CHAPTER 6

Survey Operations Procedures in PIRLS 2016

Ieva Johansone

Overview

As data-based indicators of countries' student achievement profiles and learning contexts, PIRLS assessments are crucially dependent on the quality of the data collected by each participating country and benchmarking entity. Whereas the development of the assessments is an intensely collaborative process involving all of the partners in the enterprise, the process of administering the assessments and collecting the data is uniquely the responsibility of each individual country or benchmarking participant.

To ensure the consistency and uniformity of approach necessary for high-quality, internationally comparable data, all participants are expected to follow a set of standardized operations procedures. These procedures have been developed through a partnership involving the TIMSS & PIRLS International Study Center, IEA Amsterdam, IEA Hamburg, Statistics Canada, and National Research Coordinators (NRCs) from participating countries. With each new assessment cycle, the operations procedures are updated to enhance efficiency and accuracy and reduce burden, making use of developments in information technology to automate routine activities wherever possible. Additionally, with the ePIRLS extension being administered for the first time in 2016, developing operations and procedures for this innovative assessment of online reading and integrating the workflow into the existing PIRLS operations was a significant undertaking.

In each country or benchmarking entity, the National Research Coordinator was responsible for the implementation of PIRLS 2016. Internationally, National Research Coordinators provided the country's perspective in all international discussions, represented the country at international meetings, and were the responsible contact persons for all project activities. Locally, National Research Coordinators were responsible for implementing all the internationally agreed-upon procedures and facilitating all of the national decisions regarding PIRLS, including any adaptations for the national context.

The daily tasks of the National Research Coordinators varied over the course of the PIRLS 2016 cycle. In the initial phases, National Research Coordinators participated in the PIRLS 2016 assessment framework and assessment development process (see Chapter 1) and collaborated with Statistics Canada and IEA Hamburg to develop a plan to implement the PIRLS 2016 sampling design within the country or benchmarking entity (see Chapter 5).

Following the development of the draft reading passages, achievement items, and context questionnaires, all countries conducted a full-scale field test of all instruments and operational procedures in March through April 2015 in preparation for the PIRLS 2016 data collection, which took place in October through December 2015 in Southern Hemisphere countries, and in March through May 2016 in Northern Hemisphere countries. The field test allowed the National Research Coordinators and their staff to become acquainted with the operational activities, and the feedback they provided was used to improve the procedures for the data collection. As expected, the field test resulted in some enhancements to survey operations procedures, especially for ePIRLS, which was new for the 2016 assessment cycle, and contributed to ensuring the successful execution of PIRLS 2016.

As part of ongoing efforts to improve operations, the National Research Coordinators were asked to complete a Survey Activities Questionnaire (SAQ), which sought feedback on all aspects of their experience conducting PIRLS 2016. The feedback solicited in the SAQ included an evaluation of the quality of the assessment materials and the effectiveness of the operations procedures and documentation. The results of the PIRLS 2016 Survey Activities Questionnaire are presented in the final section of this chapter.

PIRLS 2016 Survey Operations Units, Manuals, and Software

To support the National Research Coordinators in conducting the PIRLS 2016 assessments, the TIMSS & PIRLS International Study Center provided step-by-step documentation of all operational activities. Organized into a series of units, the *PIRLS 2016 Survey Operations Procedures* were made available at critical junctures of the project to ensure that National Research Coordinators had all the tools and information necessary to discharge their responsibilities. ePIRLS specific supplements to the PIRLS units were provided when necessary. Also, the *Procedures* units were accompanied by a series of manuals for use by School Coordinators and Test Administrators that National Research Coordinators could translate and adapt to their local situations. The TIMSS & PIRLS International Study Center and IEA Hamburg also provided National Research Coordinators and their staff with intensive training in constructed response item scoring and data management.

Consistent with the goal of automating and streamlining procedures wherever possible, IEA Hamburg provided National Research Coordinators with a range of custom-built software products to support activities, including sampling and tracking classes and students, administering

school, teacher, and home questionnaires, documenting scoring reliability, and creating and checking data files. IEA Hamburg was also responsible for ePIRLS Software development. The ePIRLS system was hosted on the IEA Hamburg server and consisted of a number of software modules enabling the translation and verification processes, assessment administration to students, monitoring of the ePIRLS data upload, and scoring of the ePIRLS constructed response items.

The *Survey Operations Procedures* units were crucial resources for the National Research Coordinators as the units described in detail the tasks the NRCs were responsible for conducting. In the event that some of these tasks were contracted out to other people or organizations, the units ensured that the NRCs had sufficient knowledge of these matters to supervise the activities of the people who helped conduct the assessment(s) in their countries.

The following units, manuals, and software systems were provided for administering PIRLS and ePIRLS 2016:

- *PIRLS 2016 Survey Operations Procedures Unit 1: Sampling Schools and Obtaining their Cooperation*
- *PIRLS 2016 Survey Operations Procedures Unit 2: Preparing for and Conducting the PIRLS 2016 Field Test*. Unit 2 consisted of the following sections: Preparing Achievement Booklets and Background Questionnaires with an ePIRLS supplement on preparing the ePIRLS assessment tasks, Sampling Classes and Field Test Administration, Scoring the Constructed Response Items with an ePIRLS supplement on scoring the ePIRLS constructed response items online, and Creating the Field Test Databases. Unit 2 was accompanied by field test versions of the School Coordinator Manual, "Preparing Computers for ePIRLS" instructions, Test Administrator Manuals for PIRLS and ePIRLS, and a National Quality Control Monitor Manual. Eight software systems/modules (WinW3S, ePIRLS System Check Program, ePIRLS Online Translation System, ePIRLS Software, ePIRLS Online Data Monitor, ePIRLS Online Scoring System, IEA DME, and IEA OSS—described below) were provided for the field test.
- *PIRLS 2016 Survey Operations Procedures Unit 3: Contacting Schools and Sampling Classes for the Data Collection*. Unit 3 was accompanied by the School Coordinator Manual and the Windows® Within-school Sampling Software (WinW3S) and its manual. The WinW3S software enabled PIRLS 2016 participants to randomly select classes in each sampled school and document in detail the class selection process. The software also was used to track school, teacher, student, and student-teacher linkage information; prepare the survey tracking forms (described later in this chapter); and assign assessment instruments to students, including printing labels for the assessment instruments.

- *PIRLS 2016 Survey Operations Procedures Unit 4: Preparing Achievement Booklets and Context Questionnaires*. Unit 4 was accompanied by the IEA Online SurveySystem (OSS) and its manual. The IEA Online SurveySystem supported the online administration of the school, teacher, and home (Learning to Read Survey) questionnaires.
- *ePIRLS Supplement to the PIRLS 2016 Survey Operations Procedures Unit 4: Preparing the ePIRLS Assessment Tasks*. This ePIRLS supplement was accompanied by the ePIRLS Online Translation System enabling National Research Coordinators to connect to the ePIRLS server at IEA Hamburg to translate the ePIRLS assessment tasks into their language(s) of instruction. The translated tasks were then available online for translation and layout verification by IEA Hamburg and the TIMSS & PIRLS International Study Center (see Chapter 7).
- *PIRLS 2016 Survey Operations Procedures Unit 5: Conducting the Data Collection*. Unit 5 was accompanied by the Test Administrator Manuals for PIRLS and ePIRLS, the National Quality Control Monitor Manual, and the International Quality Control Monitor Manual.
- "Preparing Computers for ePIRLS" instructions and the ePIRLS System Check Program. The instructions and software provided the necessary information and tools for countries to test computers for ePIRLS compatibility and prepare the ePIRLS compatible computers for ePIRLS administration.
- ePIRLS Software for administering the ePIRLS assessment to students. ePIRLS Software was provided for each participating country and benchmarking entity individually, containing each participant's national/translated version of the ePIRLS assessment tasks.
- *PIRLS 2016 Survey Operations Procedures Unit 6: Scoring the Constructed Response Items*. Unit 6 was accompanied by the PIRLS 2016 Scoring Guides, the IEA Coding Expert Software, the Trend Reliability Scoring Manual, and the Cross-country Reliability Scoring Manual. The IEA Coding Expert Software was used to facilitate the trend and cross-country reliability scoring tasks.
- *ePIRLS Supplement to the PIRLS 2016 Survey Operations Procedures Unit 6: Scoring the Constructed Response Items*. This ePIRLS supplement was provided with the ePIRLS Online Data Monitor and ePIRLS Online Scoring System software facilitating monitoring of the ePIRLS data upload to the IEA Hamburg ePIRLS server and scoring the ePIRLS constructed response items.
- *PIRLS 2016 Survey Operations Procedures Unit 7: Creating the Databases*. Unit 7 was accompanied by the IEA Data Management Expert (DME) software, its manual, and codebooks that specified information on the IEA DME data fields in each of the data files. The IEA DME software is used for data entry and data verification.

PIRLS 2016 Survey Tracking Forms

PIRLS uses a series of tracking forms to document class sampling procedures, assign assessment instruments, and track school, teacher, and student information, including the participation status of the respondents. The tracking forms also facilitate the data collection and data verification process. Four different tracking forms were used for PIRLS 2016:

- Class Listing Form: This form was completed for each sampled school, listing the eligible classes and providing details about the classes, such as the class stream (if applicable), the number of students, and the names of teachers.
- Student Listing Form: This form was completed for each class sampled, listing the names of the students, student birth dates, gender, and exclusion codes.
- Student Tracking Form: This form was created for each class assessed and was completed by the Test Administrators during test administration. Separate Student Tracking Forms were provided for PIRLS and ePIRLS. The Test Administrators used this form to verify the assignment of survey instruments to students and to indicate participation status, including the return status of the Learning to Read Surveys (home questionnaires).
- Teacher Tracking Form: This form was completed for each sampled school to indicate the completion of the teacher questionnaires.

Operations for Data Collection

The following sections describe the major operational activities coordinated by the National Research Coordinators:

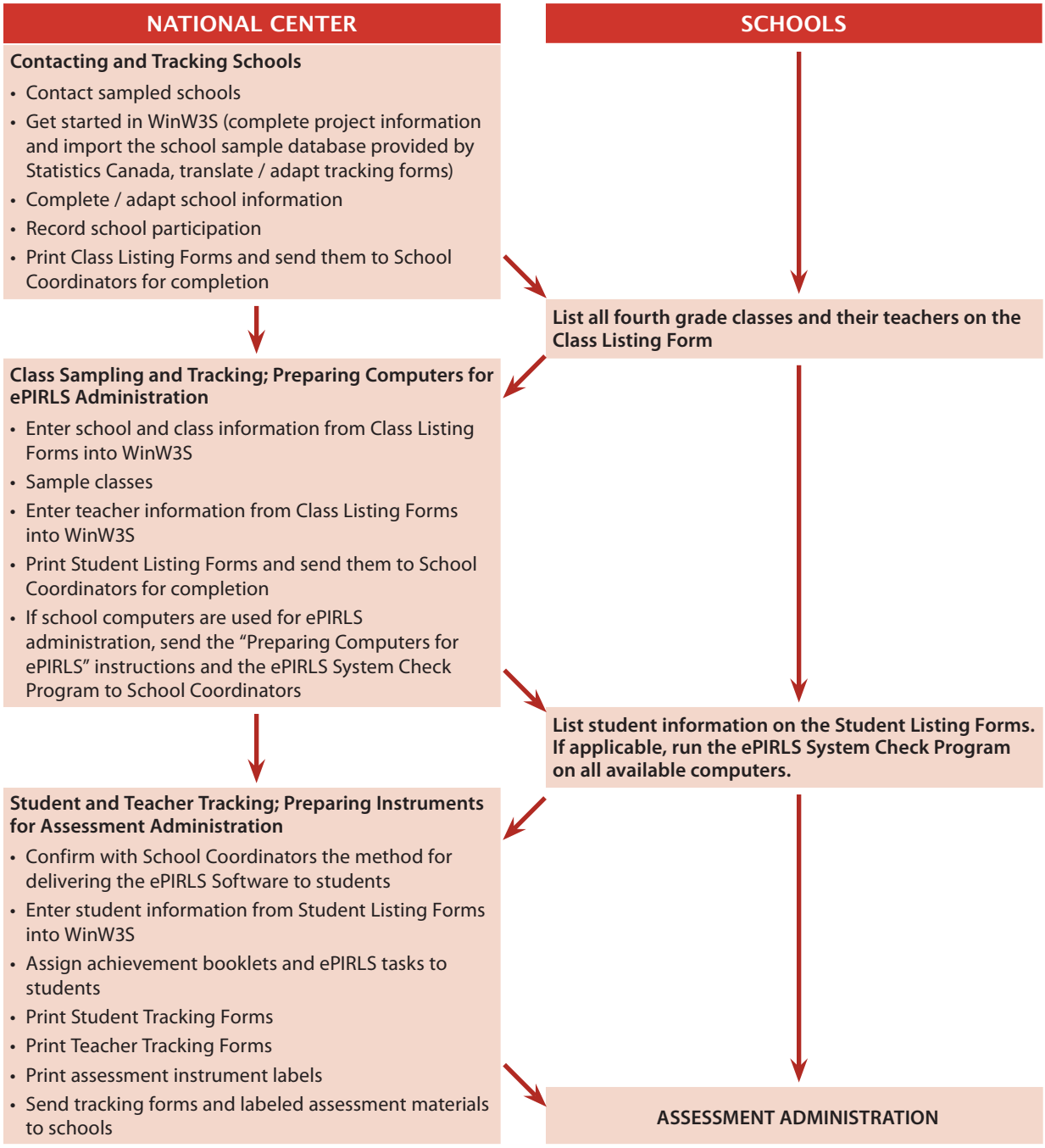
- Contacting schools and sampling classes
- Overseeing translation and preparing assessment instruments
- Managing the PIRLS 2016 assessment administration
- Scoring of the constructed response items
- Creating the PIRLS 2016 data files

Two other major PIRLS 2016 operational activities are described in separate chapters of the *Methods and Procedures in PIRLS 2016* publication—sampling schools (Chapter 3) and translation and layout verification of the assessment instruments (Chapter 7).

Contacting Schools and Sampling Classes

Exhibit 6.1 illustrates the major steps of working with schools to sample classes and prepare for the PIRLS 2016 assessment administration. Once the school samples were drawn, National Research Coordinators were tasked with contacting schools and encouraging them to take part in the assessment(s). Depending on the national context, this could involve obtaining support from national or regional educational authorities. *Survey Operations Procedures Unit 1* outlines suggestions on ways to encourage schools to participate in the assessment.

Exhibit 6.1: Diagram of the Sampling Procedures and Preparations for the Assessment Administration Implemented by National Centers and Schools



In cooperation with school principals, National Research Coordinators were responsible for identifying and training School Coordinators for all participating schools. A School Coordinator could be a teacher or guidance counselor in the school, or National Research Coordinators could appoint a member of the national center to fill this role. In some countries, a School Coordinator from the national center was responsible for several schools in an area. School Coordinators were provided with a School Coordinator Manual, describing their responsibilities. The School Coordinator Manual was prepared by the TIMSS & PIRLS International Study Center and translated/adapted by national center staff in each country.

The responsibilities of the School Coordinators included providing the national center with information on the school; coordinating the dates, times, and places for testing; identifying and training Test Administrators to administer the assessments; coordinating the completion of the tracking forms; distributing questionnaires; and when necessary obtaining parental permission. If school computers were used for ePIRLS administration, School Coordinators were provided with the "Preparing Computers for ePIRLS" instructions and the ePIRLS System Check Program in order to test the computers for ePIRLS compatibility and prepare the compatible computers for testing. School Coordinators also confirmed receipt of all assessment materials, oversaw the security of the assessment materials, and ensured the return of the assessment materials to the national center following assessment administration.

In addition, School Coordinators provided the national center with data on eligible classes in the schools. With this information, the national centers used WinW3S to sample classes within the schools. Because PIRLS samples intact classes, the School Coordinators checked that every student was listed in one and only one class. This was necessary to ensure that the sample of classes resulted in a representative sample of students, and every student at the target grade had a chance of being selected.

Overseeing Translation and Preparing Assessment Instruments

National Research Coordinators also were responsible for preparing the assessment instruments (achievement booklets, ePIRLS tasks, and context questionnaires) for their countries—a process that included overseeing the translation of the assessment instruments. The overarching goal of assessment instrument preparation is to create internationally comparable instruments that are appropriately adapted for the national context of each participating country.

Each student was assigned one achievement booklet. There are 16 PIRLS achievement booklets and 16 PIRLS Literacy achievement booklets. Each booklet contains two assessment blocks, each including a passage with a set of items. Even though each assessment block appeared in more than one booklet, from an operational perspective, each block needed to be translated only once. Countries used Adobe® InDesign® software to link the translated and adapted assessment blocks to the appropriate booklets. Automating this process through Adobe® InDesign® decreased the chances of human error in the production process.

Students participating in ePIRLS were assigned two of five ePIRLS assessment tasks. ePIRLS translations and/or adaptations were applied through the ePIRLS Online Translation System and then distributed and delivered to students via the ePIRLS Software.

As described in Chapter 1, ten new assessment blocks were developed for PIRLS and PIRLS Literacy 2016, with the new blocks replacing the ones released at the end of the previous assessment cycle. Also, five assessment tasks were developed for the new ePIRLS 2016 assessment. The new assessment blocks (PIRLS passages and ePIRLS tasks) tasks were all tried out through the field test to investigate the psychometric characteristics of the achievement items. The best assessment blocks were chosen and some edits were applied for the main data collection. Similarly, the context questionnaires were evaluated following the field test to gauge the validity and reliability of the various questionnaire scales.

All participating countries and benchmarking entities translated and/or adapted the newly developed assessment blocks into the test administration language and did the same for the questionnaires. Countries that did not participate in PIRLS/prePIRLS 2011 or PIRLS 2006 had to translate and/or adapt the assessment blocks used in previous assessments (trend blocks) into their language(s) in preparation for the 2016 assessment administration. Countries that had participated in PIRLS/prePIRLS 2011 and/or PIRLS 2006 were required to use the same translations they used in those cycles.

For both the field test and main data collection, the participating countries received the international version (English) of the achievement booklets and context questionnaires with all the necessary instrument production files, including fonts and graphics files. For ePIRLS, this was done via the ePIRLS Online Translation System. Instructions on how to use the materials to produce high-quality, standardized instruments, were included in the corresponding *Survey Operations Procedures* Unit.

Once translated and/or adapted, first for the field test and then again for the main data collection, the passage/tasks, items, and context questionnaires were submitted to IEA Amsterdam for translation verification. IEA worked with independent translators to evaluate each country's translations and, when deemed necessary, suggested changes to the text.

After the translation verification, National Research Coordinators applied the necessary changes, and copies of the instruments were submitted to the TIMSS & PIRLS International Study Center for layout verification and to review national adaptations. This review checked that each booklet, ePIRLS assessment task, and questionnaire conformed to the international format and that any adaptations made to the instruments did not unduly influence their international comparability.

Documenting National Adaptations

While preparing national assessment instruments, countries sometimes by necessity made adaptations to the international versions. All national adaptations to the international assessment instruments, other than direct translation, were documented. For the achievement booklets and context questionnaires, the National Adaptations Forms (NAFs) were used to capture this documentation. For ePIRLS, national adaptations were documented via the ePIRLS Online Translation System.

During the translation verification and layout review, the verifiers checked whether the national adaptations were likely to influence the ability to produce internationally comparable data for the items involved. Any questions raised were directed to the National Research Coordinator for consideration.

The documentation was completed and reviewed at various stages of preparing national assessment instruments. Version I of the forms and online documentation was completed during the internal translation and review process and sent along with the rest of the materials for international translation verification. After translation verification, the documentation (Version II) was updated in response to the translation verifier's comments, reflecting any changes resulting from the verification, and sent along with the national assessment instruments for layout verification. Following layout verification, the national instruments and documentation were finalized (Version III) and submitted to IEA and the TIMSS & PIRLS International Study Center.

Managing the Administration of the PIRLS 2016 Assessments

Printing, preparing, and distributing assessment materials to the participating schools required careful organization and planning on the part of the National Research Coordinators. The assessment materials were packaged and sent to the School Coordinators prior to testing, giving ample time for the School Coordinators to confirm the receipt and correctness of the materials. The School Questionnaire and Teacher Questionnaires were then distributed, and the other instruments were kept in a secure room until the testing date.

Each sampled class was assigned a Test Administrator(s) who followed procedures described in the PIRLS and/or ePIRLS Test Administrator Manual to administer the assessment and student questionnaire. Test Administrators were in most cases chosen and trained by School Coordinators, and in some cases, the School Coordinator doubled as the Test Administrator.

Test Administrators were responsible for distributing materials to the appropriate students, reading the instructions provided in the Test Administrator Manual to the students, and timing the sessions. WinW3S systematically assigned achievement booklets and ePIRLS assessment tasks and produced labels to facilitate the distribution of the assessment, and Test Administrators used the Student Tracking Form(s) and these labels to distribute the assessment instruments to the correct students and to document student participation. When a class had a participation rate

below 90 percent, it was the School Coordinator's responsibility to hold a makeup session for the absent students before returning all of the testing materials to the national center. Using the Test Administration Form, the Test Administrators documented the timing of the testing sessions and also solicited information about anything out of the ordinary that took place during assessment administration.

The PIRLS achievement booklets consisted of two sections with each containing one assessment block, and ePIRLS consisted of two parts with each containing one assessment task. To complete each part of the test, students were allowed 40 minutes, and the time was strictly enforced by Test Administrators. ePIRLS Software also automatically logged students out of the system once the 40 minutes had expired. There was a required break between the two parts of assessment administration. The break was not to exceed 30 minutes. Students who completed part 1 or part 2 of the assessment before the allotted time were not allowed to leave the testing room and were asked to review their answers or read quietly. Some Test Administrators provided activity sheets for these students.

Following the administration of the PIRLS assessment, students were provided 30 minutes to complete the student questionnaire with extra time provided to students who needed it. During administration of the student questionnaire, Test Administrators were permitted to read the questionnaire items aloud together with the students. Following the administration of the ePIRLS assessment, students also took a short computer-based questionnaire about their experiences and attitudes toward using a computer.

PIRLS, including the student questionnaire, was always administered before ePIRLS. ePIRLS was mostly administered via individual USB sticks on individual ePIRLS compatible computers. Sometimes, the server method was used via a Local Area Network (LAN), which entailed a single ePIRLS compatible computer being used as a local server and students using individual devices connected to the server computer. For ePIRLS, the Test Administrators and School Coordinators submitted the ePIRLS data after each testing session. Due to computer shortages, sometimes multiple ePIRLS testing sessions were needed for each class.

Linking Students to their Teachers and Classes

Exhibit 6.2 illustrates the hierarchical identification system codes that are used to link the data among schools, classes, students, and teachers. The school, class, and student IDs are strictly hierarchical, with classes nested within schools and students nested within classes.

Exhibit 6.2: Hierarchical Identification System Codes Used to Link Schools, Classes, Students, and Teachers

Participant	ID Components	ID Structure	Numeric Example
School	School	CCCC	0001
Class	School + Class within the school	CCCCKK	000101 000102
Student	School + Class within the school + Student within the class	CCCCKKSS	00010101 00010201
Teacher	School + Teacher within the school + Linkage number to the sampled class	CCCCTTLL	00010101 00010201

Each teacher is assigned a teacher identification number consisting of the four-digit school number followed by a two-digit teacher number. Since the same teacher could be teaching more than one class within a school, it is necessary to have a unique identification number for each teacher linked to a class. This is achieved by adding a two-digit link number to the six digits of the teacher identification number to create a unique eight-digit identification number.

Online Administration of the School, Teacher, and Home Questionnaires

Countries could choose to administer the school, teacher, and home questionnaires online. The benefits of administering the questionnaires online included saving money and time in printing, and improving the efficiency of questionnaire distribution, data entry, and data cleaning.

For the online administration of the questionnaires, IEA Hamburg provided its IEA Online SurveySystem Software that incorporates design, presentation, and monitoring components.

The design component, known as the Designer, supports the preparation of the online surveys, data management, and data output to IEA Hamburg. Through the IEA Online SurveySystem Designer component, national centers could tailor the online questionnaires to their national language. To facilitate translation and adaptation, the Designer concurrently stored the original English question text and the translations and/or national adaptations. It also stored the variable names and data validation rules. If a national center decided not to administer a particular international question or option, it could be disabled in the Designer and would not be administered during the online questionnaire administration. The Designer also included an integrated preview function to allow for a visual side-by-side comparison of the paper/PDF and online versions of the questionnaires, facilitating the layout verification process.

For the online presentation, the Web Component presents the questionnaires to the respondents. The navigation capabilities of the Web Component are designed to allow respondents to pick and choose their order of response. Buttons marked "next" and "previous" facilitated navigation between adjacent pages, so users could browse through the questionnaire in the same way that they flip through the pages of the paper questionnaire. A hyperlinked interactive "table

of contents" allowed the respondents to fluidly navigate to specific questions. Overall, these two functions permitted the respondents to answer questions in the order of their choosing, and skip questions just as they could do if they were answering the paper questionnaire. Also, the online questionnaires could be accessed through any standard Internet browser on all standard operating systems without the user needing any additional software.

Finally, the Web-based Monitor component allows for monitoring the survey responses in real time. Many national centers made extensive use of the Web-based Monitor to follow-up with non-respondents.

IEA Hamburg followed a stringent set of procedures to safeguard the confidentiality of the respondents and maintain the integrity of the data. Each respondent received a statement of confidentiality, and information on how to access the online questionnaire. For most countries, the online questionnaire administration was hosted on the IEA Hamburg customized high-performance server. This server allowed for the 24-hour availability of the questionnaires during the data collection period, and it also ensured backup and recovery provisions for the data.

Scoring the Constructed Response Items

Constructed response items represent a substantial portion of the PIRLS assessments, and because reliable and valid scoring of these items is critical to the assessment results, the TIMSS & PIRLS International Study Center provided explicit scoring guides and extensive training in their use. Also, the *Survey Operations Procedures* units specified a procedure for efficiently organizing and implementing the scoring activity. Scoring of the ePIRLS constructed response items was done online via the ePIRLS Online Scoring System, which incorporated the IEA standards and reliability procedures.

International scoring training sessions (one for the field test and two for the main data collection—one for Southern Hemisphere countries and another for Northern Hemisphere countries) were conducted where all National Research Coordinators (or country representatives appointed by the National Research Coordinators) were trained to score each of the constructed response items. At these training sessions, the scoring guide for each item was reviewed and applied to a sample set of example student responses that had already been scored. These example papers were chosen to represent a range of response types and to demonstrate the guides as clearly as possible. Following the example papers, the training participants applied the scoring guides to a different set of student responses that had not yet been scored. The scores to these practice papers were then shared with the group and any discrepancies were discussed.

Following the international scoring training, national centers trained their scoring staff on how to apply the scoring guides for the constructed response items. National Research Coordinators were encouraged to create additional example papers and practice papers from student responses collected in their country.

Documenting Scoring Reliability

Because reliable scoring of the constructed response items is essential for high quality data, it is important to document the reliability of the scoring process. A high degree of scorer agreement is evidence that scorers have applied the scoring guides in the same way. The procedure for scoring the PIRLS constructed response items provided for documenting scoring reliability within each country (within-country reliability scoring), over time (trend reliability scoring), and across countries (cross-country reliability scoring).

The method for establishing the reliability of the scoring within each country was for two independent scorers to score a random sample of 200 responses for each constructed response item. The degree of agreement between the scores assigned by the two scorers is a measure of the reliability of the scoring process. In collecting the within-country reliability data, it was vital that the scorers independently scored the items assigned to them, and each scorer did not have prior knowledge of the scores assigned by the other scorer. The within-country reliability scoring was integrated within the main scoring procedure and ongoing throughout the scoring process. The within-country reliability scoring procedure was implemented in both PIRLS and ePIRLS.

The purpose of the trend reliability scoring was to measure the reliability of the scoring from one assessment cycle to the next (i.e., from PIRLS 2011 to PIRLS 2016). The trend reliability scoring required scorers of PIRLS 2016 to score student responses collected in 2011. The scores from 2016 were then compared with the scores awarded in 2011. Trend reliability scoring was conducted using the IEA Coding Expert Software provided by IEA Hamburg.

Student responses included in the trend reliability scoring (150–200 responses per item) were actual student responses to 22 items from four of the PIRLS trend assessment blocks and/or 24 items from three of the PIRLS Literacy trend assessment blocks collected during the PIRLS/prePIRLS 2011 assessment administration in each country and benchmarking entity. These responses were scanned and provided for each participating country and benchmarking entity along with the IEA Coding Expert Software. All scorers who scored the trend assessment blocks in 2016 were required to participate in the trend reliability scoring. If all scorers were trained to score all trend items, the software divided the student responses equally among the scorers. If scorers were trained to score specific assessment blocks, National Research Coordinators were able to specify within the software which scorers would score particular blocks, and the software allocated the student responses accordingly. Similar to the within-country reliability scoring, the trend reliability scoring had to be integrated within the main scoring procedure.

Finally, cross-country reliability scoring gave an indication about how consistently the scoring guides were applied from one country to the next. The cross-country reliability scoring also was conducted using IEA Coding Expert Software. Student responses included in the cross-country reliability scoring (200 responses per item) were student responses to 22 items from four of the

PIRLS assessment blocks (the same passages and items were used for the trend scoring reliability study) that were collected from the English-speaking countries during the PIRLS 2011 assessment administration. All scorers who could score student responses written in English were required to participate in the cross-country reliability scoring, and the student responses were equally divided among the participating scorers in each country. The scoring exercise was completed immediately after all other scoring activities.

Creating the PIRLS 2016 Databases

The data entry process took place from March to May 2015 for the field test, from December 2015 to March 2016 following data collection in the Southern Hemisphere, and June to September 2016 following data collection in the Northern Hemisphere. The procedure for creating the PIRLS 2016 databases included entering sampling and assessment administration information into the WinW3S database and adding responses from the context questionnaires and achievement booklets using the IEA Data Management Expert (DME) software. IEA Hamburg provided the DME software to accommodate keyboard data entry from the paper instruments. The DME software also offers data and file management capabilities, a convenient checking and editing mechanism, interactive error detection, and quality control procedures.

Along with the DME software, IEA Hamburg provided international codebooks describing all variables and their characteristics, thus ensuring that the data files met the internationally defined rules and standards for data entry. The files within the DME database for entering the PIRLS 2016 data were based on these codebooks. However, the codebooks had to match exactly the national assessment instruments so that the answers of the respondents could be entered properly. Therefore, any adaptations to the international instruments also required adaptations to the international codebooks. The adapted national codebooks then were used to create the PIRLS 2016 data files in each country, with the responses to the context questionnaires, achievement booklets, and Reliability Scoring Sheets keyed into the DME database.

Quality control throughout the data entry process was essential to maintain accurate data. Therefore, National Research Coordinators were responsible for performing periodic reliability checks during data entry and for applying a series of data verification checks provided by both WinW3S and DME software prior to submitting the databases to IEA Hamburg. To ensure the reliability of the data entry process, the data entry staff was required to double enter at least 5 percent of each instrument type. An error rate of 1 percent or less was acceptable for the background files. An error rate of 0.1 percent or less was required for the student achievement files and the reliability scoring files. If the required agreement was not reached, retraining of the key punchers was required.

The ePIRLS assessment data were captured automatically by submitting them to the IEA Hamburg ePIRLS server immediately after the assessment administration. Countries were provided with the ePIRLS Online Data Monitor to monitor the data submission. The ePIRLS constructed

response scoring took place directly in the online database and thus did not require any manual data entry. For the PIRLS 2016 teacher, school, and home questionnaires administered online through the Online SurveySystem (OSS) via the IEA Hamburg server, the data were directly accessible by IEA Hamburg and no further data entry was required.

Both WinW3S and DME offer a data verification module identifying a range of problems, such as inconsistencies of identification codes, inconsistencies between participation status information and achievement and/or background data availability, and out-of-range or otherwise invalid codes. The data quality control procedures also verify the integrity of the linkage between the students, teachers, and schools entered into the DME database and tracking of information for those specified in WinW3S. For data captured online (i.e., ePIRLS and context questionnaires administered online), it was possible to export data availability information and apply data verification to check for inconsistencies via the WinW3S and DME data verification modules.

When all data files had passed the quality control checks, they were submitted to IEA Hamburg, along with data documentation, for further checking and processing. For information on data processing at IEA Hamburg, please refer to Chapter 9 of this publication.

PIRLS 2016 Survey Activities Questionnaire

The Survey Activities Questionnaire was designed to elicit information about National Research Coordinators' experiences in preparing for and conducting the PIRLS 2016 data collection. The questionnaire was composed of six sections and focused on the following:

- Sampling schools and classes
- Preparing assessment instruments
- Administering the assessment(s)
- Implementing the National Quality Control Program
- Preparing for and scoring the constructed response items
- Creating the databases

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

The *PIRLS 2016 Survey Activities Questionnaire* was administered online via the IEA's Online SurveySystem and was completed by a total of 52 NRCs, with 15 NRCs also providing feedback on ePIRLS administration. The following sections summarize information gathered from the Survey Activities Questionnaire.

Sampling Schools and Classes

The first section of the Survey Activities Questionnaire asked National Research Coordinators about the *Survey Operations Procedures* Units for sampling both schools and classes within the sampled schools. As shown in Exhibit 6.3, all but one of the National Research Coordinators considered that *Survey Operations Procedures* Units 1 and 3 to be clear and sufficient. Two countries reported deviating from the basic PIRLS sampling design. Their reasons for these modifications to the sampling procedures included allowing for census participation, oversampling certain regions, and specific requirements to coordinate their PIRLS 2016 sample with the TIMSS 2015 sample. Statistics Canada, in cooperation with IEA Hamburg, selected the school samples for all countries and benchmarking participants.

Exhibit 6.3: Survey Activities Questionnaire, Section One—Sampling (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "PIRLS 2016 Survey Operations Procedures Unit 1 – Sampling Schools and Obtaining their Cooperation" clear and sufficient?	51	0	1
Were there any conditions or organizational constraints that necessitated deviations from the basic PIRLS sampling design described in the "Survey Operations Procedures Unit 1"?	2	49	1
Did you use the Within-school Sampling Software (WinW3S) to sample classes?	49	2	1
Did you experience any problems or inconveniences when using the WinW3S software?	16	32	4
Was the information provided in the "PIRLS 2016 Survey Operations Procedures Unit 3 – Contacting Schools and Sampling Classes for the Data Collection" clear and sufficient?	50	1	1
Did you follow the procedures outlined in "Survey Operations Procedures Unit 3" for working with the schools to sample classes (e.g., using the appropriate tracking forms in the proposed order to obtain information from School Coordinators)?	38	13	1

Two National Research Coordinators reported not using the Windows® Within-school Sampling Software (WinW3S) provided by IEA Hamburg to select classes within the sampled schools. One of them was for a benchmarking entity of a participating country, for which WinW3S was used centrally to sample classes within schools for the whole country. National Research Coordinators did report experiencing problems using the WinW3S Software. Among the issues reported were the slow speed of the software, the software not working on a shared network, issues importing information from Excel, problems with "right-to-left" languages, and issues coordinating PIRLS and ePIRLS participation status. National Research Coordinators also suggested that the software could be improved through the addition of an export to Excel function.

Thirteen National Research Coordinators applied some modifications to the procedures outlined in the *Survey Operations Procedures Unit 3*. For example, some National Research Coordinators did not use the Class Listing Forms because all classes at the target grade were tested or because a class level database was available at the ministry, and a number of countries did not use the Teacher Tracking Forms because there was only one teacher per class. All modifications were reviewed and approved by the TIMSS & PIRLS International Study Center.

Translating, Adapting, and Producing Assessment Instruments

The second section of the Survey Activities Questionnaire asked National Research Coordinators about translating, adapting, assembling, and printing the test materials, as well as issues related to checking the materials and securely storing them. Some ePIRLS specific questions were asked in this section that were related to using the ePIRLS Online Translation System, receiving ePIRLS Software, and preparing USBs in order to deliver ePIRLS to schools and students.

As reported in Exhibit 6.4, almost all National Research Coordinators found the instructions on preparing achievement booklets, context questionnaires, and ePIRLS assessment tasks clear and sufficient. However, ten countries reported experiencing some problems using the survey instrument production materials and/or the ePIRLS Online Translation System. These problems mostly included issues with fonts and special characters (e.g., for Cyrillic alphabet), difficulty fitting longer national text in the context questionnaires, and some problems with the layout style of tables. Among the problems reported about using the ePIRLS Online Translation System were inconsistencies between PDF storyboards and the translation system, some text not exporting properly to PDF, and the inability to hyphenate words. All of the identified problems were resolved either by specialists at the national center or with assistance from IEA Hamburg and the TIMSS & PIRLS International Study Center.

All National Research Coordinators, except one for a benchmarking entity of a participating country, reported applying corrections to their survey instruments as suggested by the external translation verifier or the layout verifier.

Exhibit 6.4: Survey Activities Questionnaire, Section Two—Translating, Adapting, and Producing Assessment Instruments (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "PIRLS 1016 Survey Operations Procedures Unit 4 – Preparing Achievement Booklets and Context Questionnaires" clear and sufficient?	49	2	1
Was the information provided in the "ePIRLS Supplement to the PIRLS 2016 Survey Operations Procedures Unit 4 – Preparing the ePIRLS Assessment Tasks" clear and sufficient?	15	0	0
Did you encounter any major problems using the assessment instrument production materials (e.g., instrument production files, fonts, support materials) provided by the TIMSS & PIRLS International Study Center?	4	47	1
Did you encounter any major problems with the ePIRLS Online Translation System?	6	9	0
After the translation verification, did you correct your translations/ adaptations as suggested by the verifier in the majority of cases?			
PIRLS/PIRLS Literacy booklets	50	1	0 *(Not Answered)* 1 *(Not Applicable)*
Context Questionnaires	49	1	0 *(Not Answered)* 2 *(Not Applicable)*
ePIRLS assessment tasks	13	1	0 *(Not Answered)* 1 *(Not Applicable)*
After the layout verification, did you correct your assessment instruments as noted by the verifier in the majority of cases?			
PIRLS/PIRLS Literacy booklets	51	0	0 *(Not Answered)* 1 *(Not Applicable)*
Context Questionnaires	50	0	0 *(Not Answered)* 2 *(Not Applicable)*
ePIRLS assessment tasks	13	0	0 *(Not Answered)* 2 *(Not Applicable)*
Did you apply any quality control measures to check the achievement booklets and context questionnaires during the printing process (e.g., checking for missing pages, upside down pages, text too bright or too dark)?	49	3	0
Did you experience any problems receiving the final ePIRLS Software from IEA Hamburg and preparing the ePIRLS USB sticks?	2	13	0
Did you apply any quality control of the prepared ePIRLS USB sticks before sending them to the participating schools?	11	4	0
Did you take measures to protect the security of the assessment instruments during the translation, assembly, and printing process?	51	1	0
Did you detect any potential breaches in security of the assessment instruments?	0	51	0
Did you encounter any problems preparing the Online SurveySystem files for administering the school, teacher, and/or home (Early Learning Survey) questionnaires online?	1	22	0 *(Not Answered)* 29 *(Not Applicable)*

Nearly all of the countries conducted the recommended quality control checks during the process of printing the testing materials for PIRLS and preparing USBs for ePIRLS. The most common errors detected and fixed during the printing process were pages that were missing or in the wrong order. For ePIRLS, two countries reported issues with their initial ePIRLS Software, which was then corrected and new software provided.

One country reported that they experienced a problem with the IEA's Online SurveySystem (OSS). They reported that they could not print from the OSS web print preview.

Assessment Administration

The third section of the Survey Activities Questionnaire addressed the extent to which National Research Coordinators detected errors in the testing materials during packaging for shipment to schools. As shown in Exhibit 6.5, a small number of errors were found in the materials. Approximately half of such errors were corrected before distributing the materials to the respondents. Errors found after distribution usually were very minor, and either were fixed by School Coordinators or replacement materials were provided. The few cases where the errors could not be remedied were reported to the TIMSS & PIRLS International Study Center, where decisions were made about setting the problematic data to "Not Administered."

Exhibit 6.5: Survey Activities Questionnaire, Section Three—Assessment Administration (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "PIRLS 2016 Survey Operations Procedures Unit 5 – Conducting the Data Collection" clear and sufficient?	51	0	0
Were any errors detected in any of the following assessment materials after they were sent to schools?			
Achievement booklets	11	41	0 *(Not Answered)* 0 *(Not Applicable)*
Achievement booklet ID labels	6	45	0 *(Not Answered)* 1 *(Not Applicable)*
ePIRLS USB Sticks	0	15	0 *(Not Answered)* 0 *(Not Applicable)*
Student Questionnaires	7	44	0 *(Not Answered)* 1 *(Not Applicable)*
Student Questionnaire ID labels	5	45	0 *(Not Answered)* 2 *(Not Applicable)*
Learning to Read Surveys	3	45	0 *(Not Answered)* 4 *(Not Applicable)*
Learning to Read Survey ID labels	3	45	0 *(Not Answered)* 4 *(Not Applicable)*
Student Tracking Forms	3	48	0 *(Not Answered)* 1 *(Not Applicable)*
Teacher Questionnaires	0	50	0 *(Not Answered)* 2 *(Not Applicable)*

Exhibit 6.5: Survey Activities Questionnaire, Section Three—Assessment Administration (Numbers of NRC Responses) (Continued)

Question	Yes	No	Not Answered
Teacher Tracking Forms	0	48	0 *(Not Answered)* 4 *(Not Applicable)*
School Questionnaires	2	49	0 *(Not Answered)* 1 *(Not Applicable)*
School Coordinator Manuals	2	48	0 *(Not Answered)* 2 *(Not Applicable)*
Test Administrator Manuals	5	45	0 *(Not Answered)* 2 *(Not Applicable)*
If any errors were detected, did you correct the error(s) before the testing began?	17	18	0 *(Not Answered)* 17 *(Not Applicable)*
Does your country have a confidentiality policy that restricts putting student names on tracking forms and survey instrument covers?	13	39	0
Did you encounter any problems translating and/or adapting the School Coordinator Manual?	1	51	0
Did you encounter any problems translating and/or adapting the "Preparing Computers for ePIRLS" instructions?	0	15	0
Did you experience any software-specific problems when using the ePIRLS System Check Program to test computers for ePIRLS comparability?	2	13	0
Did you encounter any problems translating and/or adapting the Test Administrator Manual(s)?	2	50	0
Were School Coordinators appointed from within the participating schools?	44	8	0
Did you hold formal training session(s) for School Coordinators?	31	21	0
Were Test Administrators trained by School Coordinators within the participating schools?	30	22	0
Did Test Administrators document any problems or special circumstances that occurred frequently during the assessment administration (please refer to the completed Test Administration Forms)?	15	37	0
Did you require/suggest/provide an additional person to help the Test Administrator during the ePIRLS testing sessions?	14	1	0
Did you have a sufficient number of computers available for all/most schools to test all of the selected students (the whole class) at the same time?	8	7	0
Did you experience any software-specific problems when using the ePIRLS Software?	8	7	0
Did you use the individual computers/USB sticks or the server method to administer ePIRLS in your country?			
Individual computers/USB sticks	10	-	0
Server method	0	-	0
Both methods were used	5	-	0
Did you experience any software-specific problems when using the ePIRLS Online Data Monitor?	4	11	0
If you administered school, teacher, and/or home (Learning to Read Survey) questionnaires online, did any of the respondents in your country encounter any problems responding to the online questionnaires?	3	17	0 *(Not Answered)* 31 *(Not Applicable)*

Three National Research Coordinators reported difficulties translating the School Coordinator Manual and/or the Test Administrator Manual. Primarily, problems arose when the manual(s) had to be reorganized or adapted and the standardized procedures were modified (e.g., no Class Listing Forms or Teacher Tracking Forms were used).

Preparing computers for ePIRLS went smoothly—no country participating in ePIRLS reported problems translating and adapting the instructions provided and only two National Research Coordinators reported problems with the ePIRLS System Check Program. For these participants, changes had occurred on some of the computers in some schools after the initial system check, and the ePIRLS Software could not be run on these computers despite a successful initial ePIRLS compatibility test. One country reported problems with running the ePIRLS Software on Apple computers via the server method—their Apple computers could not open the ePIRLS Software.

In 44 countries, School Coordinators were appointed from within the participating schools and in the remaining countries, School Coordinators were from the national center or were contracted externally. In most countries, the National Research Coordinators organized training sessions for School Coordinators. In some, mostly larger countries, training was conducted either online or in a written form via extended manuals. In 30 countries, Test Administrators were trained by the School Coordinators within the participating schools.

Although the PIRLS administration when very well, Test Administrators occasionally reported difficulties. Among the problems documented by Test Administrators during assessment administration were the following: loud noises outside the classroom, many students asking questions, confusion about the PIRLS Reader and its booklet, too much time, not enough time, some technical problems with the ePIRLS administration, the student questionnaire being too long, confusion about the spare assessment materials, and student complaints that the test was too difficult.

In all but one country participating in ePIRLS, an additional person helped Test Administrators during the ePIRLS testing sessions. Half the ePIRLS countries had enough ePIRLS compatible computers to test all students in the participating classes at the same time, while the other half organized more than one testing session for all or some of the classes. In about half the ePIRLS countries, some problems occured when running the ePIRLS Software. These included computers freezing during the testing session, students preferring the keypad mouse instead of the suggested external mouse, USBs failing if used multiple times, and data upload being too slow. The release of Windows 10 just before the administration of the assessment also led to a number of issues. In all but a few cases, ePIRLS was successfully administered despite the reported issues. Most countries used individual computers and USBs to deliver ePIRLS, and five countries used both the USB method and the server method. No ePIRLS country used the server method exclusively.

National Quality Control Program

The fourth section of the Survey Activities Questionnaire addressed the National Quality Control Program that each country implemented during data collection. As part of the national quality assurance activities, National Research Coordinators were instructed to send National Quality Control Observers to ten percent of the participating schools to observe both PIRLS and ePIRLS test administration and to document compliance with the prescribed procedures. This was in addition to the program of International Quality Control visits conducted by IEA. Primarily due to budgetary constraints, some countries sent national monitors to less than ten percent of participating schools, and two countries did not send monitors to any of the testing sessions.

As shown in Exhibit 6.6, when applicable, almost all of the national centers conducted their quality assurance program using the National Quality Control Monitor Manual provided by the TIMSS & PIRLS International Study Center. Among the few documented problems detected by the national monitors were some students being late or absent, students complaining about the length of the student questionnaire, some ePIRLS technical issues, and students being confused about clicking on links during the ePIRLS testing. In addition, one case was noted where the national monitor felt the Test Administrator was unprepared.

Exhibit 6.6: Survey Activities Questionnaire, Section Four—National Quality Control Program (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Did you conduct a national quality control program that observed the data collection in the participating schools?	50	2	0
Did you use the National Quality Control Monitor (NQCM) Manual and the Classroom Observation Record provided by the TIMSS & PIRLS International Study Center to conduct your national quality control program?	48	2	0 *(Not Answered)* 2 *(Not Applicable)*
Did your national quality control monitors (NQCMs) document any major problems or special circumstances that occurred frequently during the assessment administration?	7	43	0 *(Not Answered)* 2 *(Not Applicable)*

Preparing for and Scoring the Constructed Response Items

Exhibit 6.7 provides data on responses to items asking National Research Coordinators about their experiences preparing for and scoring the constructed response items. All National Research Coordinators found the scoring procedures as explained in the *Survey Operations Procedures Unit 6—Scoring the Constructed Response Items,* including the ePIRLS supplement, to be clear and sufficient. Countries reporting problems with the scoring training materials asked for more "borderline" examples, including more detailed explanations within the scoring guides. Some countries also reported difficulties translating the examples both in the scoring guides and in the training materials. More than half of National Research Coordinators reported creating their own national examples and practice papers for training their scorers, as suggested by the TIMSS & PIRLS International Study Center.

About half of ePIRLS countries reported some minor problems using the ePIRLS Online Scoring System. The reported problems included the system being slow, the system not responding at times, issues with the "zooming function," a few student responses being assigned to scorers more than once, issues with the "flag function," and some countries wished to have a training module to be used before the actual scoring began.

Exhibit 6.7: Survey Activities Questionnaire, Section Five—Preparing for and Scoring the Constructed Response Items (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "PIRLS 2016 Survey Operations Procedures Unit 6 – Scoring the Constructed Response Items" clear and sufficient?	52	0	0
Was the information provided in the "ePIRLS Supplement to the PIRLS 2016 Survey Operations Procedures Unit 6 – Scoring the Constructed Response Items" clear and sufficient?	15	0	0
Did you encounter any problems using the scoring training materials, provided by the TIMSS & PIRLS International Study Center?	13	39	0
Did you create national scoring training materials in addition to the international scoring training materials?	28	24	0
Did you scan the achievement booklets for electronic image scoring?	16	36	0
Did you encounter any problems using the ePIRLS Online Scoring System?	8	7	0
Did you encounter any problems during the Trend Reliability Scoring?			
Procedural problems	3	42	0 *(Not Answered)* 7 *(Not Applicable)*
Technical, software related problems	13	32	0 *(Not Answered)* 7 *(Not Applicable)*
Did all your scorers participate in scoring student responses of the trend items?	29	16	0 *(Not Answered)* 7 *(Not Applicable)*
Did you encounter any problems during the Cross-country Reliability Scoring?			
Procedural problems	2	45	0 *(Not Answered)* 5 *(Not Applicable)*
Technical, software related problems	14	33	0 *(Not Answered)* 5 *(Not Applicable)*
Did all your scorers participate in the Cross-country Reliability Scoring?	20	27	0 *(Not Answered)* 5 *(Not Applicable)*

Sixteen countries scanned their PIRLS achievement booklets and scored student responses electronically. Some technical problems were encountered while using the IEA's Coding Expert Software for the trend and cross-country scoring. Mostly countries reported the scans displayed via the Coding Expert Software being of poor quality and difficult for the scorers to read. Because

English was used for the cross-country reliability scoring task, five countries were unable to participate. For the countries that did not participate in the previous cycle of PIRLS, the question on the trend reliability scoring procedures did not apply.

Creating the Databases

The last section of the Survey Activities Questionnaire addressed data entry of the paper assessment instruments and data quality control activities. As shown in Exhibit 6.8, almost all of the National Research Coordinators found the instructions in *Survey Operations Procedures Unit 7* to be clear and sufficient. Some National Research Coordinators expressed a wish for a more automated data entry process in WinW3S, as some issues arose with the import and export functions. Also, the administration mode of the school, teacher, and home questionnaires was set to the same mode for all respondents. If some respondents, especially parents or guardians, chose to complete their questionnaire online, this status had to be adjusted manually.

Exhibit 6.8: Survey Activities Questionnaire, Section Six—Creating Databases (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "PIRLS 2016 Survey Operations Procedures Unit 7 – Creating the Databases" clear and sufficient?	50	2	0
Did you encounter any problems entering test administration information and exporting your WinW3S database?	16	36	0
Who primarily entered the data for your country?			
National center staff	13	-	0
Temporarily hired data entry staff	20	-	0
An external data entry firm	8	-	0
Combination of the above	8	-	0
Other	3	-	0
Did you use manual (key) data entry to create the data files for your country?			
Achievement booklets	35	15 *(Optical Scanning)*	0 *(Not Answered)* 2 *(Not Applicable)*
Context questionnaires	39	12 *(Optical Scanning)*	0 *(Not Answered)* 1 *(Not Applicable)*
Did you encounter any problems using the IEA's Data Manager Expert (DME) software?	4	48	0
If you entered data manually, did you enter 5% of each survey instrument twice as a quality control measure?	35	8	0 *(Not Answered)* 9 *(Not Applicable)*
Did you apply all the data quality checks described in the "PIRLS 2016 Survey Operations Procedures Unit 7 – Creating the Databases" before submitting your data to IEA Hamburg?	51	1	0
Have you stored all achievement booklets and context questionnaires in a secure storage area until the original documents can be discarded?	52	0	0

Most countries reported hiring temporary data entry staff to enter data manually. In 13 countries, the national center staff entered data from the paper instruments. A number of countries used optical scanning instead of manual data entry. All but one country reported applying all required data quality checks. All countries reported having securely stored their original assessment instruments until all data are processed and reported, and these materials can be destroyed.