions

- Coverage is 100%
- Overall exclusion of 10.2%
- Estimated population higher than defined population (+17%) due to outdated frame based on 2018 National Population Assessment
- School-level exclusions consisted of very small schools (measure of size < 10), special needs schools, nomadic schools, Islamiyah schools, and Tsangaya/Almajiri schools
- No within-school exclusions

Sample design

- Explicit stratification by school type (public, private) and state size (big, medium, small) within public stratum
- Implicit stratification by urbanization (urban, rural) and state (37 levels) within public schools
- Sampled one class per school, large classes (measure of size > 96) were divided into separate classes and tested

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public—Big States	28	0	27	0	0	1	0
Public—Medium States	40	0	37	0	2	1	0
Public—Small States	22	0	21	1	0	0	0
Private—All	10	0	10	0	0	0	0
Total	**100**	**0**	**95**	**1**	**2**	**2**	**0**

Pakistan

Fifth Grade

Coverage and Exclusions

- Coverage is 100%
- No school-level exclusions
- Within-school exclusions consisted of students with functional disabilities

Sample design

- Explicit stratification by region (AJK—Gilgit-Baltistan, Baluchistan, Islamabad, KPK, Punjab, Sindh) and school type (public, private)
- No implicit stratification
- Sampled one class per school
- Schools were oversampled at the region and school type levels to enable meaningful comparisons

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
AJK—Gilgit-Baltistan—Private	30	0	12	8	8	2	0
AJK—Gilgit-Baltistan—Public	30	0	20	7	3	0	0
Balochistan—Private	30	0	14	8	3	4	1
Balochistan—Public	30	0	18	7	2	3	0
Islamabad—Private	30	2	19	3	6	0	0
Islamabad—Public	30	0	29	1	0	0	0
KPK—Private	30	0	21	5	2	2	0
KPK—Public	30	1	28	1	0	0	0
Punjab—Private	32	0	21	5	4	2	0
Punjab—Public	52	0	48	2	1	1	0
Sindh—Private	30	0	20	4	5	1	0
Sindh—Public	30	0	19	3	4	4	0
Total	**384**	**3**	**269**	**54**	**38**	**19**	**1**

Palestine

Fifth Grade

Coverage and Exclusions

- Coverage is 60.2%, coverage was restricted to schools located in West Bank
- School-level exclusions consisted of very small schools (measure of size < 10), special needs schools and English-speaking schools
- Within-school exclusions consisted of students with functional disabilities

Sample design

- Explicit stratification by school type (public, private, UNRWA), gender (male, female, co-ed)
- No implicit stratification
- Sampled one class per school

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
West Bank—Public—Male	28	0	25	0	0	3	0
West Bank—Public—Female	26	0	26	0	0	0	0
West Bank—Public—Co-ed	10	0	9	0	0	1	0
West Bank—Private—Male	6	0	6	0	0	0	0
West Bank—Private—Female	6	0	3	1	0	2	0
West Bank—Private—Co-ed	8	0	8	0	0	0	0
West Bank—UNRWA—Male	6	0	6	0	0	0	0
West Bank—UNRWA—Female	6	0	6	0	0	0	0
West Bank—UNRWA—Co-ed	6	0	6	0	0	0	0
Total	**102**	**0**	**95**	**1**	**0**	**6**	**0**

Senegal

Sixth Grade

Coverage and Exclusions

- Coverage is 100%
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, language other than French, and area of insecurity
- Estimated population lower than defined population (–18%) due to sampling frame based on Grade 5 students in 2022, while the assessment was conducted with Grade 6 students in 2023

Sample design

- Explicit stratification by region (16 levels)
- Implicit stratification by school type (public, private), urbanization (urban, rural)
- Sampled one class per school, large classes (measure of size >80) were divided into separate classes and tested

School Participation Status

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Dakar	8	0	8	0	0	0	0
Pikine-Guediawaye	12	0	12	0	0	0	0
Rufisque	6	0	6	0	0	0	0
Thies	18	0	18	0	0	0	0
Diourbel	8	1	7	0	0	0	0
Fatick	6	0	6	0	0	0	0
Kaffrine	6	1	5	0	0	0	0
Kaolack	8	0	8	0	0	0	0
Kedougou	6	0	6	0	0	0	0
Tamba	6	0	6	0	0	0	0
Louga	6	0	6	0	0	0	0
Matam	6	1	5	0	0	0	0
Saint-Louis	8	1	7	0	0	0	0
Kolda	6	0	6	0	0	0	0
Sedhiou	6	1	5	0	0	0	0
Ziguinchor	6	0	6	0	0	0	0
Total	**122**	**5**	**117**	**0**	**0**	**0**	**0**

Appendix C: Survey Operations and International Database

Oliver Neuschmidt and Clara Beyer

Survey Operations Procedures

As for TIMSS and PIRLS, the responsibility for administering the LaNA assessment and collecting data lies with individual countries. All participants follow standardized operational procedures to support the collection of consistent, high-quality, internationally comparable data. These procedures, initially developed for TIMSS and PIRLS, were adapted to meet the needs of countries administering the LaNA 2023 Linking Study while upholding the established quality standards of TIMSS and PIRLS.

In each country, the National Research Coordinator (NRC) played a crucial role in implementing the LaNA 2023 Linking Study, representing their country's perspective in international discussions and serving as the contact responsible for project activities. During the assessment cycle, they collaborated on sampling design and conducted a practice administration of the assessment material to test the survey operations procedures in the period of April to June 2023. The main data collection was implemented from mid-November to mid-December 2023.

To record their experiences in the field, and as a quality assurance measure, NRCs recorded their experiences in the field in a Survey Activities Questionnaire (SAQ) for the LaNA 2023 Linking Study. The SAQ gathered this feedback as a quality assurance measure and included an evaluation of the quality of assessment materials and adherence to operational procedures. The results are discussed in the section LaNA Linking Study 2023 Survey Activities Questionnaire.

Survey Operations Units, Manuals, and Software

To support NRCs and national teams during the LaNA 2023 Linking Study administration, IEA Hamburg, in close collaboration with the TIMSS & PIRLS International Study Center, provided detailed operational documentation within Survey Operations Procedures (SOP) units. These units equipped national teams with essential tools and step-by-step procedures for implementing the LaNA Linking Study, with each unit focusing on a specific stage of the assessment. Accompanying manuals were available for School Coordinators and Test Administrators, which could be translated and adapted to local contexts. Additionally, intensive training in constructed-response item scoring and data management was offered by the TIMSS & PIRLS International Study Center and IEA Hamburg, respectively.

To automate and streamline procedures wherever possible, IEA Hamburg provided NRCs with a range of custom-built software products to support a range of activities. These included the Windows® Within-School Sampling and Student Tracking Macro (W3ST Macro) to assist with sampling and tracking classes and students and the IEA Data Management Expert (IEA DME) software for creating and checking data files for all paper-based assessment instruments. These software products were accompanied by manuals to support their use.

The following units, manuals, and software systems were provided for administering the LANA 2023 Linking Study:

- SOP Unit 1: Sampling Schools and Obtaining their Cooperation
- SOP Unit 2: Sampling Classes and Main Data Collection
  - SOP Unit 2 was accompanied by a School Coordinator and a Test Administrator Manual
- SOP Unit 3: Preparing the Main Data Collection Instruments
- SOP Unit 4: Scoring the Main Data Collection Instruments
  - SOP Unit 4 was accompanied by a scoring guide and scoring training material
- SOP Unit 5: Entering the LaNA Main Data Collection Data

The aforementioned manuals pertain to the Main Data Collection phase. However, for the Practice Administration Phase, a distinct set of SOP Units, along with accompanying documents and software, was provided. SOP Unit 1 covered both study phases.

Survey Tracking Forms

The LaNA 2023 Linking Study employed a series of tracking forms to document class sampling procedures, assign assessment instruments to students, and track student information, including the participation status of respondents. These tracking forms also facilitated the data collection and verification processes. Specifically, two different tracking forms were used for the LaNA 2023 Linking Study:

- Class Listing Form: Completed for each sampled school, listing eligible classes and providing details such as class names and the number of students.
- Student Tracking Form: Created for each assessed class, this form was completed by Test Administrators during test administration. It served to verify the assignment of survey instruments to students and indicate their participation status.

Survey Operations Procedures

The NRCs coordinated several major operational activities related to the LaNA 2023 Linking Study, as described in the sections below:

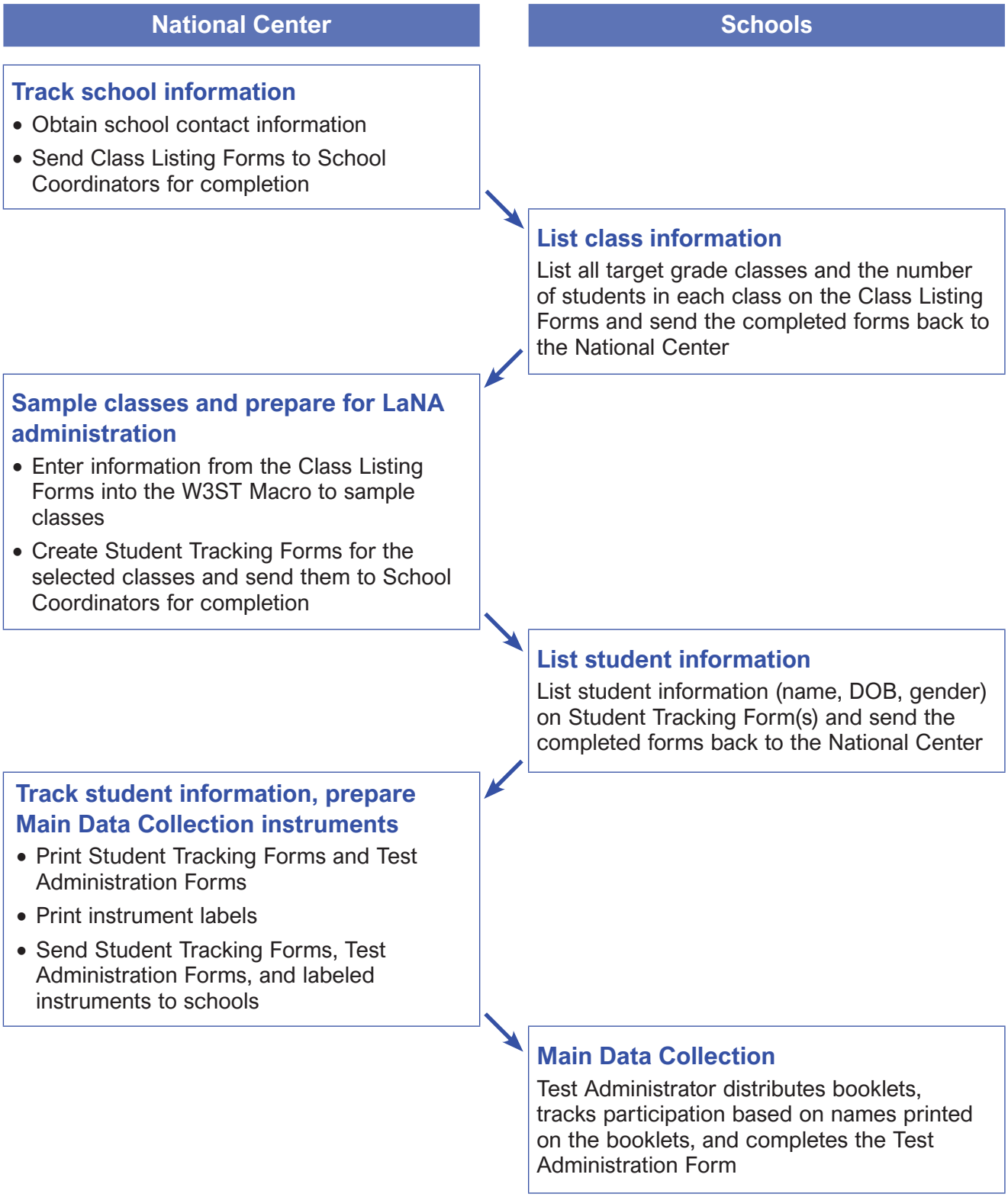
- Contacting schools and sampling classes: Reaching out to schools and selecting classes for the LaNA 2023 Linking Study with the help of the provided Within-School Sampling and Tracking macro
- Overseeing national instrument preparation: Coordinating the translation and preparation of assessment and context questionnaire materials
- Managing study administration: Preparing and conducting the administration of the assessment
- Scoring constructed-response items: Evaluating student responses to the constructed-response items administered as part of the LaNA linking material
- Creating LaNA database files: Entering and verifying responses from the LaNA paper instruments

Additionally, one other significant operational activity related to LaNA is discussed in a separate section of this publication: sampling schools (Appendix B).

Contacting Schools and Sampling Classes

Exhibit C.1 outlines the key steps in collaborating with schools to sample classes and prepare for the LaNA 2023 Linking Study administration. After drawing the school samples, NRCs were tasked with the responsibility of contacting schools and encouraging their participation in the LaNA 2023 Linking Study. Depending on the national context, this process might involve seeking support from national or regional educational authorities. SOP Unit 1 provided guidance on effective strategies to encourage schools' involvement in the assessment.

Exhibit C.1: Diagram of Sampling Procedures and Preparations for the Assessment Administration Implemented by National Centers and Schools



NRCs collaborated with school principals to identify and train School Coordinators for participating schools. School Coordinators, who could be teachers or guidance counselors, received a manual outlining their responsibilities. These included preparing for the testing sessions, coordinating testing logistics, training Test Administrators, distributing questionnaires, and ensuring assessment materials' security and return. Additionally, School Coordinators played a crucial role in providing data for the sampling process.

Overseeing National Assessment Instrument Preparation

NRCs were responsible for preparing assessment instruments (including achievement booklets and context questionnaires) for administration in their countries. This process involved overseeing translation and/or adaptation of the test content to create internationally comparable instruments tailored to each country's context. The LaNA 2023 Linking study included material for the LaNA assessment (four booklets total) as well as material from original TIMSS and PIRLS assessments that were used to establish the LaNA–TIMSS/PIRLS Link (also four booklets total). The LaNA assessment included five LaNA reading passages—two literary and three informational—along with five mathematics blocks. The Link material comprised four PIRLS reading passages—two literary and three informational—and four blocks of TIMSS items.

To facilitate the adaptation and translation phases, participating countries received National Adaptations Forms (NAFs) to record any desired adaptations to the instruments; PDFs of the international (English) versions of the instruments; and corresponding Rich Text Format (RTF) files which were used to enter their translations. Using the international versions of the instruments as a guide, they first entered any desired adaptations into NAFs, which were used by IEA Hamburg to provide feedback and indicate approval or denial of proposed adaptations. After the adaptation verification process, countries translated instruments using the RTF files. The NAFs and RTF files were then submitted to IEA Amsterdam for translation verification. IEA Amsterdam worked with independent translators to evaluate each country's translations and, when deemed necessary, suggested changes to the text.

Once the translations had been verified, the instruments could be assembled into printable PDF documents. Instructions on how to use the provided materials to produce high-quality, standardized instruments were included in the corresponding SOP Unit 3. For some countries, IEA Hamburg created the national PDF versions of the instruments using verified adaptations and translations, while other countries performed the instrument preparation in-country. After the instruments were compiled (and, if done in-country, sent to IEA Hamburg), a layout verification step was performed. This process assured that the instruments conformed to the international standards.

During the Practice Administration (PA), only one of the LaNA booklets and one of the Link booklets were prepared. The full verification process, as described above, was implemented only for the Main Data Collection phase.

Managing the LaNA 2023 Linking Study Administration

NRCs played a crucial role in organizing and planning the distribution of assessment materials. They ensured that the materials were carefully prepared and ready for distribution. These materials were then provided to School Coordinators before the testing phase. This advance distribution provided School Coordinators sufficient time to verify the receipt and correctness of the materials. Assessment instruments were securely stored until the actual testing date.

Each sampled class had a designated Test Administrator who followed specific procedures outlined in the Test Administrator Manual. Their responsibilities included distributing materials to students, reading instructions to students during the assessment sessions, and timing the sessions. The W3ST Macro, the system used for this process, facilitated distribution by assigning achievement booklets to students in a rotated design and producing identification labels.

Each achievement booklet was divided into two parts. In LaNA booklets, one part contained two blocks of LaNA mathematics items and two blocks of LaNA reading passages. In Linking Booklets, one part contained two blocks of TIMSS Mathematics items and one part contained an easy PIRLS Reading passage. The full test design is described in Appendix A. Students were allotted 40 minutes to complete each part (i.e., 80 minutes total), and Test Administrators were responsible for monitoring the timing.

Between the two parts of the achievement test, there was a required break, which was not to exceed 30 minutes. Students who completed Part 1 or Part 2 before the allotted time were not allowed to leave the testing room. Instead, they were asked to review their answers or read quietly.

Following the achievement test, students were given an additional 30 minutes to complete the student questionnaire. Test Administrators were allowed to read the questionnaire items aloud to the students during this time.

Linking Students to their Classes

Exhibit C.2 illustrates the identification system codes that were used to link the data among schools, classes, students, and teachers. The school, class, and student IDs were hierarchical, with classes nested within schools and students nested within classes.

Exhibit C.2: Hierarchical ID System Codes

Participant	ID Components	ID Structure	Numeric Example
School	School	CCCC	1001
Class	School + Class within the school	CCCCKK	100101
Student	School + Class within the school + student within the class	CCCCKKSS	10010101

Scoring the Constructed-Response Items

Ensuring reliable and valid scoring of constructed-response items is crucial for accurate assessment results. The TIMSS & PIRLS International Study Center provided scoring guides for each individual item that required evaluation by one or two human scorers, along with comprehensive training on their use. Additionally, the SOP units outlined an efficient procedure for organizing and implementing the scoring process.

Two international scoring training sessions were held, one for the Practice Administration and one for the main data collection. During these sessions, NRCs (and their appointed representatives) were trained to score each of the constructed-response items. Participants reviewed the scoring guide for each item and applied it to a set of example student responses that had already been scored. These examples covered a range of response types and demonstrated how to apply the guidelines clearly. Afterward, participants applied the scoring guides to a different set of student responses that had not yet been scored. Any discrepancies in scores were discussed within the group. Following international scoring training, national centers trained their scoring staff on applying the guidelines for constructed-response items. NRCs were also encouraged to create additional example papers and practice materials using student responses collected in their respective countries.

Documenting Scoring Reliability

Ensuring reliable scoring of constructed-response items was essential for obtaining high-quality data. To achieve this, it was important to document the reliability of the scoring process through the implementation of within-country reliability scoring. A high degree of agreement among scorers indicated that they had consistently applied the scoring guidelines. For this purpose, in each country, two independent scorers evaluated a random sample of the student responses to constructed-response items in about half of the linking booklets. The level of agreement between their assigned scores served as a measure of the scoring process's reliability. Importantly, scorers worked independently and were unaware of each other's scores. The within-country reliability scoring was integrated throughout the main scoring procedure.

Creating the LaNA 2023 Linking Study Database Files

The process of creating the LaNA 2023 Linking Study database file involved several steps. First, tracking form information as well as responses from paper context questionnaires and achievement booklets were entered in electronic files using IEA's Data Management Expert (IEA DME) software. IEA Hamburg provided the IEA DME software, which allowed for keyboard data entry from paper instruments and provided data and file management capabilities, making it a convenient tool for handling large amounts of data efficiently.

IEA Hamburg provided international codebooks describing all variables and their properties. These codebooks ensured that data files produced with the system adhered to internationally

defined rules and standards for data entry. Before use, the international codebooks were updated to accommodate any national adaptations to the data collection instruments. These adapted national codebooks were then used to create data files in each country, with responses from paper context questionnaires, achievement booklets, tracking forms, and Reliability Scoring Sheets manually entered into the IEA DME database.

Quality control during data entry was crucial to maintaining accurate data. NRCs periodically performed reliability checks during data entry and applied data verification checks provided with the IEA DME software before submitting the databases to IEA Hamburg. Data entry staff independently reentered at least 5% of records from each instrument type to ensure reliability. Acceptable error rates were 1% or less for questionnaire files and 0.1% or less for student achievement files and reliability scoring files. If required agreement was not reached, data entry staff underwent retraining and affected records were re-entered.

IEA DME also provided a data verification module that checked for various issues, such as inconsistent identification codes, discrepancies between participation status information and achievement/questionnaire data availability, invalid codes, and out-of-range values for numerical answers. The module also verified the integrity of the linkage between students and schools entered into the IEA DME database.

Once all data files passed quality control checks, they were submitted to IEA Hamburg along with data documentation for further checking and processing. Details on data processing at IEA Hamburg are provided in the section Data Verification and creation of the International Database below.

Survey Activities Questionnaire

The SAQ was designed to elicit information about NRCs' experiences in preparing for and conducting the LaNA data collection. The questionnaire was composed of four sections and focused on the following:

- Survey operations procedures, sampling schools and classes
- Translating, adapting, and producing the assessment instruments
- Administering the assessments
- Scoring the constructed-response items and creating the databases and documentation

All items in the SAQ included accompanying comment fields, in which NRC respondents were encouraged to explain their responses, provide additional information, and suggest improvements to the process.

The LaNA 2023 Linking Study SAQ was administered via MS Forms and was completed by all six NRCs. The following sections summarize information gathered from the Survey Activities Questionnaire.

Survey Operations in general, Sampling Schools and Classes

The initial section of the SAQ which is summarized in Exhibit C3 assessed the clarity of survey procedures, challenges in school participation, and challenges NRCs faced with implementing the sampling design.

All NRCs found the SOP units to be clear and sufficient. Two countries reported difficulties in convincing schools to participate in LaNA. One NRC faced special challenges when dealing with private schools not overseen by government authorities, while the second NRC occasionally relied on assistance from local and national authorities. One country reported challenges with the administration regarding conflict in the region but mentioned that related issues could be solved before the administration. All countries implemented the prescribed within-school sampling design, meaning that they could collect data about classes and students on time before the test implementation and hence follow the step-by-step approach described in Exhibit C.1. All countries but one regarded the newly developed Within-School Sampling and Tracking Macro (W3ST Macro) easy to work with, while one NRC reported issues with entering school and class data. IEA Hamburg provided support as needed and successfully helped resolve the issues.

Exhibit C.3: SAQ, Section One—Survey Operations Procedures (SOP), Sampling Schools and Classes (Number of NRC Responses)

Question	Yes	No
Was the information about the survey operations procedures provided in the SOP Units 2–5 clear and sufficient?	6	0
Did you have any difficulties in convincing schools to participate?	2	4
Were you prevented from administering LaNA in certain schools due to conflict or instability in your country?	1	5
Were there any conditions or organizational constraints that necessitated deviations from the standard within-school sampling design?	0	6
Did you find the W3ST Macro easy to work with?	5	1

Translating, Adapting, and Producing Assessment Instruments

The second section of the SAQ asked NRCs about translating, adapting, assembling, and printing the test materials, as well as issues related to checking the materials and securely storing them.

In Exhibit C.4, all NRCs indicated their understanding of the procedures for adapting survey instruments and documenting national deviations. However, half of the NRCs found it challenging to adapt the international test instruments to their local context. Additionally, one NRC reported difficulties in adapting background questionnaires. These challenges primarily

related to material that was unfamiliar to students, such as a story about an octopus (which might not be well-known in the region). Furthermore, the thorough translation review process occasionally involved unexpected delays. IEA Hamburg and IEA Amsterdam maintained constant communication with countries to assist them in overcoming technical and content-related challenges throughout the translation, adaptation, and instrument production process.

Exhibit C.4: SAQ, Section Two—Translating, Adapting, and Producing Assessment Instruments

Question	Yes	No
Was the procedure of adapting the instruments and documenting adaptations in the National Adaptations Form (NAF) clear to you?	6	0
Question	**Somewhat**	**Not at all**
How difficult was it to adapt the international source versions of the test booklets?	0	6
How difficult was it to adapt the international source versions of the questionnaires?	5	1

Assessment Administration

The third section of the SAQ asked NRCs about the assessment administration procedures.

Exhibit C.5: SAQ, Section Three—Assessment Administration

Question	Yes	No
Did you have any difficulties with identifying, appointing or contacting School Coordinators?	0	6
Did School Coordinators and/or Test Administrators report any issues with regard to the timing of the test?	2	4
Did School Coordinators and/or Test Administrators report any issues with the test items?	2	4
Did School Coordinators and/or Test Administrators report any issues with the rotation of the test booklets?	1	5
Were there any defective test booklets (e.g., printing errors, missing booklets, incomplete booklets)?	3	3
Are you aware of any circumstances under which Test Administrator did not properly complete the Student Tracking Form and/or the Test Administration Form?	3	3
In your country, were there many instances of mismatches between Students' listed age as listed on the Student Questionnaire vs. the Student Tracking Forms?	4	2

Exhibit C.5: SAQ, Section Three—Assessment Administration (Continued)

Question	Yes	No
In your country, were there many instances of mismatches between Students' listed gender as listed on the Student Questionnaire vs. the Student Tracking Forms?	3	3
To your knowledge, did any Test Administrators deviate from the standardized administration script?	1	5
Did the duration and/or timing of any testing sessions differ from the standardized test duration and timing?	0	6
Question	**Somewhat**	**Not at all**
How difficult was it to recruit Test Administrators?	2	5

Exhibit C.5 indicates that none of the countries encountered difficulties in selecting School Coordinators, but two NRCs found it somewhat challenging to recruit Test Administrators. In certain countries, both School Coordinators and Test Administrators faced issues related to timing (in two countries), test items (in two countries), or the rotation of test booklets (in one country). Regarding timing, it was reported that some students needed more time due to lengthy texts. Additionally, in one country, two items were identified as mistranslated, and an NRC noted deviations from the prescribed booklet rotation. These deviations occurred because test administrators were unfamiliar with prelabeled booklets and inadvertently assigned booklets meant for absent students to subsequent present students. During data cleaning, rotation issues were identified and countries were consulted to ensure accuracy in ID linkages. One of the mistranslated items was corrected during the data cleaning process, while another had to be excluded from the analysis.

Three countries reported some issues with defective booklets and tracking forms, and three countries reported that tracking forms and test administration forms were not always completed properly. Most issues were resolved during test administration, while any information that was unavailable or could not be clearly attributed to a specific student was coded as "not administered" during the cleaning process.

Four countries reported challenges with harmonizing birth information between student tracking forms and background questionnaires, while three countries reported a similar issue with gender information. As a reason for these inconsistencies, countries reported challenges in consolidating the student data on time before the test that was administered in most countries rather at the beginning of the school year. Moreover, the more pronounced discrepancy in students' ages often arises due to students' lack of awareness about their birth dates. During the data cleaning process, birthdate information was cross-validated across multiple sources and, whenever possible, corrected in close collaboration with the NRCs.

While the timing of the testing session was adhered to by all countries, one country reported deviations from the standardized procedures by a different use of spare booklets and issues with following the rules for filling out the participation status.

Scoring the Constructed Response Items and Creating and Submitting the Databases and Documentation

The fourth section of the SAQ asked NRCs about the procedures for scoring, data entry, data verification, and data submission.

Exhibit C.6: SAQ, Section Four—Scoring, and Creating and Submitting the Databases and Documentation

Question	Yes	No
Did you experience difficulties scoring the LaNA Linking Booklet items?	1	1
Did you have any difficulties with unacceptable levels of error during the double-entry of data (e.g., due to the incorrect handling of certain data entry rules by particular staff)?	0	6
Did you have any difficulties performing the quality control checks of the data (e.g., record consistency check)?	0	6

Exhibit C.6 shows that only one country reported challenges with scoring the LaNA Linking items. It was reported here that scorers in some cases struggled with selecting the appropriate two-digit invalid code for some of the mathematics link items. IEA Hamburg and the ISC helped to clarify scoring issues in mutual country contacts. None of the countries reported any issues with performing the different IEA DME within-country data quality checks.

Data Verification and Creation of the International Database

The LaNA 2023 Linking Study International Database is available to researchers, analysts, and any individuals interested in the data collected and analyzed as part of the LaNA 2023 Linking Study. The database includes student achievement data and contextual data collected from students and school principals for the six participating countries.

Preparing the LaNA 2023 International Database and ensuring its integrity was a major undertaking, requiring extensive collaboration between the LaNA Consortium (i.e., IEA Hamburg and the TIMSS & PIRLS International Study Center) and the national centers of participating countries. In creating a large database that contains internationally comparable data collected across a diverse group of countries, languages, cultures, and educational contexts, it is important that standardized procedures are followed and that both the process and content are well-documented for secondary users of the data.

Participating countries followed standardized survey operations procedures to prepare instruments, administer the achievement test, collect context questionnaire data, and prepare data files and documentation to submit to IEA Hamburg for processing and validation. Once the data from the countries was received, standardized international data management and verification procedures to check for errors and inconsistencies were followed, and standardized data and summary reports for each country were created. These procedures addressed unique aspects for each instrument (achievement test and context questionnaires) and accounted for national adaptations to the context questionnaires made by the participating countries.

After countries prepared and submitted their national data and documentation, data processing included a uniform standard cleaning procedure for data structure, identification variables, linkages, and context data. The international data management process included verification of the following aspects:

- All information in the final database conforms to the internationally-defined data structure.
- The content of codebooks and documentation appropriately reflects national adaptations to questionnaires.
- All variables used for international comparisons are comparable across countries (after harmonization, where necessary).

IEA Hamburg checked the data files submitted by each country, applying standardized data cleaning rules to verify the accuracy and consistency of the data and documenting any deviations from the international file structure.

National Research Coordinators (NRCs) from each participating country collaborated with IEA Hamburg to resolve any queries that emerged during the data cleaning process and checked interim versions of their database(s) during this process. The TIMSS & PIRLS International Study Center provided NRCs with univariate data almanacs containing summary item statistics on each variable so that the national centers could evaluate their data from an international perspective.

The TIMSS & PIRLS International Study Center also conducted all operational psychometric analyses of the achievement and context questionnaire data. It produced mathematics and reading achievement scores, context questionnaire scores, and other derived variables based on the context data. The Sampling Unit at IEA Hamburg calculated the sampling weights, population coverage, and school and student participation rates.

The *User Guide for the LaNA 2023 Linking Study International Database* describes all data files and their contents, along with documentation about the achievement items and context questionnaire items.

Preparing and Submitting National Data and Documentation

Participating countries collected and entered the LaNA 2023 Linking Study data and applied first quality control measures, before submitting data to IEA Hamburg for processing and cleaning. After processing and cleaning, data from paper achievement booklets and context questionnaires went to the TIMSS & PIRLS International Study Center for verification and analysis.

Data Entry and Verification of Instruments

Each national center was responsible for entering the responses collected into data files using the IEA Data Management Expert (IEA DME) software. IEA DME is a software system developed by IEA Hamburg that facilitates data entry and includes validation checks to identify inconsistencies. National centers were instructed to enter data for any questionnaire that contained at least one valid response and to discard unused or empty instruments.

National centers entered responses from the instruments into data files using a predefined national codebook. The codebook defines the structure of the data to be entered and contains information about the names, lengths, labels, and missing codes of variables; valid response ranges for continuous measures or counts; and valid values for nominal or ordinal questions. National codebooks were created to match the structure of the national instruments and account for any adaptations.

To ensure consistency across participating countries, the basic rule for data entry into DME required national staff to enter data "as is" without any interpretation, correction, truncation, imputation, or cleaning.

The guiding principles for data entry included the following:

- Responses to closed-response items were coded as "1" if the first option was marked, "2" if the second option was marked, and so on.
- Responses to open-response questions, for example number of students in the sampled class, were entered "as is" even if the value was outside the originally expected range.
- Nonresponse, ambiguous responses, responses given outside of the expected format, or conflicting responses (e.g., selection of two options in a multiple-choice question) were entered as "omitted."

As each respondent ID number was entered, the DME software checked it for alignment with a five-digit checksum generated by the W3ST Macro. A mistype in either the ID or the checksum resulted in an error message prompting the person entering the data to check the entry. The data verification module of the DME also checked for other issues, such as inconsistencies in identification codes and out-of-range or otherwise invalid codes. When such issues were flagged by the software, the data entry staff were prompted to resolve the inconsistency before resuming data entry.

Double Data Entry

To check data entry reliability in participating countries, national centers were required to have 5% of each survey instrument (achievement booklet or questionnaire) entered a second time by a different data entry person, operating independently from the first. It was recommended that countries begin the double-data-entry process as early as possible during the data capture period to identify possible systematic misunderstandings or mishandlings of data entry rules, and to initiate appropriate remedial actions such as retraining or replacing data entry staff.

Although it was desirable that every discrepancy be resolved before submission of the complete dataset, the acceptable level of disagreement between the originally entered and double entered data was established at 1% or less for questionnaire data and at 0.1% or less for achievement data. Values above these levels required resolution of the discrepancy and re-entry of data.

The level of disagreement between the originally entered and double-entered data was evaluated, and it was found that in general, the margin of error observed for processed data was well below the required thresholds.

Data Verification at the National Centers

Before sending the data to IEA Hamburg for further processing, national centers carried out mandatory validation and verification steps on all entered data and implemented corrections as necessary.

While the questionnaire data were being entered, the data manager or other staff at each national center used the information from the tracking forms to verify the completeness of the materials. Student participation information (e.g., whether a student participated in the assessment or was absent) was entered into a special dataset in the IEA DME.

The validation process was supported by quality control checks in the IEA DME, including the generation of an inconsistency report listing discrepancies between variables recorded from the tracking forms during the test administration processes and the achievement and background data.

Data managers were requested to check and, whenever possible, resolve issues before data submission. When inconsistencies remained, and the national center could not solve them, IEA Hamburg asked the center to provide documentation on these problems.

Extensive documentation regarding the test administration process was requested from all participating countries. Test administrators were required to complete Test Administration Forms for each test session. These forms could be filled out directly online using MS Forms or as Excel documents, which were later entered into MS Forms by national center staff. Additionally, a report on procedural activities at the national level was collected through an online Survey Activities Questionnaire.

Data Processing and Data Cleaning

To ensure the integrity of the international database, IEA Hamburg followed uniform data cleaning procedures as part of processing national data, involving regular consultation with NRCs. The main objectives of the data cleaning process were to ensure that the data adhered to international formats, that school and student information could be linked across different survey files, and that the data reflected the information collected within each country in an accurate and consistent manner.

After each country submitted its data, codebooks, and documentation, IEA Hamburg, in collaboration with the NRCs, conducted a four-step cleaning procedure upon the submitted data and documentation:

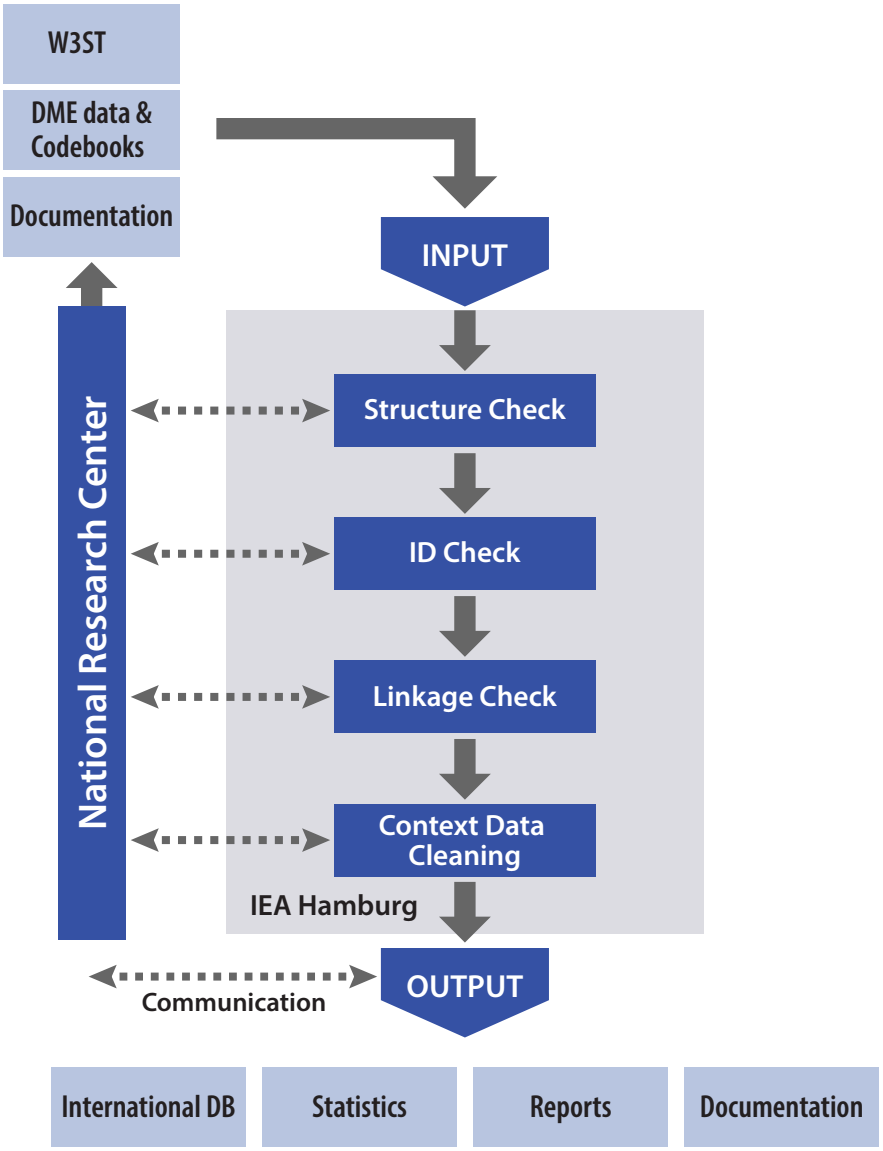
- Checking structure and documentation
- Checking identification (ID) variables
- Checking linkages
- Context data cleaning

Data processing included numerous iterations of the four-step cleaning procedure and was completed for each national data set in close collaboration with national centers. This iterative multi-step cleaning ensured that all data were properly cleaned and that any new errors that could have been introduced during the data cleaning were rectified. Any inconsistencies detected during the cleaning process were resolved in collaboration with national centers, and all corrections made during the cleaning process were documented in the cleaning report produced for each country. The cleaning process was repeated as many times as necessary until all data were made consistent and comparable.

After the final cleaning iteration, each country's data was sent to the IEA Sampling team for the calculation of sampling weights. Then the data, including sampling weights, were sent to the TIMSS & PIRLS International Study Center so that the psychometric analyses could be conducted. The NRCs were provided with interim data products to review at different points in the process.

As illustrated in Exhibit C.7, the program-based data cleaning consisted of a set of activities explained in the following subsections.

Exhibit C.7: Data Processing Workflow



Checking Structure and Documentation

For each country, data cleaning began with a review of data file structures and data documentation, including a review of National Adaptation Forms and Student Tracking Forms. After the review, and as a first cleaning step, IEA Hamburg re-ran the DME inconsistency checks that also were provided to the countries for within-country data verification purposes. In case of any remaining unclarified inconsistencies, cleaning reports and accompanying documentation were created, and country teams were requested to resolve or explain the remaining issues.

In a subsequent step, all DME data was imported into the SAS Cleaning system.

The first checks identified differences between the international and the national file structures. To keep track of adaptations and the approval process, national centers were required to complete National Adaptation Forms. For the LaNA 2023 Linking Study, countries' national instruments adhered to the international data structure, and adaptations were limited to cultural adaptations to the test and questionnaire items. Two countries required the addition of a variable to track the language of the test and questionnaire administration. The different data sources for these two countries were harmonized into the corresponding international variable.

Variables created purely for verification purposes during data entry were discarded, and provisions were made for adding new variables necessary for analysis and reporting, including reporting variables, derived variables, sampling weights, and scale scores.

In a subsequent step, a series of standard data cleaning rules were applied for further processing and implemented in the SAS data processing system.

Each potential problem flagged at this stage was identified by a unique problem number and then described and recorded in an Excel report document. The action taken by the cleaning program or IEA Hamburg staff with respect to each problem was also recorded.

IEA Hamburg referred problems that could not be rectified automatically to the responsible NRC so that national center staff could check the original data collection instruments and tracking forms to trace the source of these errors. Wherever possible, a remedy was suggested, and national centers were asked to either accept it or propose an alternative. If a national center could not solve the issue through verification of the instruments or forms, IEA Hamburg applied a general cleaning rule to the files to rectify the error. When all automatic updates had been applied, SAS recoding scripts were used to directly apply any remaining corrections to the data files.

Checking Identification Variables

Each record in a data file requires a unique identification number. The existence of records with duplicate ID numbers in a file implies an error of some kind. If two records in the LaNA 2023 Linking Study database shared the same ID number and contained the same data, IEA Hamburg deleted one of the records and kept the other one in the database. In the rare cases where records contained different data and IEA staff found it impossible to identify which record contained the more reliable or complete version of the data, national centers were asked which record to keep. Duplicate IDs rarely occurred in the LaNA 2023 Linking Study, as the IEA DME prevented the data entry personnel from mistakenly entering duplicate ID and checksum combinations.

In addition to checking the unique student ID number, variables pertaining to student participation and exclusion statuses were checked. Confirming students' birth dates and dates of testing was also crucial to correctly calculate student age at the time of testing.

Checking Cross-File Linkages

As data on students and schools appeared in different data files, a process of linkage cleaning was implemented to ensure that the data files would correctly link together. The linking of the data files followed a hierarchical system of identification codes that included school, class, and student components. These codes linked the students with their class and school membership.

Linkage cleaning consisted of several checks to verify that student entries matched across achievement files, student context questionnaire data files, and scoring reliability files. In addition, at this stage, checks were conducted to ensure that student records were linked correctly to the appropriate schools. The Student Tracking Forms were crucial in resolving any anomalies. IEA Hamburg also liaised with NRCs about any problematic cases, providing the national centers with standardized reports listing all inconsistencies identified within the data.

As mentioned previously, IEA Hamburg conducted all cleaning procedures in close cooperation with the national centers. After national center staff had cleaned the identification variables and linkages, the cleaned databases with information about student participation and exclusion were passed on to the IEA Hamburg Sampling team, which used this information to calculate students' participation rates, exclusion rates, and student sampling weights.

Context Data Cleaning: Resolving Inconsistencies in Context Questionnaire Data

IEA Hamburg staff also checked questionnaire data for consistency across the responses given and unusual values for numerical background questionnaire items. For example, Question 6E in the school questionnaire asked whether there are computers or laptops available for students, and Question 8 asked for the number of computers schools have for use by students. Logically, if Question 6E is answered with "No," we would not expect a number listed for the computers in Question 8. Similarly, it would not make sense if a school is open for instruction for more than 365 days (Question 5A). Inconsistencies of this kind were flagged, and the national centers were asked to review these issues. After consulting with the respective NRC and the TIMSS & PIRLS International Study Center, remaining implausible values were recoded as "invalid."

IEA Hamburg also applied split variable checks to questions where the answer was coded into several variables. For example, Question 5 in the Student Questionnaire asked students the following: "Do you have any of these things at your home?" Student responses were captured in a set of five variables, each one coded as "Yes" if the corresponding "Yes" option was filled in and "No" if the "No" option was filled in. Occasionally, students checked some "Yes" boxes in the set but left the rest unchecked. In these cases, the options left unchecked were assumed to be "No" responses and were coded accordingly.

In addition, student responses to items on gender and age in the student questionnaire were checked against the tracking information provided by the School Coordinator or Test Administrator during the within-school sampling and test/questionnaire administration process. When information on gender or birth year and month was missing in the student questionnaire, this information, when available, was copied over from the tracking data to the questionnaire. If

discrepancies were found in gender and/or age between student questionnaire responses and existing tracking data, IEA Hamburg queried the case with the national center, and the national center investigated possible issues with the booklet assignment and the source of information was correct. If booklets were judged as being correctly assigned and the case could not be resolved, tracking data was used rather than questionnaire data since several NRCs indicated that students in their countries do not know their precise date of birth.

Handling of Missing Data

Two overarching types of entries were possible during the LaNA 2023 Linking Study data capture: valid response values and missing response values. During data capture, missing responses were assigned a value of "omitted/invalid" or "not administered." IEA Hamburg applied additional missing data codes to facilitate further analyses. This process led to three distinct types of missing response codes in the international database:

- Omitted or invalid: The respondent had a chance to answer the question but did not do so, leaving the corresponding item or question blank. This code was also used if the response was uninterpretable or considered as out-of-range.
- Not administered: The item or question was not administered to the respondent. The "not administered" code was used for those student test items that were not in the set of assessment blocks administered to a student, either deliberately (due to the rotation of assessment blocks), or in rare cases, due to technical issues such as misprints. This code was also used for response variables for those records that were included in the international database but did not contain a single response to one of the assigned questionnaires.
- Not reached: This applied to individual items of the student achievement test. It indicated those items that the student did not attempt at the end of the booklet when it was assumed the student stopped responding before reaching them. "Not reached" codes were assigned to all but the first omitted response after the last response was provided to the test. For example, in the response pattern "1 9 4 2 9 9 9 9 9 9", the value 2 was the last response, the response immediately following the 2 is considered omitted, and the five remaining omitted responses (9) are considered as not reached.

Data Cleaning Quality Control

The LaNA 2023 Linking Study was a complex study with high standards for data quality. Maintaining these standards required an extensive set of interrelated data checking and data cleaning procedures. To ensure that all procedures were conducted in the correct sequence, that no special requirements were overlooked, and that the cleaning process was implemented independently of the persons in charge, the data quality control process included the following:

- Thoroughly testing all data cleaning programs: Before applying the programs to real datasets, they were applied to simulation datasets containing all possible problems and inconsistencies.
- Registering all incoming data and documents in a dedicated database: The date of arrival was recorded as well as specific issues requiring attention.
- Carrying out data cleaning according to strict rules: Deviations from the cleaning sequence were not possible, and the scope for involuntary changes to the cleaning procedures was minimal.
- Documenting all systematic data recoding applied to all countries: All changes to data in the comprehensive cleaning documentation were recorded and provided to national centers.
- Logging every "manual" correction to a country's data files in a recoding script: Logging these corrections, which occurred only occasionally, allowed IEA Hamburg staff to undo specific changes or redo the whole manual cleaning process at any later stage of the data cleaning process.
- Repeating, on completion of data cleaning for a country, all cleaning steps from the beginning: This step allowed IEA Hamburg to detect any problems that might have been inadvertently introduced during the data cleaning process.
- Working closely with national centers at various steps of the cleaning process: IEA Hamburg provided national centers with the processed data files and accompanying documentation so that center staff could thoroughly review and correct any identified inconsistencies.

Interim Data Products

Before the LaNA 2023 Linking Study International Database was finalized, two major interim versions of the data files were sent to each country. For the two interim versions, each country only received its own dataset and did not have access to other countries' data.

The first version of the interim data files was sent as soon as the data could be considered "clean" regarding identification codes, linkage issues, and context data inconsistencies, and a second version of the data files was sent to countries when the weights and achievement plausible values were available and had been merged with the data files. This version, sent to the countries in October 2024, was structurally equivalent to the files to be published in the LaNA 2023 Linking Study International Database. It contained only those student records that were used in the analysis by the TIMSS & PIRLS International Study Center and satisfied the sampling standards.

Interim data products were accompanied by detailed data processing and cleaning documentation, a draft User Guide, codebooks, and summary statistics. The summary statistics were created by the TIMSS & PIRLS International Study Center and included univariate statistics

for all questionnaire variables for each country. For categorical variables, representing the majority of variables, the percentages of respondents choosing each of the response options were displayed.

For continuous numeric variables, various descriptive statistics were reported, including the minimum, maximum, mean, median, mode, and percentiles. For both types of variables, the percentages of missing data were reported. Additionally, for the achievement items, the TIMSS & PIRLS International Study Center provided item analysis and reliability statistics listing information such as the number of valid cases, percentages, percentage correct, Rasch item difficulty, and scoring reliability. These statistics were used for a more in-depth review of the data at the international and national levels in terms of plausibility, unexpected response patterns, etc.

Final Product—The LaNA 2023 Linking Study International Database

The extensive data cleaning effort described in this appendix helps ensure that the LaNA 2023 Linking Study International Database contains high-quality, internationally-comparable data. More specifically, the process ensures that:

- Information coded in each variable conforms to the international scheme
- National adaptations are reflected appropriately in all variables
- All entries in the database can be successfully linked across students and schools
- Sampling weights and student achievement variables are available for international comparisons

Citation

Neuschmidt, O. & Beyer, C. (2025). Appendix C: Survey operations and international database. In M. von Davier, L. Khorramdel, K. Reynolds, C. E. A. Aldrich, A. Bookbinder, A. Kennedy, & E. Gonzalez, *LaNA 2023 Linking Study Results* (pp. C.1-C.22). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

Appendix D: Psychometric Analysis

Liqun Yin, Ummugul Bezirhan, and Lale Khorramdel

Scaling the Achievement Data

Overview

The LaNA 2023 Linking Study assessment design relied on multiple matrix-sampling where each student received a subset of the entire pool of items, and no item was administered to all participants. Furthermore, there were two sets of four booklets with no common items between these two sets. The first set of booklets consisted of LaNA items (LaNA booklets), and the second set consisted of TIMSS and PIRLS linking items (TIMSS and PIRLS linking booklets). Each student was administered either a LaNA booklet or a linking booklet. See more information on the booklet design in Appendix A.

To provide comparable estimates of student achievement for mathematics and reading from such a design and to be able to relate these estimates to contextual variables, the LaNA scaling approach relies on latent regression population models, item calibration based on item response theory (IRT), and multiple imputations to obtain plausible values representing proficiency in mathematics and reading. The estimation of achievement on a comparable scale is established through a combination of common item linking within each of the two sets of booklets and common population linking between the two sets of booklets.

This section describes the psychometric analyses of the LaNA 2023 Linking Study achievement data. The TIMSS & PIRLS International Study Center implemented the IRT calibration and latent regression population modeling and conducted related analyses to ensure the quality and validity of the results. The psychometric models and population modeling applied in LaNA 2023 are the same as those used in TIMSS and PIRLS. A detailed description of how these methodologies are applied in TIMSS and PIRLS can be found in these projects' most recent technical reports (e.g., Bezirhan et al., 2023; von Davier, 2020). For general introductions to IRT, please refer to textbooks such as Embretson & Reise (2000), or encyclopedic works such as van der Linden (2017). For a formal and, alternatively, a more intuitive introduction to latent regression IRT models and plausible values, the reader is referred to von Davier & Sinharay (2013) and von Davier, Gonzalez, & Mislevy (2009), respectively.

Three main objectives were addressed in the psychometric analyses of the LaNA 2023 Linking Study achievement data: 1) Calibrating the new LaNA items and evaluating the fit of the TIMSS 2019 and PIRLS 2021 item parameters for linking items (for linking LaNA to TIMSS and PIRLS), 2) population modeling and drawing plausible values, and 3) transforming the

plausible values onto the TIMSS mathematics and PIRLS reading metrics. The TIMSS & PIRLS International Study Center conducted four major analysis phases to achieve these goals.

1. Item calibration: In the first phase, item parameters were estimated for the new LaNA items using multiple-group IRT models, while the fit of the TIMSS 2019 and PIRLS 2021 item parameters were evaluated for the linking items using fixed item parameter calibration.
2. Principal component analysis: In the second phase, principal components were extracted from the context data for each country for use in the population modeling.
3. Latent regression population modeling: In the third phase, latent regression population models were estimated (conditioning) for each country's data, and plausible values (PVs) were drawn for mathematics and reading.
4. Scale transformation: Finally, the generated plausible values were transformed onto the TIMSS 2019 and PIRLS 2021 metrics using the corresponding linear transformations.

The psychometric analyses of the LaNA 2023 Linking Study achievement data are described under four subsections according to these phases. Several quality checks and analyses were conducted iteratively throughout the analysis process. These analyses and their outcomes are described later in the section of this chapter.

Before describing each step of the scaling implementation, the treatment of item-level nonresponses is described in the following section.

Treatment of Item-Level Nonresponse (Omitted and Not-Reached)

Given the matrix-sampling design used in the LaNA 2023 Linking Study, students were randomly administered a subset of the available items. The responses to the items not administered were treated as missing at random and, therefore, ignored in the analysis (Little & Rubin, 1987; Rubin, 1976). However, nonresponse also occurred for other reasons, such as a student choosing not to answer an item, omitting it by mistake, or running out of time during the assessment period.

In scaling the LaNA 2023 achievement data, the treatment of omitted and not-reached items followed a strength-of-evidence approach initially introduced in PIRLS 2021 (Yin et al. 2023). When estimating parameters for the LaNA items, omitted and not-reached responses were treated as "not administered" and, therefore, not considered in the item parameter estimation. However, to account for nonresponses in proficiency estimation, two nonresponse indicator variables were calculated for each student: one for LaNA mathematics items and one for LaNA reading items. The variables indicated whether each student answered all items (1) or had at least one missing response (0). The nonresponse indicators were included alongside achievement items as predictors in the population models. The nonresponse of TIMSS and PIRLS linking items were treated the same way they were treated during their respective

operational analysis. Details on the treatment of the missing responses are described in the corresponding technical report (Foy et al., 2020; Yin et al., 2023).

To mitigate any potential undesirable effects of the nonresponse indicators on item parameter estimation, a stepwise approach was adopted to estimate parameters for the achievement items and the nonresponse indicators. First, only achievement items were included in the IRT calibration. In the second step, a 3PL IRT model was used to estimate the parameters for the nonresponse indicators, with all parameters for the achievement items fixed to those obtained in the first step.

IRT Item Calibration

Item Response Theory (IRT) is a framework for modeling the relationship between individuals' latent traits (e.g., mathematics and reading skills) and their responses to test items and allows for the estimation of item and person characteristics in the form of item and person parameters (Lord & Novick, 1968). It has become one of the most important tools of educational measurement as it provides a flexible framework for estimating proficiency scores from students' responses to test items. IRT is particularly well suited to handle data collection designs in which not all students are tested with all items.

Before item calibration and population modeling were conducted for the LaNA 2023 achievement data, an extensive item-by-item review was conducted based on classical item statistics calculated for all countries. The review was to evaluate the quality of the assessment items and to identify any unexpected or suspect item behaviors. As a result, two items (LL1A11 and LN11156) were excluded from the analyses for all countries. Items with problems in a specific country were excluded for generating plausible values for that specific country.

The IRT item calibration was conducted using the open-source MIRT package (Chalmers, 2012) in the R statistical programming language (R Core Team, 2020). Item parameters for the LaNA items were estimated using multiple-group IRT models. The TIMSS and PIRLS link item parameters were fixed to the values obtained in their corresponding operational assessments. This approach places the LaNA-specific mathematics items on the same metric as the TIMSS 2019 less difficult item parameters and LaNA reading items on the PIRLS 2021 item parameter metric. In the multiple-group IRT model, country was used as the grouping variable. Item parameters were estimated freely, but with equality constraints across groups. However, the ability distributions were estimated freely for each country. A two-parameter logistic (2PL) IRT model (Birnbaum, 1968) was used for all dichotomous items, while the generalized partial credit model (GPCM; Muraki, 1992) was applied to constructed-response items with more than one score point.

After initial calibration, the item fit was evaluated. Any TIMSS and PIRLS linking items that displayed poor fit were re-estimated and assigned new item parameters. In total, seven TIMSS and eight PIRLS linking items were identified as misfitting and subsequently treated as LaNA-specific items for which new item parameters were estimated. The treatment of nonresponse

remained unchanged. The item parameters and the estimates of nonresponse indicators estimated from the final item calibration are presented in this table.

Exhibits D.1 and D.2 provide the number of items included in the LaNA 2023 Linking Study calibration for mathematics and reading, respectively.

Exhibit D.1: Items for the LaNA 2023 Calibration — Mathematics

Item Type	Points	Items from LaNA linking booklets		Items from LaNA booklets		Total	
		Items	Points	Items	Points	Items	Points
Multiple Choice	1	21	21	79	79	100	100
Constructed Response	1	29	29			29	29
	2	2	4			2	4
Total		**52***	**54**	**79**	**79**	**131**	**133**

* 7 Mathematics Link Items estimated in the final model due to misfit.

Exhibit D.2: Items for the LaNA 2023 Calibration — Reading

Item Type	Points	Items from LaNA linking booklets		Items from LaNA booklets		Total	
		Items	Points	Items	Points	Items	Points
Multiple Choice	1	31	31	54	54	85	85
Constructed Response	1	26	26			26	26
	2	6	12			6	12
	3	2	6			2	6
Total		**65***	**75**	**54**	**54**	**119**	**129**

* 8 Reading Link Items estimated in the final model due to misfit.

Exhibit D.3 displays the sample sizes used in item calibrations. Cases were included in the calibration as long as they had at least one valid response in the corresponding subject. All student samples were weighted so that each participating country contributed equally to the calibration process. Six countries contributed data for the calibration, with an unweighted size of 30,509 students for mathematics and 30,255 students for reading.

Exhibit D.3: Sample Sizes for the LaNA 2023 Calibration

Country	Item Calibration	
	Mathematics	Reading
Burkina Faso	4,910	4,904
Egypt	5,382	5,334
Nigeria	4,158	4,048
Pakistan	8,713	8,669
Palestinian Nat'l Auth.	2,733	2,713
Senegal	4,613	4,587
Total	**30,509**	**30,255**

A multiple-group IRT model was utilized for item calibration, specifying country groups, resulting in 6 groups for each of the calibration models, one model for mathematics items, and one for reading items. While the item parameters were estimated to be equal across groups, the model allowed for estimating distinct ability distributions by country, to account for achievement differences between them properly. The parameters resulting from this calibration were then used to estimate student proficiency for all countries in the LaNA 2023 linking study.

Evaluating Item Fit

A series of IRT-based checks were performed during the item calibration phase to ensure reliable and accurate item parameter estimates of the LaNA achievement data. Several methods were used to verify the fit of the IRT model item parameters to the data. These included graphical model fit assessments based on item characteristic curves (ICCs) which are graphical representations of item response functions. ICCs can be used to visually compare the empirical and fitted item response functions and the empirical curves. In addition, quantitative checks of item fit were conducted using the root mean square difference (RMSD) and mean deviation (MD) statistics.

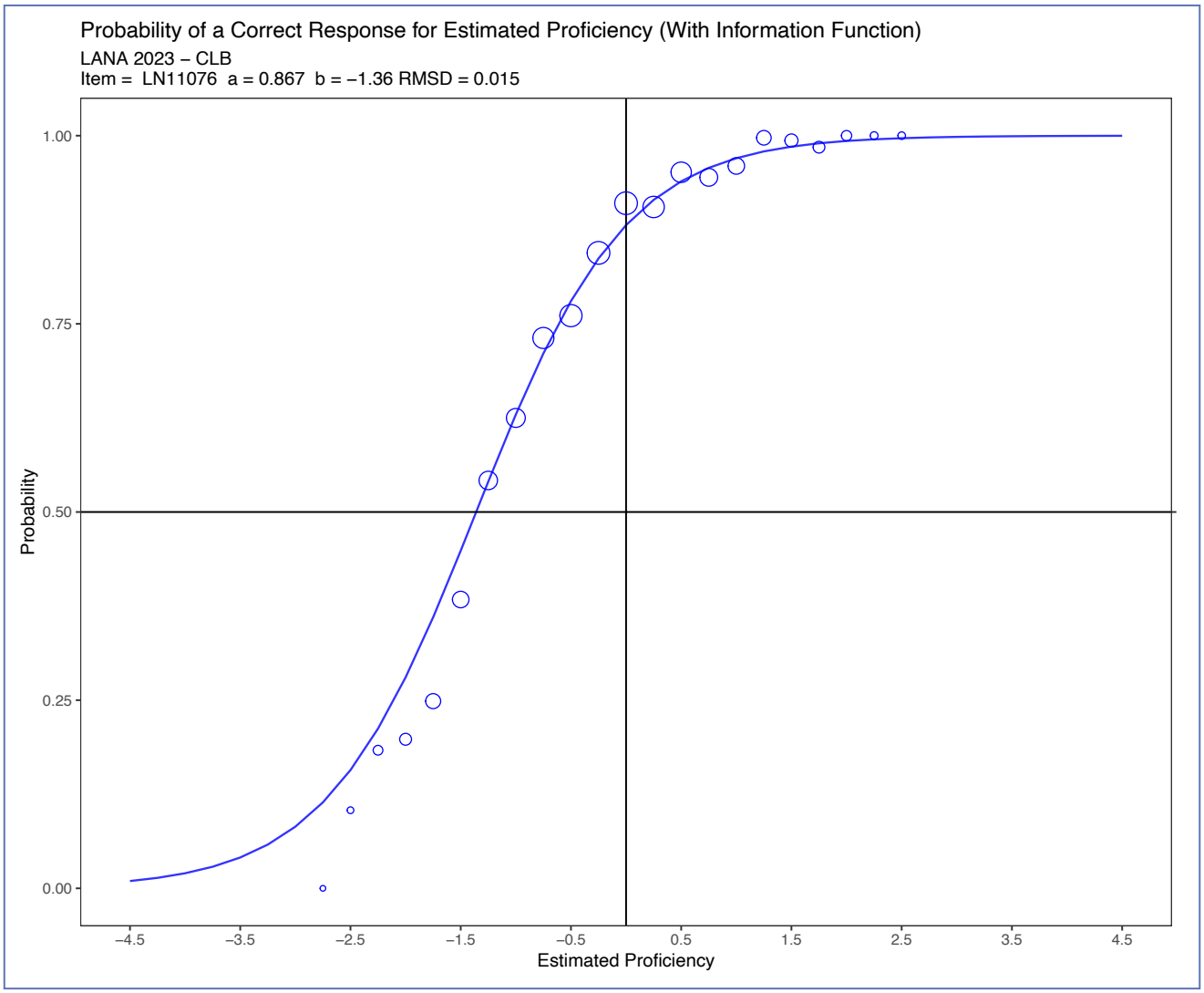
Graphical Model Fit using Item Characteristic Curves

Item fit was evaluated by visually comparing the fitted ICCs, generated based on the estimated item parameters, with the empirical ICCs, generated based on the observed response data. The empirical item response functions are based on an estimated latent ability distribution that uses the IRT model and, therefore, are also referred to as item response functions based on pseudo counts. When the empirical ICCs for an item fall near the fitted ICC, the IRT model fits the data well and provides an accurate and reliable measurement of the underlying proficiency scale.

ICC plots were examined for the TIMSS and PIRLS linking items and the LaNA items. The plot in Exhibit D.4 shows an example of the empirical and fitted ICCs for a dichotomously scored well-fitting item. The horizontal axis represents the proficiency scale on the logit metric, and

the vertical axis represents the probability of a correct response. The fitted ICC based on the estimated item parameters is shown as a solid line. Circles represent empirical results based on pseudo counts. The empirical results are obtained by first dividing the logit proficiency scale into intervals of equal size and then calculating the proportion of respondents within each of these segments that answer the item correctly. In the exhibit, the center of each circle represents this empirical proportion of correct responses. The size of each circle is proportional to the number of students contributing to the empirical proportion correct in its corresponding interval.

Exhibit D.4: Example Item Characteristic Curve for a Dichotomous Item



Item-Fit Statistics

In addition to the graphical model fit assessment, item fit was evaluated using the root mean square difference (RMSD) and mean deviation (MD) item-fit statistics. The RMSD is the square root of the average of squared differences between model-based item function and pseudo

counts, and the MD is the average of these differences between the empirical curve, shown as circles in Exhibit D.4, and the fitted curve, weighted by the number of students at each ability interval. The RMSD measures the overall magnitude of deviations in observed data from the estimated ICC, while the MD quantifies the direction and magnitude of these deviations for each item. The MD is particularly sensitive to differences in observed and model-based item difficulty, while the RMSD is sensitive to deviations in both item difficulty and item discrimination. RMSD and MD values were computed for all TIMSS and PIRLS linking items and LaNA items with RMSDs reported in the item parameter tables.

Misfitting items were identified using the median absolute deviation (MAD) outlier detection method on the RMSD and MD statistics. MAD is a robust measure of dispersion used as a flagging rule instead of an arbitrary cut-off value (von Davier & Bezirhan, 2022). This method flags an item as a possible misfit if its distance from the median of the absolute distances of all other observations exceeds a predetermined threshold. A MAD threshold of 4 was used to identify items that needed further evaluation. This choice reflects a conservative approach, as Miller (1991) defines cutoffs of 3 as conservative, 2.5 as moderate, and 2 as liberal. By selecting 4, we aimed to flag only the most significant deviations while minimizing false positives.

As discussed, after the initial calibration LaNA linking items were examined to identify the potential misfit by using the MAD outlier detection method. These linking items have fixed item parameters obtained from TIMSS 2019 and PIRLS 2021. Because these parameters may not align perfectly with the LANA data, the item fit was examined, and outliers were flagged. In the case of outliers, item parameters were freed in the final model, and new item parameters were estimated based on the LaNA data.

The RMSD values of the final IRT model (after any outlier treatment) are displayed graphically in Exhibits D.5 and D.6 for mathematics and reading, respectively. The MD values of the final

IRT model (after any outlier treatment) are similarly visualized in Exhibits D.7 and D.8. In these exhibits, the LaNA and linking items are sorted from smallest to largest item-fit statistic values.

Exhibit D.5: RMSD Statistics for Items in the LaNA 2023 Calibration—Mathematics

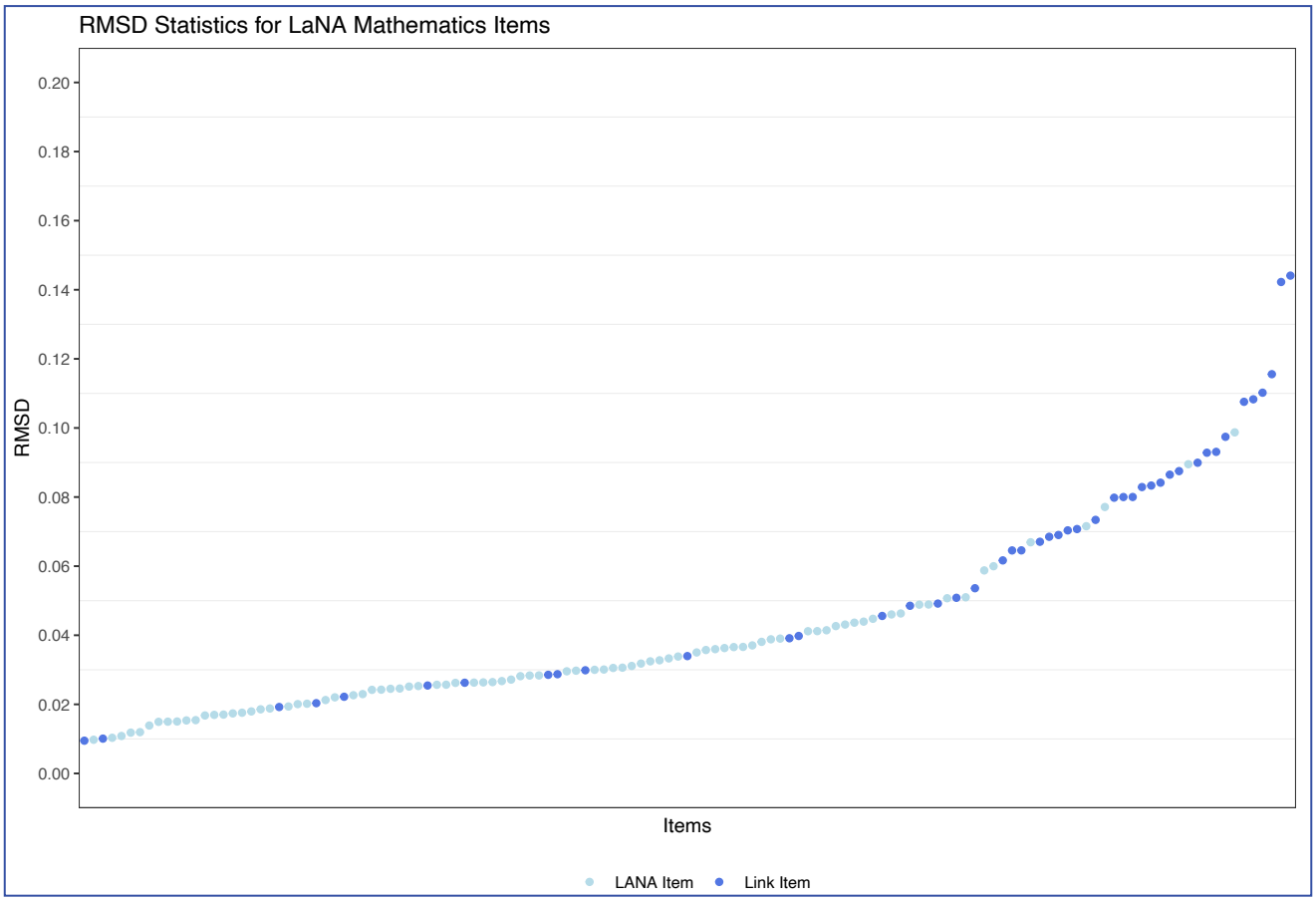


Exhibit D.6: RMSD Statistics for Items in the LaNA 2023 Calibration—Reading

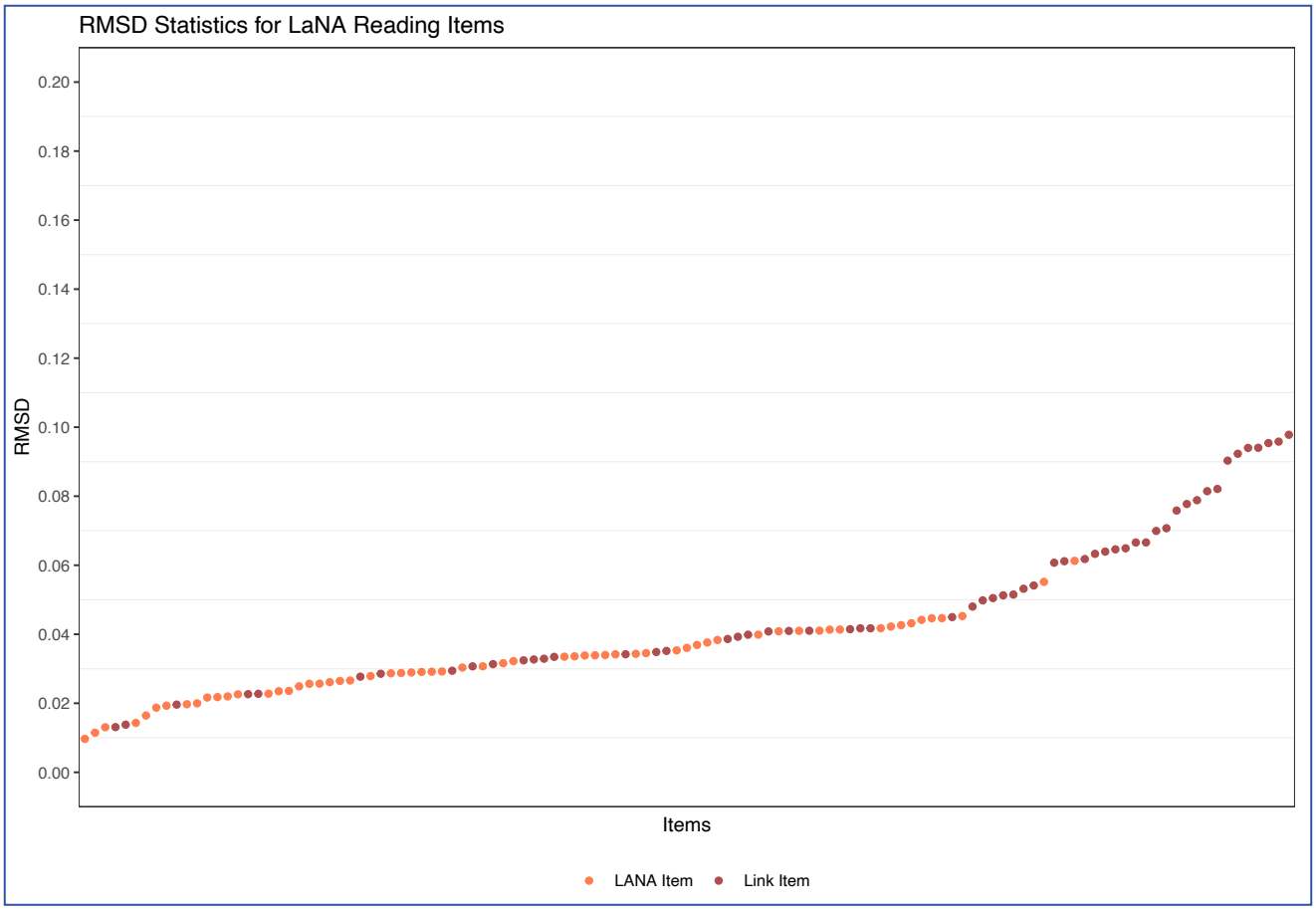


Exhibit D.7: MD Statistics for Items in the LaNA 2023 Calibration—Mathematics

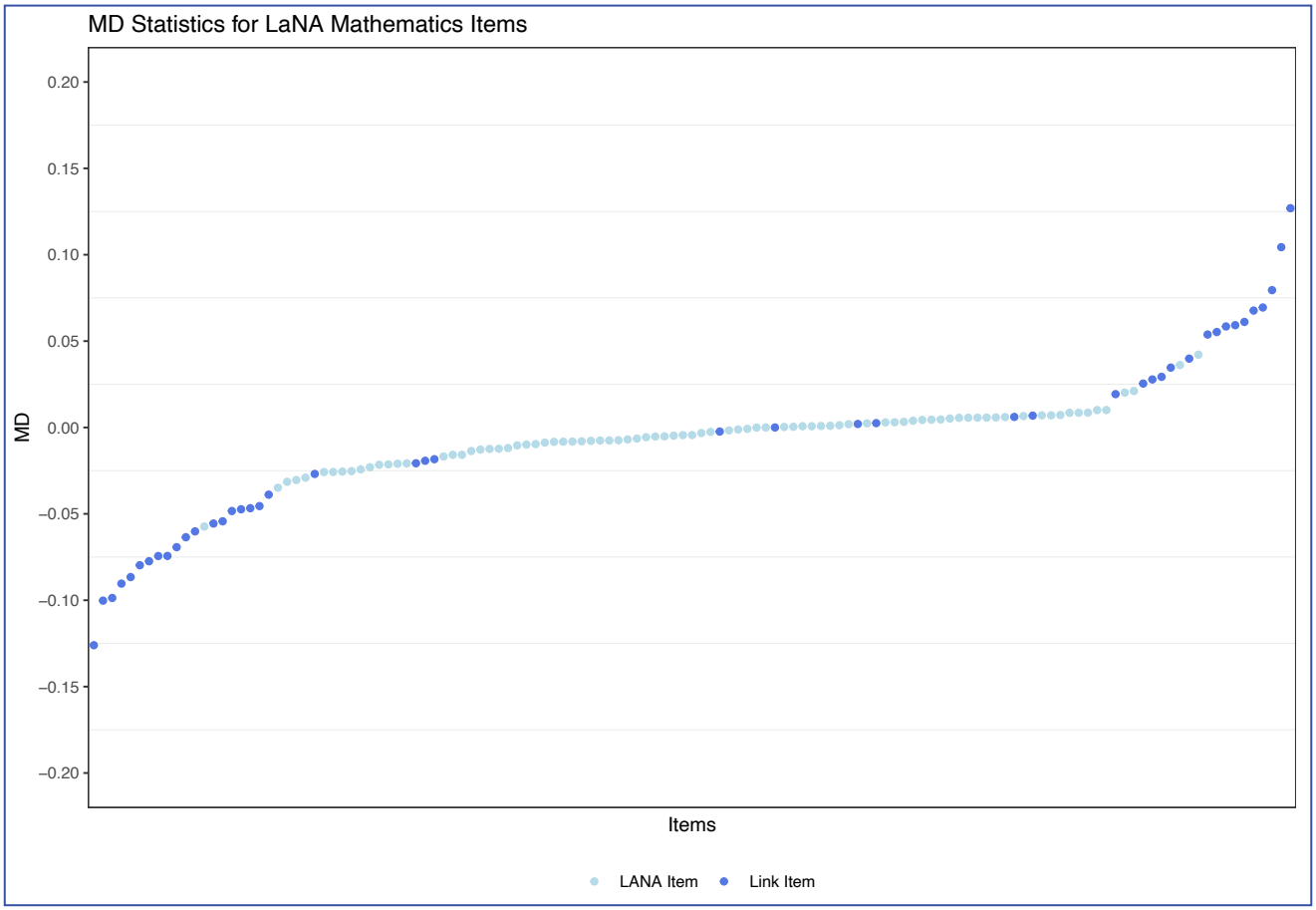
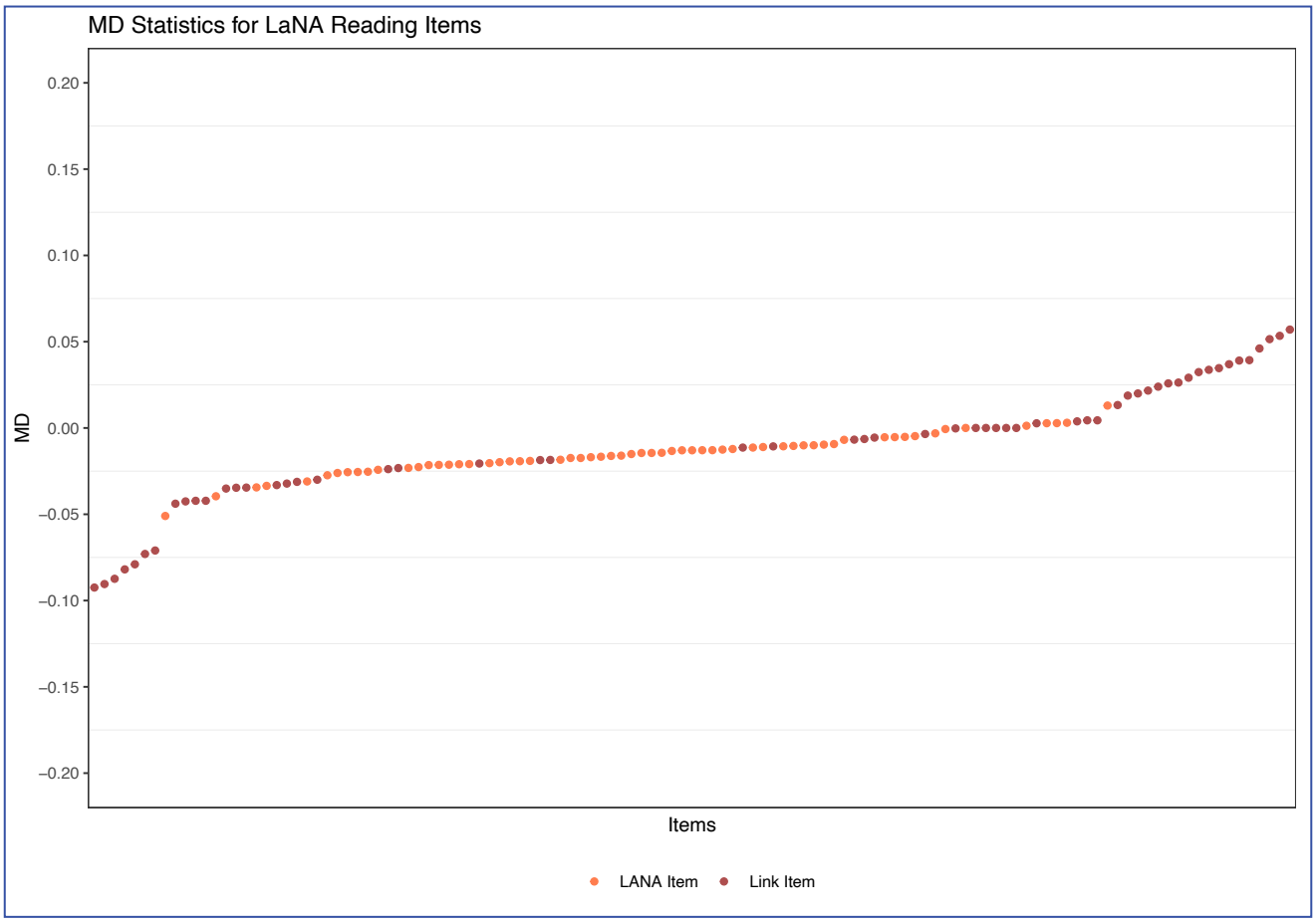


Exhibit D.8: MD Statistics for Items in the LaNA 2023 Calibration—Reading



Principal Component Analysis

The LaNA 2023 Linking Study used a latent regression or population model to estimate distributions of proficiencies. The population model is based on the likelihood function of an IRT model and a linear, latent regression of the proficiency on contextual data collected in background or context questionnaires (von Davier et al., 2006; von Davier et al., 2009). This approach can be viewed as an imputation model for the unobserved proficiency distribution that aims at obtaining unbiased group-level proficiency distributions by utilizing information about the extent to which background or context variables are related to the proficiency variable. Population models utilize a large number of context variables in the latent regression to avoid omission of any useful information (von Davier et al., 2006; von Davier et al., 2009; von Davier & Sinharay, 2013).

To reduce the number of context variables and avoid overparameterization, a principal component analysis (PCA; e.g., Hotelling, 1933; see also Jackson, 1991; Joliffe, 2002) was conducted to eliminate collinearity by identifying a smaller number of orthogonal predictors that account for most of the variation in the background variables.

The LaNA 2023 Linking Study followed the practice established by TIMSS and PIRLS of using a PCA and selecting principal components that account for 90% of the background variance (e.g., Foy, et al., 2020; Yin, et al., 2023). For the PCA, categorical variables with a small number of response options (usually eight or fewer) were dummy-coded to represent all response options, including responses coded as "not administered," "not applicable," and "omitted." The year of the student's birth was recoded using criterion scaling (Beaton, 1969) by replacing the response value with mean interim achievement scores, EAPs of mathematics and reading, of all students with that response. The PCA was conducted separately for each of the participating countries.

In addition to the principal components, students' gender (dummy coded), the language of the test (dummy coded), and an indicator of the classroom in the school to which a student belongs (criterion scaled) were included as primary conditioning variables.

Exhibit D.9 provides the number of variables used in the conditioning models for the proficiency estimation of the LaNA 2023 Linking Study data.

Exhibit D.9: Conditioning Model Dimensionality for the LaNA 2023 Data

Country	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained
Burkina Faso	2	94	56
Egypt	3	94	59
Nigeria	2	94	58
Pakistan	4	95	57
Palestinian Nat'l Auth.	2	94	58
Senegal	2	94	56

Estimating the Population Model and Generating Plausible Values

ETS's MGROUP programs (Rogers et al., 2006; Sheehan, 1985) were used to estimate the latent regression model separately for each participating country and to generate plausible values (PVs) for the LaNA 2023 data. These programs take the students' responses to the items, the item parameters, the estimates of indicators estimated at the calibration stage, and the conditioning variables as input. They then generate as output the estimated regression coefficients, the residual variance-covariance matrix, and plausible values that represent the posterior distribution of student proficiency given their achievement and contextual data (e.g., Mislevy, 1991; Thomas, 1993; von Davier et al., 2006; von Davier & Sinharay, 2013).

To improve the reliability of each scale, and to account for the correlation between the two subjects (mathematics and reading), a two-dimensional latent regression model was used to generate plausible values simultaneously for the mathematics and reading scales. Following the

estimation of the latent regression model for imputation, five plausible values were generated (imputed) for each student in both subjects. All participating students were included in the population model. Population models were estimated separately for each LaNA 2023 Linking Study country.

Scale Transformation

After the plausible values were generated, they were transformed onto TIMSS and PIRLS reporting metrics through a series of linear transformations. The TIMSS mathematics reporting metric was originally established in TIMSS 1995 by setting the mean of the pooled, fourth-grade sample to 500, and the standard deviation to 100. The subsequent TIMSS cycles were linked to 1995 through several steps that involved concurrent calibrations and scale transformations. In addition, the extension of the regular TIMSS assessment and TIMSS 2019 less difficult mathematics was also linked to the TIMSS trend reporting metric. Similarly, the PIRLS reporting metric was established in PIRLS 2001 by setting the mean of the national average scores for all countries that participated in 2001 to 500 and the standard deviation to 100. The subsequent PIRLS cycles, including PIRLS 2021, were then placed on the PIRLS reporting metric.

For LaNA mathematics, the scale transformation constants established for the TIMSS 2019 less difficult mathematics data were used to transform the corresponding plausible values on the TIMSS reporting metrics (Foy et al., 2020). To link LaNA reading to the PIRLS reporting metrics, the transformation constants established for the PIRLS 2021 paper data (Yin et al., 2023) were used to transform the reading plausible values generated from the LaNA assessment.

How to Use Plausible Values

As described in the previous section, for each subject (mathematics and reading), five plausible values representing the proficiency variable are drawn. It should be noted that plausible values are not individual test scores in the traditional sense and should only be used for analyses to report summary statistics for groups of students.

The general procedure for analyzing plausible values follows the basic principle of working with multiple imputations. That is, performing any statistical analysis five times — once for each set of plausible values — and aggregating the five sets of results (Mislevy et al., 1992).

For any given achievement-based statistic t, such as the average reading achievement for a country, estimating that statistic from each plausible value yields five estimates t_m, m = 1,…,5. The final estimate of that statistic, t_0, is the average of these five estimates:

$$t_0 = \frac{1}{5}\sum_{m=1}^{5} t_m$$

Uncertainty about students' proficiency is a function of the number of items administered and the interaction of the item characteristics and student proficiency, among other factors.

Proficiency estimates have an associated variability due to measurement error. These plausible values are proficiency estimates that incorporate the portion of measurement uncertainty that can be quantified. LaNA follows the customary procedure of imputing multiple plausible values for each student and using the variability among them as a measure of that uncertainty.

To quantify the uncertainty of that statistic, t_0, the standard error associated with that statistic is estimated, which serves to calculate confidence intervals (CI). For statistics reporting student achievement, which are based on plausible values, standard errors are estimated based on two components: the sampling variance and imputation variance. The first reflects the uncertainty due to generalizing from a student sample to the entire student population from which it was drawn. The second reflects uncertainty due to inferring students' achievement estimates from their observed performance on a subset of achievement items and their background data. The procedures for calculating sampling variance and imputation variance are the same as the methods used in TIMSS and PIRLS (Foy & Almaskut, 2023; Foy & LaRoche, 2020; Foy & Yin, 2015). The sampling variance in this context is the average of the sampling variances from the five plausible values, as follows:

$$Var_{jrr}(t_0) = \frac{1}{5}\sum_{m=1}^{5} Var_{jrr}(t_m)$$

where

$$Var_{jrr}(t_m) = \frac{1}{2}\sum_{h=1}^{125}\sum_{j=1}^{2}(t_{mhj} - t_m)^2$$

and is the appropriate Jackknife Repeated Replication (JRR) estimate based on plausible value m computed using the set of replicate sampling weights from sampling zone h when school j is included (Foy & Yin, 2015; Foy & LaRoche, 2020; Foy & Almaskut, 2023), where h could be up to 125.

The imputation variance of the statistic t_0 is simply the variance of the five results from the plausible values, computed as follows:

$$Var_{imp}(t_0) = \frac{6}{5}\sum_{m=1}^{5}\frac{(t_m - t_0)^2}{(5-1)}$$

where the factor $\frac{6}{5}$ is a correction factor required by the multiple imputation methodology. The associated total variance estimate of the statistic t_0 is the sum of the sampling variance and imputation variance as follows:

$$Var_{tot}(t_0) = Var_{jrr}(t_0) + Var_{imp}(t_0)$$

The square root of the total variance is then the estimated standard error for the statistic, t_0, based on plausible values:

$$SE(t_0) = \sqrt{Var_{tot}(t_0)}$$

In the LaNA 2023 Linking Study, the achievement results for countries or sub-groups are reported by using the 95% confidence interval to reflect the uncertainty of the estimates of statistics of interest. The confidence interval is constructed by using

$$CI = t_0 \pm Z_{crit} * SE(t_0)$$

Where Z_{crit} is the value from the Z distribution, or Z score, that is used to determine the critical region for a normal distribution, to be consistent with previous reporting in TIMSS and PIRLS. The 95% confidence interval (CI) of the statistic, t_0, is expressed as follows:

$$((t_0 - 1.95996 * SE(t_0), (t_0 + 1.95996 * SE(t_0))$$

The use and interpretation of CIs in the LaNA 2023 Linking Study is described in the introduction of the international results section at the beginning of this report.

Scaling the Context Questionnaire Data and Deriving Context Variables

In addition to the data on students' mathematics and reading achievement, LaNA collects data about the home, school, and classroom contexts that relate to learning. Some of the items in the LaNA 2023 Context Questionnaires were combined into scales measuring a common and dominant underlying latent construct related to achievement. To facilitate interpretation of the context scale results, questionnaire respondents (students or principals) are classified based on their responses into regions corresponding to high, middle, and low levels of the construct.

There are several steps followed when analyzing the TIMSS context questionnaire responses, including item calibration, evaluation of item fit, estimation of scale scores, scale transformation onto a reporting metric, creation of scale regions for reporting, and validation of the results. The derived context scales are included in the LaNA 2023 Database as continuous scale scores and categorical scale region variables.

The next sections describe the procedures for constructing, interpreting, validating, and reporting the LaNA 2023 Context Questionnaire Scales.

Constructing Context Scales

To construct a context scale in LaNA, first, the items in the context questionnaires that measure a common latent construct were identified to be used as component variables in each scale.

Six context scales were created in the LaNA 2023 Linking Study. The number of items per scale varies from three to 10 items. The LaNA context scales were analyzed using the Rasch partial credit model (PCM; Masters, 1982). For each context scale, the item parameters of each item were estimated using the pooled sample of responses, item fit was evaluated, and the estimated item parameters were used to estimate the scale score on the latent construct for each respondent. These scale scores were then transformed onto a reporting metric.

The estimation of the item parameters was conducted using the ConQuest 2.0 software (Wu et al., 2007) on the combined data from the six participating countries. Each country contributed equally to the calibration through the use of "senate weights." Only cases with two or more valid item responses on a scale were included for the calibration without any weight adjustments.

Exhibit D.10 presents the item parameters for the *Students' Attitudes Toward Reading and Mathematics* scale as an example. For each item, the delta parameter δ_i shows the estimated overall location of the item on the scale, and the tau parameters τ_{ij} show the location of the steps, expressed as deviations from delta ($b_{ij} = \delta_i - \tau_{ij}$). To remove scale indeterminacy in the calibration, the "items constraint" in ConQuest was used to set the mean of the item difficulty (location) parameters to zero.

Exhibit D.10: Item Parameters for the Students' Attitudes Toward Reading and Mathematics Scale

Item	delta	tau_1	tau_2	tau_3	Infit	Q-Index
ASBG11A	–0.3374	0.0251	–0.4839	0.4588	1.02	0.10
ASBG11B	0.1456	–0.2360	–0.6952	0.9312	1.01	0.02
ASBG12A	–0.1189	–0.0291	–0.5805	0.6096	1.01	0.08
ASBG12B	0.3107	–0.4687	–0.7181	1.1868	0.96	0.05

Two item fit statistics were used to evaluate how well the model fits the data for each context scale: the Rasch infit statistic (Wright & Masters, 1982) and the Q-index (Rost & von Davier, 1994). These are shown in the last two columns in Exhibit D.10. The Rasch infit item statistic is a residual-based measure of how well the estimated item parameters fits the data, with values ranging from 0 to infinity. A value of 1.0 corresponds to the optimal fit to the Rasch model. A value of 1.4 was used as an upper bound to indicate potential misfit in the LaNA 2023 Linking Study, based on Adams and Khoo (1996) and Preuschoff (2010). When using the Q-index, the fit of an item is evaluated with regard to the conditional probability of its observed response vector and does not depend on the item parameters. The Q-index is standardized and ranges from 0 to 1, with a smaller value indicating a better fit. Overall, all the items in the LaNA context scales have Rasch infit statistics close to 1 and Q-index value below 0.2, both indicating a good fit.

In addition to item fit evaluations, the item response patterns were also examined before calibration. The *School Climate* scale has very sparse responses in the "very low" category. These sparse responses provided limited information when estimating item parameter for that

category. Therefore, two item response categories, “low” and “very low”, were collapsed to create the scale.

Once the calibration was completed and item fit was evaluated, individual scores for each respondent were estimated using weighted maximum likelihood estimation (Warm, 1989). All cases with valid responses for at least two items on a scale were assigned scale scores. These scale scores produced by the weighted likelihood estimation were on the logit metric with estimated values ranging from approximately –5 to +5.

To convert to a more convenient reporting metric, a linear transformation was applied to the distribution of logit scores for each scale so that the resulting distribution across all countries had a mean of 10 and a standard deviation of 2. Exhibit D.11 presents the scale transformation constants applied to the LaNA countries’ distribution of logit scores to transform them to the (10,2) reporting metric using the *Students’ Attitudes Toward Reading and Mathematics* scale as an example. Each scale was transformed using a different set of transformation constants to put the estimates on the reporting scale. A link to the transformation constants is provided at the end of this appendix.

Exhibit D.11: Scale Transformation for the Students’ Attitudes Toward Reading and Mathematics Scale

Scale Transformation Constants	
A = 7.054199	Transformed Scale Score = 7.054199 + 1.692110 • Logit Scale Score
B = 1.692110	

Reporting Context Scales and Creating Scale Regions

To enable reporting and interpretation of the context scales in meaningful categories, the scores on the LaNA context scales were divided into high, middle, and low regions. The scale region cutpoints were established by assigning a numeric value to each response category such that each respondent’s responses to the scale items could be expressed as a raw score. For example, assigning 0 to “Disagree a lot,” 1 to “Disagree a little,” 2 to “Agree a little,” and 3 to “Agree a lot,” results in raw scores on the *Student Attitudes Toward Reading and Mathematics* scale ranging from 0 (“disagree a lot” with all four statements) to 12 (“agree a lot” to all four statements).

The particular response combinations that defined the regions’ boundaries, or cutpoints, were identified using a data-driven method, Latent Class Analysis for cut scores (LCA-CS) (Brown, 2007; Jiao et al., 2011; Yin et al., 2024). The LCA-CS method builds on the classical Latent Class Analysis (Lazarsfeld, 1955), a latent variable modeling technique for categorical data that identifies groups based on a statistical optimality criterion.

A property of a Rasch scale is that each raw score has a unique scale score associated with it. Once the raw cutpoints were determined using the LCA-CS method, the corresponding scale

scores on the scale metric were then identified. To facilitate reporting, the highest cut point was rounded down to the first decimal place to ensure that those with an unrounded scale score at the cut point were included in the highest region. Similarly, the lower cut point was rounded up. As shown in Exhibit D.12 on the example of the *Student Attitudes Toward Reading and Mathematics* scale, a raw score of 7 corresponds to a scale score of 7.3 (rounded up) and a raw score of 11 corresponds to a scale score of 10.0 (rounded down).

Exhibit D.12: Equivalence Table of Raw and Transformed Scale Scores for the Student Attitudes Toward Reading and Mathematics Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.2276	
1	4.5395	
2	5.1920	
3	5.6670	
4	6.0717	
5	6.4468	
6	6.8229	
7	7.2041	7.3
8	7.6406	
9	8.1750	
10	8.8992	
11	10.0257	10.0
12	12.2581	

Reliability and Validity of Context Scales

As evidence that the context questionnaire scales provide reliable measurement across countries, Cronbach's Alpha reliability coefficients (Peterson, 1994; Taber, 2017) were computed for each scale for every country, and a PCA of the responses to the scale was conducted within each country. Exhibit D.13 presents the results of this analysis for the *Student Attitudes Toward Reading and Mathematics* scale as an example. The Cronbach's alpha measure of internal consistency for this scale was generally acceptable. The exhibit also shows that the percentage of variance among the scale items accounted for by the first principal component was sufficiently high in most countries, indicating that the items measuring the contextual construct could be adequately represented by a single scale. Moreover, the component loadings of each item in the principal component analysis are positive and substantial, indicating a strong correlation between each item and the scale in every country.

Exhibit D.13: Cronbach's Alpha Reliability Coefficient and PCA of the Items in the Student Attitudes Toward Reading and Mathematics Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item			
			ASBG11B	ASBG11C	ASBG12B	ASBG12C
Burkina Faso	0.72	54.88	0.76	0.74	0.75	0.71
Egypt	0.66	49.71	0.66	0.71	0.69	0.76
Nigeria	0.78	61.00	0.77	0.76	0.80	0.80
Pakistan	0.62	47.38	0.63	0.66	0.71	0.76
Palestinian Nat'l Auth.	0.59	44.83	0.58	0.70	0.62	0.76
Senegal	0.63	48.34	0.65	0.73	0.69	0.71

Detailed information on each of the LaNA 2023 Linking Study Context scales, including item parameter estimates and scale statistics described above, can be found in this table.

Citation

Yin, L., Bezirhan, U., & Khorramdel, L. (2025). Appendix D: Psychometric analysis. In M. von Davier, L. Khorramdel, K. Reynolds, C. E. A. Aldrich, A. Bookbinder, A. Kennedy, & E. Gonzalez, *LaNA 2023 Linking Study Results* (pp. D.1–D.19). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

References

Adams, R. J.and Khoo, S-T. (1996). ACER Quest: the interactive test analysis system. [Computer software]. Version 2.1. Camberwell, Victoria: Australian Council for Educational Research. https://research.acer.edu.au/measurement/3/

Birnbaum, A. (1968). Some latent trait models and their use in inferring a student's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Beaton, A. E. (1969). Criterion scaling of questionnaire items. *Socio-Economic Planning Sciences, 2*, 355–362.

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

Brown, R. S. (2007). Using latent class analysis to set academic performance standards. *Educational Assessment, 12*(3–4), 283–301.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Embretson, S.E., & Reise, S.P. (2000). Item Response Theory for Psychologists (1st ed.). *Psychology Press*. https://doi.org/10.4324/9781410605269

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

Foy, P., Fishbein, B., von Davier, M., & Yin, L. (2020). Implementing the TIMSS 2019 scaling methodology. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report* (pp. 12.1–12.146). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/methods/chapter-12.html

Foy, P., & LaRoche, S. (2020). Estimating standard errors in the TIMSS 2019 results. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report* (pp. 14.1–14.60). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/methods/chapter-14.html

Foy, P., & Yin, L. (2015). Data analyses with IEA's TIMSS and PIRLS international databases. *Contemporary Educational Research Quarterly, 23*(4), 41–61. http://contemporary.cere.ntnu.edu.tw/sites/default/files/public/23%284%29_2_0.pdf

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*(6), 417–441. https://doi.org/10.1037/h0071325

Jackson, J.E. (1991) A User's Guide to Principal Components. John Wiley & Sons, New York. https://doi.org/10.1002/0471725331

Jiao, H., Lissitz, R. W., Macready, G., Wang, S., & Liang, S. (2011). Exploring levels of performance using the mixture Rasch model for standard setting. *Psychological Test and Assessment Modeling, 53*(4), 499–522.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer. https://doi.org/10.1007/b98835

Lazarsfeld, P. F. (1955). Recent developments in latent structure analysis. *Sociometry, 18*(4), 391–403. https://doi.org/10.2307/2785875

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: J. Wiley & Sons. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119013563

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. https://doi.org/10.1007/BF02296272

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*, 177–196. https://doi.org/10.1007/BF02294457

Mislevy, R. J., Beaton, A., Kaplan, B. A., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29*(2), 133–161.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159–177. https://doi.org/10.1002/j.2333-8504.1992.tb01436.x

Peterson, A. R. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research, 21*(2), 381–391. http://dx.doi.org/10.1086/209405

Preuschoff, A. C. (2010). *Using TIMSS and PIRLS to construct global indicators of effective environments for learning* [Doctoral Dissertation, Boston College, Lynch School of Education]. http://hdl.handle.net/2345/1947

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

Rogers, A., Tang, C., Lin, M.-J., & Kandathil, M. (2006). DGROUP [computer software]. Princeton, NJ: Educational Testing Service.

Rost, J. & von Davier, M. (1994). A conditional item-fit index for Rasch models. *Applied Psychological Measurement, 18*(2), 171–182. https://doi.org/10.1177/014662169401800206

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592. https://doi.org/10.2307/2335739

Sheehan, K. M. (1985). M-Group: Estimation of group effects in multivariate models [computer software, Version 3.2]. Princeton, NJ: Educational Testing Service.

Taber, K. (2017). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Thomas, N. (1993). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics, 2*, 309–322. https://www.jstor.org/stable/1390648

van der Linden, W. J. (2017). *Handbook of Item Response Theory, Volume One: Models*. (Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences). Chapman and Hall/CRC. https://doi.org/10.1201/9781315374512

von Davier, M. (2020). TIMSS 2019 scaling methodology: Item Response Theory, population models, and linking across modes. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report* (pp. 11.1–11.25). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/methods/chapter-11.html

von Davier, M., & Bezirhan, U. (2022). A robust method for detecting item misfit in large-scale assessments. *Educational and Psychological Measurement, 7*(2), 110–114. https://doi.org/10.1080/15366360903117079

von Davier, M., Gonzalez, E., & Mislevy, R.J. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Eds.), *IERI Monograph Series: Issues and Methodologies in Large Scale Assessments* (Vol. 2, pp. 9–36). Retrieved from https://ierinstitute.org/fileadmin/Documents/IERI_Monograph/Volume_2/IERI_Monograph_Volume_02_Chapter_01.pdf

von Davier, M., & Sinharay, S. (2013). Analytics in international large-scale assessments: Item response theory and population models. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 155–174). Boca Raton, FL: CRC Press. https://doi.org/10.1201/b16061

von Davier, M., Sinharay, S., Oranje, A., & Beaton, A. (2006). Statistical procedures used in the National Assessment of Educational Progress (NAEP): Recent developments and future directions. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. 26: Psychometrics). Amsterdam, Netherlands: Elsevier.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450. https://doi.org/10.1007/BF02294627

Wu, M. L., Adams, R. J, Wilson, M. R., & Haldane, S. (2007). Conquest 2.0 [computer software]. Camberwell, Australia: Australian Council for Educational Research.

Yin, L., Bezirhan, U., Fishbein, B., & Foy, P. (2023). 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., Bezirhan, U., & von Davier, M. (2024). Improving context scale interpretation using LCA for cut scores [Paper presentation]. International Meeting of the Psychometric Society 2024, Prague, Czech Republic. https://www.psychometricsociety.org/imps-2024

Appendix E: Scale Anchoring and Description of New Basic LaNA Benchmarks

Lale Khorramdel, Charlotte E. A. Aldrich, and Ann Kennedy

Scale Anchoring

Following the TIMSS and PIRLS procedures, LaNA achievement results are summarized using extensions of item response theory (IRT) and reported on the LaNA achievement mathematics and reading scales (see Appendix D), which are linked to the TIMSS mathematics and the PIRLS reading achievement scales. Country-level score distributions provide information about how students' mathematics and reading achievement compares across countries and whether achievement is improving or declining, overall or for certain subgroups, over time. However, comparing average achievement internationally is only one part of the picture. The range of mathematics and reading achievement within countries, paired with information about what students at different levels of the scale know and can do, provide a much more nuanced picture of student ability to policymakers compared to only providing an estimate of the average scale score per country.

To provide this information, it is important to understand the mathematics and reading competencies associated with different regions along the LaNA achievement scales. For example, what does it mean for a country to have a confidence interval of average mathematics or reading achievement that ranges from 313 to 351, or from 386 to 407? How are these students within these achievement ranges characterized? What mathematics or reading skills or competencies can students within different achievement ranges expected to demonstrate? LaNA emphasizes characterizing the progression of student achievement through a scale anchoring approach to create descriptions of achievement at established international benchmarks.

This appendix describes the scale anchoring process for creating descriptions of the new LaNA 2023 Basic International Benchmarks for Mathematics and Reading achievement, which complement the existing TIMSS and PIRLS International Benchmarks. Together, these benchmarks help contextualize LaNA results by providing information about what students know and can do at different points along the LaNA achievement scales and the TIMSS and PIRLS achievement scales through the newly created link.

While descriptions exist for the TIMSS 2019 and PIRLS 2021 International Benchmarks, descriptions for the newly established LaNA 2023 Basic International Benchmarks were needed. More specifically, TIMSS and PIRLS have identified four points along the TIMSS mathematics and PIRLS reading achievement scales to use as International Benchmarks of achievement:

Advanced International Benchmark (625), High International Benchmark (550), Intermediate International Benchmark (475), and Low International Benchmark (400). The LaNA 2023 Linking Study builds on these efforts by extending the scales through an additional benchmark for each subject: the Basic International Benchmark (325).

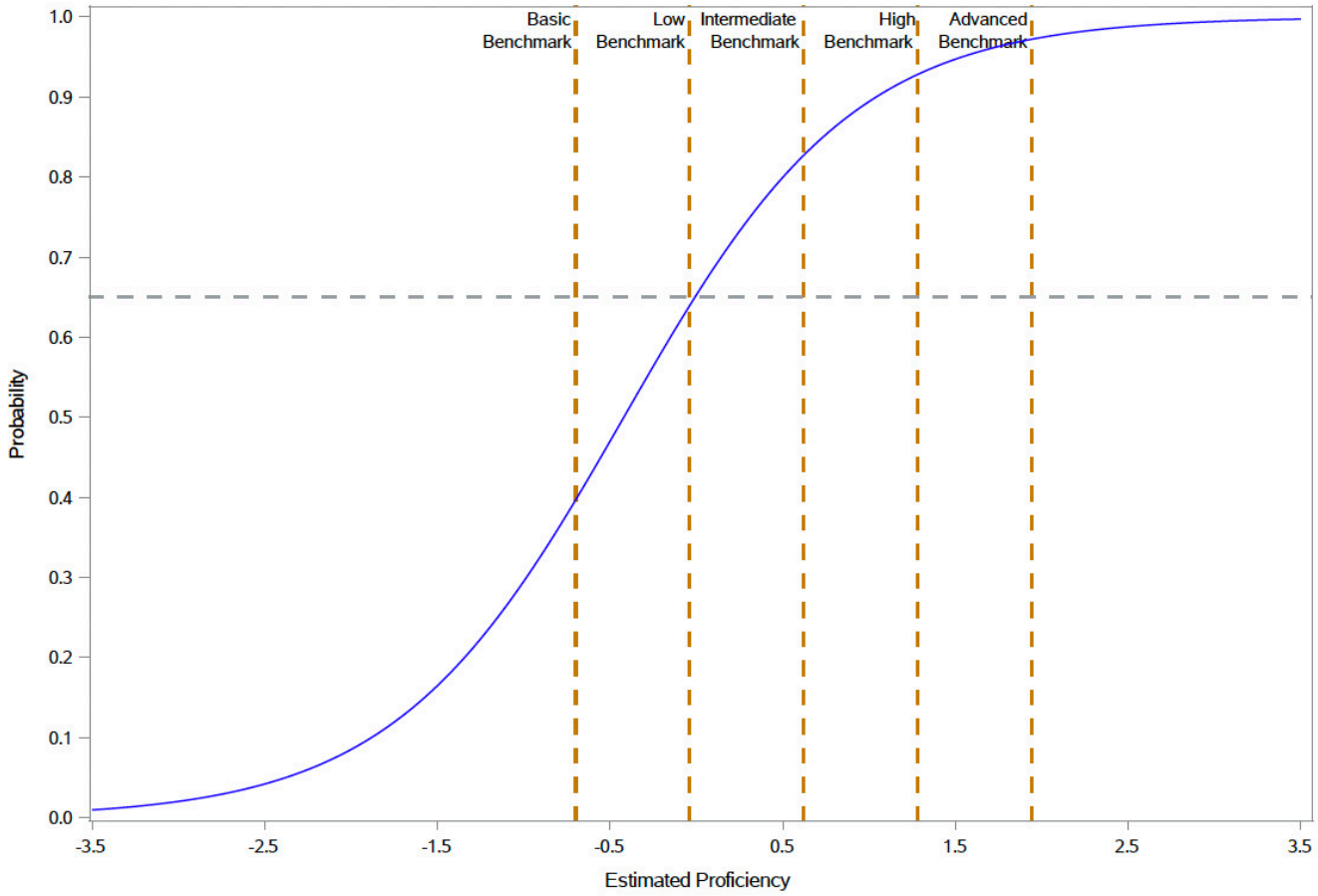
The following sections of this chapter summarize the TIMSS and PIRLS scale anchoring method, which has also been used to develop the LaNA 2023 Basic International Benchmark descriptions for Mathematics and Reading. This includes the classification of items through scale anchoring analysis and the writing of the benchmark descriptions. Additionally, the descriptions of the LaNA Basic International Benchmarks are provided. The scale anchoring method is described in detail in the *TIMSS 2019 Methods and Procedures* Chapter 15 (Mullis & Fishbein, 2020) and the *PIRLS 2021 Methods and Procedures* Chapter 14 (Wry et al., 2023). The descriptions of the existing TIMSS 2019 and PIRLS 2021 International Benchmarks for Mathematics and Reading for fourth-grade students can be found in the TIMSS 2019 and PIRLS 2021 international reports (Mullis et al., 2020, 2023).

Classifying the Items

The LaNA 2023 scale anchoring followed the steps outlined in the TIMSS and PIRLS methods and procedures (Chapter 15 and Chapter 14, respectively) and was conducted separately for mathematics and reading. An important feature of the scale anchoring method is that it yields descriptions of the mathematics and reading competencies demonstrated by students reaching each of the International Benchmarks, reflecting the TIMSS mathematics and the PIRLS reading frameworks.

The first stage in the scale anchoring process involves identifying items that anchor at each International Benchmark based on the IRT parameters estimated as part of achievement scaling for the LaNA 2023 Linking Study (described in Appendix D). Exhibit E.1 provides an example item characteristic function that illustrates the relationship between the model-based probability of correctly responding to a mathematics or reading item and the estimated student ability on the achievement scale. The vertical reference lines indicate the location of each International Benchmark in the IRT logit metric, and the values on the vertical axis correspond to the probability of answering the item correctly.

Exhibit E.1: Example Item Characteristic Function for Scale Anchoring Analysis



As can be seen, the probability of a correct response increases from basic to higher International Benchmarks. Students with achievements that meet or exceed the Advanced Benchmark have the highest probability, while students located at lower benchmarks are less likely to give a correct response.

The next step in the scale anchoring method involves applying criteria that consider performance at adjacent benchmarks. This step identifies items that students at each International Benchmark are likely to answer correctly and differentiates achievement between consecutive benchmarks. These criteria help ensure that the performance descriptions at each benchmark reflect progressively higher levels of accomplishment by students reaching each successive benchmark. The LaNA 2023 Linking Study uses the same criteria as TIMSS and PIRLS.

For multiple-choice items, a probability of 0.65 (i.e., 65% of expected correct answers) is used as the criterion for anchoring at each benchmark. In addition, a criterion requiring less than 50% of expected correct answers is applied such that there is a clear distinction that students at lower benchmarks are more likely to have answered the item incorrectly than correctly. A

somewhat less strict criterion was used for the constructed-response items because students are less likely to guess when responding. For constructed-response items, the criterion of a probability of 0.50 of answering correctly (50% expected correct answers) was used without any discrimination criterion for the next lower benchmark.

The classification of items that "almost anchor" is used for multiple-choice items that meet slightly less stringent criteria for the IRT-based percent correct estimates. A multiple-choice item is considered "too difficult" to anchor if those at the Advanced Benchmark ability have less than an expected 60% correct responses. A constructed-response item is considered "too difficult" to anchor with less than 50% of students expected to answer correctly at the Advanced Benchmark. All items are used in the anchoring process to provide information about content domains and cognitive processes in mathematics and different skills in reading that might not otherwise be represented by those items that stringently meet the anchoring criteria.

Writing the International Benchmark Descriptions

Based on the analysis and resulting classification, the items were organized according to the benchmark at which they anchor. Each item was summarized to describe the kind of knowledge, skill, or cognitive process demonstrated by students who responded correctly to the item. For linking items, the item descriptions from TIMSS and PIRLS were used. For items newly developed for LaNA, the TIMSS & PIRLS International Study Center drafted the item descriptions. The new item descriptions were reviewed by a mathematics expert from the TIMSS 2023 Science and Mathematics Item Review Committee (SMIRC) and a reading expert from the PIRLS 2026 Reading Development Group (RDG). Then, the TIMSS & PIRLS International Study Center and the mathematics and reading experts wrote the new LaNA 2023 Basic International Benchmark summaries for mathematics and reading based on the confirmed item descriptions.

This process resulted in distinct descriptions of student achievement at the LaNA 2023 Basic International Benchmarks, yielding a content-referenced interpretation of the achievement results.

LaNA 2023 Linking Study Scale Anchoring

The TIMSS 2023 scale anchoring analysis was conducted using the item parameters for linking items coming from the TIMSS 2019 and PIRLS 2021 operational IRT calibrations, the item parameters for new items estimated as part of the LaNA 2023 Linking Study achievement scaling implementation (see Appendix D), and the expected correct responses from the fifth- and sixth-grade students who participated in the LaNA 2023 Linking Study. Exhibit E.2 presents the number of mathematics items anchoring at each International Benchmark. Exhibit E.3 presents the number of reading items anchoring at each International Benchmark.

Exhibit E.2: Number of Mathematics Items Anchoring and Almost Anchoring at Each International Benchmark for Mathematics

Block Type	Basic (325)	Low (400)	Intermediate (475)	High (550)	Advanced (625)	Above Advanced
LaNA	36	18	19	3	1	2
Linking	5	14	17	17	1	0
Mathematics Total	**41**	**32**	**36**	**20**	**2**	**2**

Exhibit E.3: Number of Reading Items Anchoring and Almost Anchoring at Each International Benchmark for Reading

Block Type	Basic (325)	Low (400)	Intermediate (475)	High (550)	Advanced (625)	Above Advanced
LaNA	32	17	2	2	0	0
Linking	15	26	23	11	1	0
Reading Total	**47**	**43**	**25**	**13**	**1**	**0**

The scale anchoring results show that most of the new LaNA items anchor at the new Basic benchmarks for mathematics and reading. Some new LaNA mathematics items anchor at the Low and Intermediate benchmarks, and only very few anchor at the High and Advanced benchmarks. Some of the new LaNA reading items anchor at the low benchmark, only very few reading items anchor at the Intermediate and High benchmarks, and no new items anchor at the Advanced benchmark.

The benchmark descriptions for the new LaNA 2023 Basic International Benchmarks as part of the scale anchoring were written together with a mathematics expert from the TIMSS 2023 SMIRC and one reading expert from the PIRLS 2021 RDG in September 2024.[1] In preparation for this, staff at the TIMSS & PIRLS International Study Center created detailed documentation for each item that included a draft item description, the framework classification, and the answer key or scoring guide, along with the scale anchoring analysis results and international average percent correct.

The item-by-item documentation was grouped by International Benchmark, then by content domain and topic area for mathematics and by reading purpose for reading, and finally by anchoring criteria for both mathematics and reading: items that anchored, followed by items that almost anchored, then by items that met only the 60–65% criteria. Item descriptions provided

1 The role of the SMIRC and RDG are detailed in Chapter 1 of the *TIMSS 2023 Technical Report: Methods and Procedures* and Chapter 1 of the *Methods and Procedures: PIRLS 2021 Technical Report*.

a short summary of the student competencies demonstrated by a correct (or partially correct) response to each item.

The SMIRC and RDG members performed two major tasks: 1) reviewed each item to finalize the description of the student competencies demonstrated by a correct (or partially correct) response, and 2) reviewed and edited the descriptions of the basic benchmarks for mathematics and reading drafted by the TIMSS & PIRLS International Study Center that summarized the proficiency demonstrated by students reaching each of the benchmarks. The benchmark descriptions are presented in Exhibits 1.1.3a and 1.1.3b for mathematics and Exhibits 1.2.3a and 1.2.3b for reading (see the report's main sections).

Citation

Khorramdel, L., Aldrich, C. E. A., & Kennedy, A. (2025). Appendix E: Scale anchoring and description of new basic LaNA benchmarks. In M. von Davier, L. Khorramdel, K. Reynolds, C. E. A. Aldrich, A. Bookbinder, A. Kennedy, & E. Gonzalez, *LaNA 2023 Linking Study Results* (pp. E.1-E.6). Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.bk4721

References

Mullis, I. V. S., & Fishbein, B. (2020). Using scale anchoring to interpret the TIMSS 2019 achievement scales. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 Technical Report* (pp. 15.1–15.60). Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/methods/chapter-15.html

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science.* Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/international-results/

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

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

Appendix F: Organizations and Individuals Responsible for LaNA 2023 Linking Study

Introduction

The LaNA 2023 Linking Study was a collaborative effort involving individuals around the world. This appendix acknowledges the individuals and organizations for their contributions. Given that the work leading up to the administration of LaNA spanned nearly ten years and involved so many people and organizations, this list may not include all who contributed. Any omission is inadvertent.

This report also acknowledges the students, teachers, and school principals who contributed their time and effort to the study. It would not be possible without their time, attention, and best efforts.

Management and Coordination

LaNA was a major undertaking to extend the work of TIMSS and PIRLS to be a useful measure for even more students globally. IEA continues its significant contributions in international large-scale assessment from regular cycles of studies of TIMSS (assessed every four years) and PIRLS (every five years).

LaNA was a close collaboration between the IEA, headed by Executive Director Dr. Dirk Hastedt, and the TIMSS & PIRLS International Study Center, headed by Executive Director Dr. Matthias von Davier (and formerly by Ina V.S. Mullis, through 2023, and by Michael O. Martin, through 2022). The TIMSS & PIRLS International Study Center is located in Boston College's Lynch School of Education and Human Development; it is responsible for the overall direction and management of TIMSS and PIRLS and led the efforts for LaNA in assessment design, instrument development, and adaptation, human scoring training, psychometric analysis of the data, and reporting. IEA Hamburg was responsible for the overall project management of LaNA and its implementation in the participating countries; it worked closely with participating countries to organize sampling and data collection operations and to initially check all data for accuracy and consistency within and across countries. IEA Amsterdam was responsible for managing country participation and overseeing the verification process of all translations produced by the participating countries.

To work with the international team and coordinate within-country activities, each participating country designates an individual to be the National Research Coordinator (NRC). The NRCs had the challenge of implementing LaNA in their countries following the guidelines and procedures of the LaNA 2023 Linking Study, which are based on TIMSS and PIRLS quality

assurance guidelines and procedures. The quality of the LaNA 2023 Linking Study data depends on the work of the NRCs and their colleagues in carrying out the complex sampling, data collection, and scoring tasks. This report commends the NRCs who performed their many tasks with dedication, competence, energy, and goodwill and demonstrated continued commitment to the project.

Funding

Funding for LaNA was provided by the participating countries at the national level, with additional support from the World Bank and from IEA. Boston College also is gratefully acknowledged for its generous financial support and stimulating educational environment.

Pilot Administrations

The LaNA 2023 Linking Study is based on the foundational data gathered as part of three pilot rounds beginning in 2016. The TIMSS & PIRLS International Study Center and IEA would like to acknowledge the participation of Haiti (2016 and 2020), Punjab, Pakistan (2017), Serbia, North Macedonia, and Nigeria in those highly informative administrations.

TIMSS & PIRLS International Study Center at Boston College

Matthias von Davier, Executive Director

Ina V.S. Mullis, Executive Director (through 2023)

Michael O. Martin, Executive Director (through 2022)

Pierre Foy, Senior Director, Sampling, Psychometrics, and Data Analysis (through 2023)

Eugenio Gonzalez, Senior Research Director, Statistics, Psychometrics, and Assessment Innovations

Ieva Johansone, Senior Research Director, Digital Systems, Operations and Quality Verification

Ann Kennedy, Senior Research Director, Project Management

Paul T. Connolly, Director, Graphic Design and Publications

Lale Khorramdel, Research Director, Digital Assessment, Development and Special Projects

Bethany Fishbein, Associate Research Director, Psychometrics and Data Analysis

Katherine Reynolds, Associate Research Director, Context Questionnaires and Reporting

Ummugul Bezirhan, Assistant Research Director, Advanced Psychometrics and Machine Learning Applications

Charlotte E.A. Aldrich, Assistant Research Director, TIMSS Mathematics and Special Projects

Allison Bookbinder, Assistant Research Director, TIMSS Science and Special Projects

Erin Wry, Senior Research Specialist, PIRLS Coordinator

Susan Flicop, Research Specialist, Operations and Quality Control
Liqun Yin, Senior Research Psychometrician
Jessie Bristol, Research Specialist, Measurement and Data Analysis
Marcie Bligh, Manager, Events and Administration
Joan King, Manager of Finance
Alicia Bouchard, Administrative Assistant
Ruthanne Ryan, Senior Graphic Designer
Mario Pita, Lead Graphic Designer

IEA

Dirk Hastedt, Executive Director

IEA Amsterdam

Andrea Netten, Director
Paulína Koršňáková, Senior Research and Liaison Adviser
Jan-Peter Broek, Financial Director
Isabelle Gémin, Senior Financial Officer
David Ebbs, Senior Research Officer
Lauren Musu, Head of TIQ
Jan-Philipp Wagner, Research Officer (through 2023)
Philippa Elliott, Publications Manager
Katie Zuber, Head of Communications

IEA Hamburg

Juliane Hencke, Director
Heiko Sibberns, Senior Research Advisor
Eva Feron, Head of International Studies Unit
Oliver Neuschmidt, Senior Research Analyst, Lead of Capacity Building Team
Clara Beyer, Research Analyst, International Data Manager
Sebastian Meyer, Research Analyst, International Data Manager
Chen-Ho Lee, Research Analyst
Mark Cockle, Research Analyst, International Data Manager
Sabine Meinck, Senior Research Analyst, Head of Research, Analysis and Sampling
Duygu Savaşcı-Smith, Research Analyst, Sampling Unit
Sabine Tieck, Research Analyst, Team Lead Sampling

Rodrigo Leyton, Research Analyst, Sampling Unit
Meng Xue, Head of Software Unit
Maike Junod, Software Developer
Delnaz Mohebi, Software Developer
Svetoslav Velkov, Software Tester
Wolfram Jarchow, Research Analyst

Sampling Referee

Jill DeMatteis, Vice President, Westat, Inc.

International Benchmark Contributors

Mathematics

Ray Philpot, Australian Council for Educational Research

Reading

Julian Fraillon, IEA

LaNA 2023 Linking Study National Research Coordinators

Burkina Faso

Étienne Ouedraogo

Egypt

Abd Al Kareem Ahmed Badran

Nigeria

Julius Adedoja

Palestinian National Authority

Mohammed O. Matar Mustafa

Pakistan

Kanwal Ejaz

Senegal

Alioune Badara Diop

BOSTON
COLLEGE
timssandpirls.bc.edu

IEA TIMSS & PIRLS
BOSTON COLLEGE