

Operations and Quality Assurance

Ieva Johansone

Overview

As data-based indicators of countries' student achievement profiles and learning contexts, TIMSS and PIRLS assessments are crucially dependent on the quality of the data collected by each participant. Whereas the development of the 2011 assessments were an intensely collaborative process involving all of the partners in the enterprise, the process of administering the assessments and collecting the data was 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, TIMSS and PIRLS have developed a set of standardized operations procedures that all participants are expected to follow. These procedures have been developed over successive cycles of TIMSS and PIRLS through a partnership involving the TIMSS & PIRLS International Study Center, the IEA Data Processing and Research Center (DPC), the IEA Secretariat, Statistics Canada, and National Research Coordinators (NRCs). With each new TIMSS or PIRLS cycle, the operations procedures have been updated to enhance efficiency and accuracy and reduce burden, making use of developments in information technology to automate routine activities wherever possible.

In each country or benchmarking entity, the National Research Coordinator was responsible for the implementation of TIMSS and/or PIRLS in 2011. Internationally, National Research Coordinators provided the country's perspective in all international discussions, represented the country at international meetings for TIMSS and/or PIRLS, 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 TIMSS

and/or PIRLS, including any adaptations for the national context. The daily tasks of the NRCs evolved over the course of the TIMSS and/or PIRLS 2011 cycle. In the initial phases of the TIMSS and PIRLS 2011 cycle, National Research Coordinators participated in the TIMSS and/or PIRLS 2011 framework and assessment development process (see [Assessment Framework and Instrument Development](#)), and collaborated with Statistics Canada and the IEA DPC to develop a plan to implement the TIMSS and/or PIRLS sampling design within the country or benchmarking entity (more information can be found in [Sample Design in TIMSS and PIRLS](#)).

Following the development of the draft achievement items and context questionnaires, all countries conducted a full-scale field test of all instruments and operational procedures in March/April 2010, in preparation for the data collection, which took place in October-December 2010 in Southern Hemisphere countries and in March-May 2011 in Northern Hemisphere countries. This allowed the National Research Coordinators and their staff to become acquainted with the activities and provide feedback that was used to improve the procedures for the data collection. As expected, the field test resulted in some small modifications to survey operations procedures and most definitely contributed significantly to the successful execution of the TIMSS and/or PIRLS 2011 assessments.

To support the National Research Coordinators in conducting TIMSS and/or PIRLS 2011, the TIMSS & PIRLS International Study Center provided step-by-step documentation of all operational activities. Organized into a series of units, the Survey Operations Procedures were made available at critical junctures of the projects to ensure that NRCs had all the tools and information necessary to discharge their responsibilities (see [TIMSS and PIRLS 2011 Survey Operations Procedures Units and Manuals](#)). 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. Consistent with the goal of automating and streamlining procedures wherever possible, the IEA DPC provided NRCs with a range of custom-built software products to support activities, including sampling and tracking classes and students, administering school and teacher questionnaires, documenting scoring reliability, and creating and checking data files (see [TIMSS and PIRLS 2011 Survey Operations Procedures Software](#)).

The TIMSS & PIRLS International Study Center and the IEA DPC also provided NRCs and their staff with intensive training in constructed-response item scoring and data management.

The 2011 assessment cycle was unique in terms of the Survey Operations Procedures in that both TIMSS 2011 and PIRLS 2011 assessment cycles were administered within the same timeframe, and countries had the opportunity to administer both fourth-grade assessments to the same students. Where operations were similar for the two studies, combined PIRLS and TIMSS units were provided. However, whenever differences were considerable and crucial, the TIMSS & PIRLS International Study Center provided three versions—a combined PIRLS and TIMSS version for countries assessing the same students in both studies, as well as separate versions for PIRLS and TIMSS.

Complementing the effort to equip National Research Coordinators with the skills and tools necessary to conduct their TIMSS and PIRLS operational tasks, the TIMSS & PIRLS International Study Center worked with the IEA Secretariat, the IEA DPC, and Statistics Canada to implement quality assurance measures to monitor compliance with operational procedures and document the quality of the data collection effort in each country and benchmarking entity. In particular, TIMSS and PIRLS conducted a program of school visits by International Quality Control Monitors (IQCMs). The IQCMs visited a sample of fifteen schools per assessment population to observe the administration of the assessment, collect data on the administration process, and evaluate the quality of the testing session. For the 2011 cycle, taking TIMSS and PIRLS together, the IQCMs observed nearly 2,000 assessment sessions (see [Quality Control Observations of the TIMSS 2011 Data Collection](#) and [Quality Control Observations of the PIRLS/prePIRLS 2011 Data Collection](#)). In addition to the international program of school visits, each country or benchmarking participant was expected to implement a national quality control program that also included school visits.

Finally, as part of ongoing efforts to improve TIMSS and PIRLS operations, the National Research Coordinators were asked to complete a Survey Activities Questionnaire (SAQ), which sought feedback on all aspects of their experience conducting TIMSS 2011 and PIRLS 2011. 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 (see [Results of the TIMSS and PIRLS 2011 Survey Activities Questionnaire](#)).

Operations for Data Collection

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

- ◆ contacting schools and sampling classes,
- ◆ overseeing translation and preparing assessment instruments,
- ◆ managing the administration of the TIMSS and/or PIRLS 2011 assessments,
- ◆ scoring of the constructed-response items, and
- ◆ creating the TIMSS and PIRLS 2011 data files.

Two major TIMSS and PIRLS 2011 operational activities, sampling schools and translating the assessment booklets and background questionnaires, are described in separate sections of *Methods and Procedures in TIMSS and PIRLS 2011* (see [Sample Design in TIMSS and PIRLS](#) and [Translation and Translation Verification](#)).

Contacting Schools and Sampling Classes

Once the school samples were drawn (see [Sample Design in TIMSS and PIRLS](#)), one of the essential, first steps for the TIMSS and PIRLS 2011 National Research Coordinators was to establish good working relationships with the selected schools. NRCs were responsible for 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. Securing school participation in the assessment(s) was an essential part of the NRC's role.

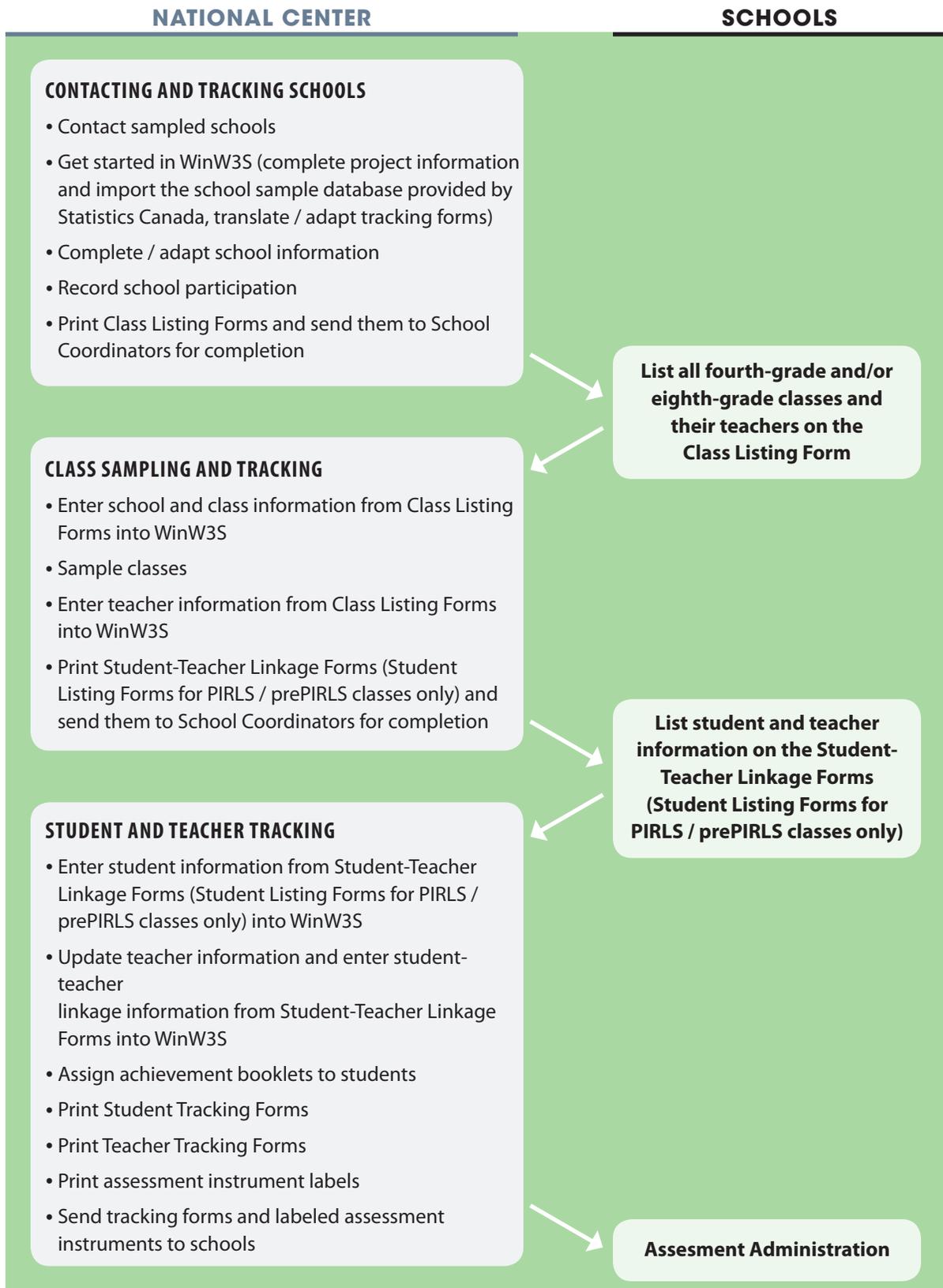
In cooperation with school principals, National Research Coordinators were responsible for identifying and training School Coordinators for all participating schools. A School Coordinator could be a teacher or guidance counselor in the school, or NRCs could appoint a member of the national center to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

The responsibilities of the School Coordinator included providing the national center with information on the school; coordinating the date, time, and place for testing; identifying and training a Test Administrator to administer the assessment; coordinating the completion of the [TIMSS 2011 and/or PIRLS 2011 tracking forms](#); distributing questionnaires; and obtaining parental permission (if necessary). 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 the administration of the assessment(s).

School Coordinators also played a critical role in providing information for the sampling process, providing the national center with data on eligible classes in the school. With this information, the national centers used the Within-school Sampling Software (WinW3S) to sample class(es) within the school. WinW3S tracked school, teacher, and student information. The software also generated the necessary tracking forms and instrument labels which facilitated the assessment administration and checking during the data cleaning process (see [TIMSS and PIRLS 2011 Survey Operations Procedures Software](#)).

As intact classes were sampled, one of the roles of the School Coordinator was to ensure that every student in the school was listed in one and only one class (course). This was necessary to ensure that the sample of classes results in a representative sample of students, and every student at the target grade has a chance of being selected. At fourth grade in most countries, students are taught mathematics, science and reading in the same classroom, and therefore the fourth-grade classroom was designated as the sampling unit. At the eighth grade, however, in many countries students are grouped differently for mathematics and science instruction. In other words, a student takes his/her mathematics class with one group of students and his/her science class(es) with a different group of students. As the sampling required one set of students who could be considered a classroom, eighth-grade classrooms usually were defined on the basis of mathematics instruction for the purposes of sampling.

The following exhibit illustrates the major steps of working with schools to sample classes and prepare for the TIMSS and/or PIRLS 2011 assessment administration.



Overseeing Translation and Preparing Assessment Instruments

National Research Coordinators also were responsible for preparing the assessment instruments (achievement booklets and questionnaires) for their countries. The overarching goal of assessment instrument preparation was to create internationally comparable achievement booklets and background questionnaires that are appropriately adapted for the national context.

For both TIMSS and PIRLS, each student completed one of the achievement booklets (see the [PIRLS 2011 Assessment Framework](#) and the [TIMSS 2011 Assessment Frameworks](#) for more information on the matrix sampling design). The achievement booklets for both TIMSS and PIRLS are composed of blocks of assessment items, repeated in more than one booklet. From an operational perspective, each block needed to be translated only once, even though it could be included in several booklets. Adobe®InDesign® software is used by countries to link the translated and adapted assessment blocks to the appropriate booklets. By automating this process through the use of Adobe®InDesign® software, the possibility of human error in the production process was decreased.

Twelve new assessment blocks at each grade level were developed for TIMSS 2011 (six mathematics and six science) and four new assessment blocks were developed for PIRLS 2011. These new assessment blocks replaced the ones released at the end of the previous assessment cycles. Also, eight new assessment blocks were developed for prePIRLS. The assessment blocks were tried out through the field test in order to investigate the psychometric characteristics of the achievement items and make well-informed decisions about the best replacements. Similarly, the background 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 and background questionnaires into their language(s) for the field test (see [Translation and Translation Verification](#)). After the field test, the best replacement blocks were chosen for the main data collection and some edits were applied to the items and background questionnaires. National Research Coordinators were responsible for applying these changes to the translated assessment blocks and questionnaires. Countries taking TIMSS or PIRLS for the first time had to translate and/or adapt the assessment blocks used in previous assessments (trend blocks) into their language(s) in preparation for the 2011 assessment

administration. Countries that had participated in previous assessment cycles used the same translations as in those assessments.

For both the field test and main data collection, the participating countries received the international version (English) of the achievement booklets and background questionnaires with all the necessary instrument production files, including fonts and graphic files. Instructions on how to use the materials to produce high quality, standardized instruments, were included in the corresponding *Survey Operations Procedures* unit.

As 13 countries were planning to administer the TIMSS 2011 assessments in Arabic, the IEA Secretariat and the TIMSS & PIRLS International Study Center also provided a generic Arabic version of the TIMSS 2011 assessment booklets and background questionnaires for both fourth and eighth grades. Individual countries adapted the generic version to local usage.

Once translated and/or adapted, first for the field test and then again for the main data collection, the contents of the achievement booklets and background questionnaires were submitted to the IEA Secretariat for an independent translation verification. The IEA Secretariat worked with independent translators to evaluate each country's translations and when deemed necessary suggest changes to the text (see [Translation and Translation Verification](#) for more information on translation and national adaptations of the assessment instruments).

TIMSS & PIRLS International Study Center Review of the Assessment Instruments

After the translations and adaptations had been verified by the IEA Secretariat, National Research Coordinators assembled the achievement booklets and background questionnaires using Adobe® InDesign® software, and print-ready copies of the booklets were sent to the TIMSS & PIRLS International Study Center for layout verification and a final review of national adaptations. This review checked that each booklet and questionnaire conformed to the international format and that any adaptations made to the instruments did not unduly influence their international comparability.

During layout verification, the national assessment instruments were checked against the international version to identify any deviations from the international format. Verifiers checked for any discrepancies in pagination, page breaks, item sequence, response options, text formats, graphics, etc. Naturally, the translated assessment instruments from the participating countries did not have exactly the same format as text length varies between languages. The

international versions, however, were designed with this in mind, and extra space was provided in the margins of the pages to facilitate the use of a longer text and different size paper without necessitating extensive changes to the layout of the instruments.

All national adaptations to the international assessment instruments were documented using the National Adaptations Forms. During the review, the TIMSS & PIRLS International Study Center checked if the national adaptations to the national background questionnaires could influence the ability to produce internationally comparable data for the affected questions. Any questions raised by the TIMSS & PIRLS International Study Center were directed to the NRC for consideration.

Managing the Administration of the TIMSS and/or PIRLS 2011 Assessments

Printing assessment materials and distributing them to the participating schools required careful organization and planning on the part of the National Research Coordinator. The TIMSS and PIRLS achievement booklets were assigned to students in a systematic rotation so that each achievement block was assigned to an equal number of students. This process is facilitated by the labels and tracking forms generated by WinW3S. Each student also was assigned a Student Questionnaire, which was labeled so that it could be linked to the achievement booklet. In PIRLS, the student's parents were assigned the Learning to Read Survey, which also was linked to the student's achievement booklet. In addition, an individually labeled Teacher Questionnaire was sent to each teacher listed on the Teacher Tracking Form and a School Questionnaire was sent to the principal. These materials were packaged and sent to the schools (School Coordinators) prior to the testing date, giving ample time for the School Coordinators to confirm the receipt and correctness of the materials. The School Questionnaire and Teacher Questionnaire were then distributed, while the other instruments were kept in a secure room until the testing date (for more information on linking please see [Linking Students to their Teachers and Classes](#)).

Each sampled class was assigned a Test Administrator who followed procedures described in the Test Administrator Manual to administer the assessment (achievement booklets) and Student Questionnaires. This person was chosen and trained by the School Coordinator. In many cases, the School Coordinator doubled as the Test Administrator. The Test Administrator was

responsible for distributing materials to the appropriate students, reading to the students the instructions provided in the Test Administrator’s manual, and timing the sessions. The Test Administrator documented the timing of the testing sessions on the Test Administration Form. The Test Administration Form also solicited information about anything out of the ordinary that took place during the assessment administration.

Both TIMSS and PIRLS achievement booklets contained two sections, and the time allotted for each section of the assessment was standardized and strictly enforced by the Test Administrator. There was a required break in between the two sections of the assessment administration, and this break was not to exceed 30 minutes. To complete each part of the TIMSS achievement test, fourth grade students were allowed 36 minutes and eighth grade students were allowed 45 minutes. For the PIRLS achievement test, students were allotted 40 minutes to complete each section. If a student completed part 1 or part 2 of the assessment before the allotted time, the student was not allowed to leave the testing room. Students completing the assessments early were asked to review their answers or read quietly. Some test administrators provided an activities sheet for these students. To complete the Student Questionnaire, students were given at least 30 minutes, but extra time was given when necessary. Also, for fourth grade students, the Test Administrator was permitted to read the questionnaire items aloud together with the students. For fourth-grade students assessed for both TIMSS and PIRLS, only one combined Student Questionnaire was administered, usually after the administration of the first assessment.

The Test Administrator was required to use the Student Tracking Form and labels to distribute the booklets to the correct students and to document student participation. If the participation rate was below 90 percent in any class, 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.

Online Administration of the TIMSS 2011 and PIRLS 2011 Teacher and School Questionnaires

For the first time in 2011, countries could choose to administer the TIMSS and/or PIRLS school and teacher questionnaires online or on paper. The benefits of administering the questionnaires online included saving money and time in printing and improving the efficiency of the distribution of the questionnaires. Online administration also facilitated data entry and data cleaning.

The online questionnaires were administered using the IEA DPC’s Online SurveySystem Designer and, for most countries, the online questionnaire

administration was hosted on the IEA DPC’s server. Altogether, 13 countries chose to administer their school and teacher questionnaires online (for more information see [Online Teacher and School Questionnaire Administration](#)). Curriculum Questionnaires and Survey Activities Questionnaires also were administered online using the IEA’s Online SurveySystem. However, these questionnaires were administered in English and did not require any translation/adaptation.

Scoring the Constructed-response Items

Constructed-response items represent a substantial portion of the TIMSS and PIRLS assessments, and the reliability and validity of scoring is critical to the assessment results. To achieve the goal of reliable and valid scoring, the TIMSS & PIRLS International Study Center provided detailed, explicit scoring guides for each item and extensive training in their use. Also, the [Survey Operations Procedures units](#) specified a procedure for organizing and implementing the scoring activity.

Two international scoring training sessions—one for the field test and another for the main data collection—were held for each assessment, where all National Research Coordinators (or the country representative appointed by the NRC) 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 actual student answers from pilot testing in several English-speaking countries, and 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 then were responsible for training 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

Reliable scoring of the constructed response items is essential for high quality TIMSS and/or PIRLS data. A high degree of scorer agreement is evidence that scorers have applied the scoring guides in the same way. A high degree of scorer agreement is evidence that scorers have applied the scoring guides in the same way. The procedure for scoring the TIMSS 2011 and/or PIRLS 2011 constructed-response items provided for documenting scoring reliability within each country (within-country reliability scoring), across countries (cross-country reliability scoring), and over time (trend 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 twice. 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 purpose of the trend reliability scoring was to measure the reliability of the scoring from one assessment cycle to the next (i.e., from 2007 to 2011 for TIMSS and from 2006 to 2011 for PIRLS). The trend reliability scoring required scorers of the current assessment to score student responses collected in the previous cycle. The scores of the current cycle were then compared with the scores awarded in the previous assessment cycle. Trend reliability scoring was conducted using the Trend Scoring Reliability System (TSRS) provided by the IEA DPC (see [TIMSS and PIRLS 2011 Survey Operations Procedures Software](#)).

Student responses included in the trend reliability scoring (200 responses per item) were actual student responses collected during the previous assessment cycle in each country and benchmarking entity. These responses were scanned and provided for each participating country and benchmarking entity on individually prepared DVDs. All scorers who scored the trend assessment blocks in 2011 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 item blocks, National Research Coordinators were able to specify within the software which scorers would score particular item 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.

The trend reliability scoring software also included a database with student responses to be scored as a qualification round in order to determine whether scorers were sufficiently trained to begin the actual scoring. Using the qualification database, each scorer was required to score 25 responses to each item assigned for him or her. Then, the National Research Coordinator (or the scoring supervisor appointed by the NRC) calculated the scorer's degree of agreement with the scores given by the previous cycle scorers. If the degree of agreement was at 85% or above, the scorer could begin the actual 2011 scoring. If not, the scorer was required to receive additional training, and then score another 25 responses to items assigned for him or her. If the degree of agreement still was below 85%, the NRC had to consider whether this scorer should be involved in the scoring.

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 was conducted using the Cross Country Reliability Scoring System (CCRS) provided by the IEA DPC (see [TIMSS and PIRLS 2011 Survey Operations Procedures Software](#)). To begin the process, the IEA DPC compiled actual responses of students from English speaking countries participating in previous cycles as well as from English-speaking Southern Hemisphere countries participating in the 2011 cycle (as the Southern Hemisphere collected data in the autumn of 2010, their student responses were available for this exercise). For PIRLS 2011, 26 constructed-response items were included for the cross-country reliability scoring. For TIMSS 2011, there were 38 items included at the fourth-grade and 50 items included at the eighth-grade. Two hundred student responses for each item were scanned by the IEA DPC and provided to countries and benchmarking entities on DVDs. 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 TIMSS and PIRLS 2011 scoring activities.

Creating the TIMSS and PIRLS 2011 Data Files

The procedure for creating the TIMSS and/or PIRLS 2011 data files included entering sampling and assessment administration information into the WinW3S database and entering the actual responses from the background questionnaires and achievement booklets using the WinDEM data entry software. The data entry process took place March-May 2010 for the field test, from December 2010-March 2011 following data collection in the Southern Hemisphere and June-September 2011 following data collection in the Northern Hemisphere. The IEA Data Processing and Research Center provided Windows Data Entry Manager (WinDEM) software to accommodate keyboard data entry and data verification. This program works in conjunction with WinW3S software, and the tracking information from the sampling process can be integrated with the assessment data. The WinDEM software also offers data and file management capabilities, a convenient checking and editing mechanism, interactive error detection, and reporting and quality-control procedures (see [TIMSS and PIRLS 2011 Survey Operations Procedures Software](#)). Moreover, for the TIMSS 2011 and/or PIRLS 2011 background questionnaires administered online on the IEA DPC's server, the data were directly accessible by the IEA DPC and no further data entry was required.

One of the important benefits of using WinDEM is that it incorporates international codebooks describing all variables and their characteristics, thus ensuring that the data files meet the internationally defined rules and standards for data entry. There were separate codebooks for the achievement booklets, the Reliability Scoring Sheets, and each of the background questionnaires. The files for the TIMSS and/or PIRLS 2011 data were based on these codebooks. However, the codebooks had to match the national assessment instruments exactly 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 TIMSS and/or PIRLS 2011 data files in each country, with the responses to the background questionnaires, achievement booklets, and Reliability Scoring Sheets keyed directly into WinDEM (see [TIMSS and PIRLS 2011 Survey Operations Procedures Software](#)).

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 the data entry, and for applying a series of data verification checks provided by WinDEM software prior to

submitting the data files to the IEA Data Processing and Research Center. As part of this process, the data-entry staff was required to double enter at least 5 percent of each instrument type to ensure the reliability of the data entry process. 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.

Additionally, the data verification module of WinDEM identified a range of problems, such as inconsistencies of identification codes and out-of-range or otherwise invalid codes. WinDEM software also allowed for verification of the integrity of the linkage between the students, teachers, and schools entered into the WinDEM data files and tracking of information for those specified in WinW3S.

When all data files had passed the WinDEM quality control checks, they were submitted to the IEA DPC, along with data documentation for further checking and processing. For information on data processing at the IEA DPC, please refer to the [Creating the International Databases](#) section.

Quality Assurance in the TIMSS and PIRLS 2011 Data Collection

As described above, considerable effort was made to develop standardized materials and survey operations procedures that would ensure that the TIMSS and PIRLS data met high quality standards, and that valid comparisons of student achievement across and within the participating countries could be made. The TIMSS & PIRLS International Study Center, working with the IEA Secretariat, developed and implemented an International Quality Assurance Program of school visits to document data collection activities and ensure that the standardized TIMSS and PIRLS operations were properly applied. In addition, National Research Coordinators were responsible for developing and implementing a national quality control program.

Implemented by the IEA Secretariat and the TIMSS and PIRLS International Study Center, the International Quality Assurance Program was an independent effort. However, National Research Coordinators were asked to nominate an International Quality Control Monitor (ICQM) to conduct the school visits. In quite a few countries and benchmarking entities, IEA had established good working relationships with quality control monitors from the

previous TIMSS and/or PIRLS cycles or from other IEA's studies and contracted them again for TIMSS 2011 and/or PIRLS 2011. For the remaining countries, National Research Coordinators assisted the IEA Secretariat in nominating an IQCM. In order to ensure the independence of the IQCM, the nominated person could not be a member of the national center or a family member or a personal friend of the NRC. Often, this person was a school inspector, a ministry official, or a retired school teacher. The IQCM was required to be fluent in both English and the language(s) spoken in the country.

The IQCMs were mandated to attend an extensive training session in their tasks and responsibilities. The training sessions were conducted by the IEA Secretariat and the TIMSS & PIRLS International Study Center. During the training, the IQCMs were introduced to the survey operations procedures and the design of the achievement booklets and background questionnaires. Each IQCM was provided with the necessary materials for completing the quality control tasks. In addition to the training, IQCMs were provided with a manual describing their role and responsibilities.

The major task of the IQCMs was to conduct site visits during the data collection. In each country, the IQCM visited a sample of 15 participating schools at each grade during assessment administration. When there were one or more benchmarking participants from the same country, with a centrally organized national center responsible for all aspects of data collection, five school visits were required within each benchmarking entity. During their school visits, IQCMs observed the testing session and recorded their observations. They were asked to note if any changes were made to the standardized administration script, timing, or procedures. In addition, IQCMs interviewed School Coordinators about their experiences with the TIMSS 2011 and/or PIRLS 2011 assessments.

As a final task, IQCMs checked whether or not the comments and suggestions made by the international translation verifier had been integrated into the final assessment instruments as documented in the National Adaptations Forms.

Where necessary, the IQCMs were permitted to recruit assistants in order to efficiently cover the territory and testing timetable. For TIMSS 2011 and PIRLS 2011, a total of 110 IQCMs were trained across the 69 participating countries and 17 benchmarking participants. In addition, the IQCMs trained more than 350 assistant monitors. Altogether, quality control monitors observed

776 fourth-grade TIMSS testing sessions, 692 eighth-grade TIMSS testing sessions, and 521 PIRLS/prePIRLS testing sessions. (see [Quality Control Observations of the TIMSS 2011 Data Collection](#) and [Quality Control Observations of the PIRLS/prePIRLS 2011 Data Collection](#)).

Each country's national quality control program was the responsibility of the National Research Coordinator who appointed National Quality Control Monitors (NQCMs) to visit a sample of schools. The NQCMs could be members of the national center or personnel external to the national center with a professional background, preferably with experience in school settings.

National Quality Control Monitors visited a random sample of at least 10% of the participating schools. To assist national centers with their quality control programs, the TIMSS & PIRLS International Study Center provided NRCs with a National Quality Control Observer Manual, which could be modified to suit national systems. The NQCMs were required to prepare a brief written report on their site visits that included a summary of their observations as well as any feedback from the Test Administrators and School Coordinators about their experience with the assessment and background questionnaire administration.