Assessment Framework and Instrument Development

Ina V.S. Mullis
Kathleen Trong Drucker
Corinna Preuscho
Alka Arora
Gabrielle M. Stanco

Unique Characteristics of the 2011 TIMSS and PIRLS Assessments

is description of the methods and procedures used to develop the assessment instruments for TIMSS and PIRLS developed in 2011 for each program.

Although the general approach to instrument development is similar for TIMSS and PIRLS, and similar from assessment cycle to assessment cycle, there also are di erences. Each assessment cycle tends to have some unique characteristics that influence the instrument development approach, and 2011 was no exception. Besides providing additional measures on trend lines monitoring changes in educational achievement, 2011 also was remarkable for

year, an event that would occur only every twenty years if the programs remain on the same cycles.

It was the inaugural year of prePIRLS. Countries whose students are not yet prepared to take PIRLS can still participate in this important international project by administering a parallel but less di cult reading assessment.

TIMSS and PIRLS in 2011

The alignment of TIMSS and PIRLS in 2011 provided countries a unique opportunity at the fourth grade to administer one comprehensive assessment of three essential subjects—reading, mathematics, and science. Countries participating in both programs received data about the relative strengths and weaknesses of their fourth grade students in reading, mathematics, and science. However, instrument development—assessment items and an array of background questionnaires—for the two assessments on the same schedule was challenging. Although most of the test item development activities were separate, it was necessary to coordinate background questionnaire development across TIMSS and PIRLS.

The opportunity to compare educational effectiveness across reading, mathematics, and science provided a strong impetus to update the contextual frameworks underlying the background data collection and focus attention on each and every questionnaire item.

As explained later in this publication, 2011 questionnaire development was guided through joint meetings of committees of TIMSS and PIRLS National Research Coordinators (NRCs). Also, a modular approach was adopted for the questionnaires to reduce response burden and facilitate data collection.

The prePIRLS Assessment of Reading Comprehension

For a variety of reasons, there are countries where most children in the fourth grade are still developing fundamental reading skills. So IEA o ers several options for matching the PIRLS assessment to the country's educational development. For some countries, participation in PIRLS at the h or sixth

TIM

Typically, each assessment cycle includes the following questionnaires:

The *Student Questionnaire* asks students about their home and school experiences in learning mathematics and science (TIMSS) and how to read (PIRLS).

The *PIRLS Learning to Read Survey (Home Questionnaire)* asks students' parents or guardians about home resources for fostering literacy and about their reading habits, highest level of education, and employment

Subsequently the test and questionnaire instruments are translated by participating countries into their languages of instruction with the goal of creating high quality translations that are appropriately adapted for the national context and at the same time are internationally comparable. Therefore, a signi cant portion of the development and review e ort by National Research Coordinators is dedicated to ensuring that the passages, test items, and questionnaires can be translated accurately.

The Item Development Process

The TIMSS & PIRLS International Study Center at Boston College uses a collaborative process to develop the new items needed for the mathematics,

TIMSS and PIRLS Assessment Frameworks

e ects. e remaining two PIRLS blocks (one literary and one informational) are presented in a magazine format in the *PIRLS Reader.* e *PIRLS Reader* has been in color since 2001, and, beginning in 2011, the entire assessment is moving to color to better mirror students' reading experiences.

Updating the Frameworks

Updating the frameworks is central to planning for each assessment cycle. For example, recommendations for updating content and cognitive domains can involve modifying content areas and their weightings (but no more than 5 percent); adding, deleting, or modifying topics within content areas to keep current with research ndings and ensure that the number of topics re ects the content area weighting; rewriting to improve clarity for item writers; and perhaps combining some topic areas to reduce redundancy. (Please click to view the **IM 2011** or **PIRL / PIRL 2011** **D** **A** .)

Recommendations about updating the contextual areas for teaching and learning that should be emphasized in collecting background data typically

2011, some PIRLS countries expressed a desire to expand PIRLS to include a component assessing web-based reading. Considerable development work was accomplished on the web-based reading initiative, with two web-based texts and questions undergoing development and review. However, not enough countries were able to secure funding for the project to continue. (Please click to view a description of the **PIRL 2011** .)

Identifying Reading Passages for PIRLS

Readers make meaning from a text in a variety of ways, depending not only on the purpose for reading but also on the di culty of the text and the readers' prior knowledge. us, identifying appropriate passages for the PIRLS and prePIRLS

Writing, Reviewing Field Test Items and Scoring Guides

e TIMSS & PIRLS International Study Center uses a collaborative process involving the participating countries to develop test items and scoring guides

e Coordinators and TIMSS & PIRLS International Study Center sta use the item writing guidelines in providing training to the teams on item writing

items.[1] Generally, one or two extra items were developed for each scale to allow for attrition a er eld testing.

Almost 40 background scales were developed for the TIMSS and PIRLS 2011 eld tests. At the same time, however, many questions from the 2006

a larger sample of 180 schools is selected and a systematic sample of 30 schools is selected from the 180 schools.

While countries are translating and nalizing their eld test instruments for data collection, the TIMSS and PIRLS International Study Center prepares materials for training countries in how to evaluate students' answers to constructed-response items in the eld test.

It is important to prepare scoring training materials for the newly developed constrntd-rete0120 595.776n tps5120 5(o)12.1(n)8a- aoiosecttems01(l)1the rewly1

Finalizing the Assessment Instruments

e TIMSS & PIRLS International Study Center reviews and analyzes the eld test data. Data almanacs containing summary item statistics are prepared for each test item and questionnaire item. e data almanac for a test question contains, row by row for each country: the sample size, the item di culty and discrimination, the percentage of students answering each option (multiple-choice) or in each score category (constructed-response), and the point-biserial correlation for each multiple-choice option or constructed-response category. For constructed-response items, the degree of scoring agreement also is provided.

The data almanacs for questionnaire items contain the percentage of students responding to each question option, together with the corresponding average student achievement in mathematics, science, or reading, respectively. In addition, the background questionnaire item sets (scales) are evaluated for unidimensionality, internal consistency, and relationship with achievement.

questionnaire committee (the QIRC or QDG). ese expert committees review each questionnaire item for clarity, examine the data to make sure the options are providing useful information, and make suggestions for re nements in preparation for data collection.

e expert committees' recommendations are implemented by sta , and the penultimate assessment blocks and questionnaires are sent to the NRCs for review. NRCs have the opportunity to review the recommended materials in light of the eld test results and within the security of their own countries. Each country also checks any unusual national results that might be an indication of translation errors and corrects the translation as necessary or recommends revisions to accommodate translation. Finally, there is an NRC meeting to reviee102.7(e1 c)-1.9(r)-6.2(n)7v(et819(t 1(a)12.9(a)-5.9(y)722(ts t)-68(o)11.1.2(e)1.85(s)1.)11.1.2(e)11(o

even in English-speaking countries) and production of the materials for printing. At the same time, countries should be making nal arrangements for data collection, including the host of activities necessary to obtain school participation, implement test administration, and score the responses to the tests and questionnaires.

A er each cycle approximately 40 percent of the assessment items are released to the public to illustrate the content of the assessments and for educational purposes, such as further research into students' strengths and weaknesses in reading or as part of instructional materials about topics in mathematics or science. e remaining assessment items are kept secure to be readministered in subsequent cycles as the basis for measuring trends.

The release plans provide for each assessment to incude items from several cycles—three for TIMSS and four for PIRLS. In TIMSS, approximately 40 percent of the items are newly developed for each cycle, 40 percent from the previous cycle, and 20 percent from two cycles before. For PIRLS also, approximately 40 percent of the items are newly developed for each cycle. However, beginning in 2016, PIRLS will have 20 percent of its items from the previous cycle, 20 percent from two cycles before, and 20 percent from three cycles before. us, for each upcoming cycle, it is necessary to replace a speci c portion of the achievement items. (Please click to learn more about the **IM 2011** or **PIRL / PIRL 2011 L R P** , including when each assessment block was rst used and when it will be released to the public.) All of the TIMSS and PIRLS procedures following instrument development, overall and for 2011 in particular, are described in subsequent sections of this publication.