- Questionnaires would be administered to the tested students, their current reading teachers, and their school principals – to collect contextual data with which to interpret achievement.

In 1999, planning for the study began with a meeting among representatives from the IEA Secretariat, the International Study Center (ISC) at Boston College, Statistics Canada, and the National Foundation for Educational Research in England and Wales. At this meeting, it was established that – in addition to incorporating the General Assembly's basic principles – PIRLS would try to collect data from children's parents about literacy activities in the home, and also collect data about early reading instruction in schools to provide additional information on reading instruction (beyond what the current-year teachers would provide). These basic goals were supported by the Reading Development Group (RDG) and representatives from the participating countries (the National Research Coordinators).

The development of PIRLS spanned two years, beginning in 1999 and continuing until early 2001, when the final reading test and questionnaires were approved by the participating countries. As part of development, 30 countries conducted a field test of the test and questionnaires. Ultimately, 35 countries participated in the main data collection.

1.2   Participating Countries

Thirty-five countries joined together to

provide countries participating in both studies with achievement and background data for three subjects at the same grade level.

In each country, representative samples of students were selected using a two-stage sampling design. In the first stage, at least 150 schools were selected using probability-proportional-to-size sampling. Countries could incorporate in their sampling design important reporting variables (for example, urbanicity or school type) as stratification variables. At the second stage, one or two fourth-grade classes were randomly sampled in each school. This resulted in a sample size of at least 3,750 students in each country. Some countries opted to include more schools and classes, enabling additional analyses, which resulted in larger sample sizes.

1.4   Assessment Dates

PIRLS was administered near the end of the school year in each country. In countries in the Northern Hemisphere (where the school year typically ends in May or June) the assessment was conducted in April, May, or June 2001. In the Southern Hemisphere, the school year typically ends in November or December; so in these countries, the assessment was conducted in October or November 2001.

1.5   Study Management and Organization

PIRLS is directed in the United States by Ina V.S. Mullis and Michael O. Martin, at the International Study Center at Boston College; they also direct the IEA's TIMSS. The PIRLS International Study Center was responsible for the design, development, and implementation of the study – including developing the instruments and survey procedures; ensuring quality in data collection; and analyzing and reporting the study results. The International Study Center worked closely with the organizations responsible for particular aspects of the study, the PIRLS advisory committees, and representatives of the participating countries.

Each country appointed a National Research Coordinator (NRC) who, together with staff members at the PIRLS national center, was responsible for all aspects of the study within that country. The PIRLS ISC organized meetings of the NRCs several times a year to review study materials and procedures, and to receive training in scoring constructed-response items, and in entering the data using the prescribed software.

The IEA Secretariat provided guidance in all aspects of the study, and was responsible for managing the ambitious translation verification effort conducted for the field test and main assessment. Statistics Canada was responsible for all aspects of sampling – including working with countries to ensure that the international procedures are followed; adapting the international

design to national conditions; documenting

across test booklets 1 through 9, and that
the fourth literacy and informational blocks
a

The remaining half will be kept secure and included in future PIRLS assessments so trends in achievement can be measured. After data collection in 2001, one literary and one informational block (as well as the blocks in the Reader) were released. Effectively, four blocks (or half of the assessment) were available to the public.

1.8   PIRLS Background Questionnaires

By gathering information about children's experiences together with reading achievement on the PIRLS test, it is possible to identify the factors or combinations of factors that relate to high reading literacy. An important part of the PIRLS design is a set of questionnaires targeting factors related to reading literacy. PIRLS administered four questionnaires: to the tested students, to their parents, to their reading teachers, and to their school principals.

1.8.2   Student Questionnaire

Each student taking the PIRLS reading assessment completes the student questionnaire. The questionnaire asks about aspects of students' home and school experiences – including instructional experiences and reading for homework, self-perceptions and attitudes towards reading, out-of-school reading habits, computer use, home literacy resources, and basic demographic information.

1.8.3   Learning to Read (Home) Survey

The learning to read survey is completed by the parents or primary caregivers of each student taking the PIRLS reading assessment. It addresses child/parent literacy interactions, home literacy resources, parents' reading habits and attitudes, home/school connections, and basic demographic and socioeconomic indicators.

1.8.4   Teacher Questionnaire

The reading teacher of each fourth-grade class sampled for PIRLS completes a questionnaire designed to gather information about classroom contexts for developing reading literacy. This questionnaire asks teachers about characteristics of the class tested (such as size, reading levels of the students, and the language abilities of the students). It also asks about instructional time, materials and activities for teaching reading and promoting the development of their students' reading literacy, and the grouping of students for reading instruction. Questions about classroom resources, assessment practices, and home/school connections also are included. The questionnaire also asks teachers for their views on opportunities for professional development and collaboration with other teachers, and for information about their education and training.

1.8.5   School Questionnaire

The principal of each school sampled for PIRLS responds to the school questionnaire. It asks school principals about enrollment and school characteristics (such as where the school is located, resources available in the surrounding area, and indicators of the socioeconomic background of the student body), characteristics of reading education in the school, instructional time, school resources (such as the availability of instructional materials and staff), home/school connections, and the school climate.

1.9 Translation and Verification of Instruments

The PIRLS reading tests and questionnaires
were

1.11   Scoring the Constructed-Response Items

Because almost two-thirds of the score points came from constructed-response items, PIRLS needed to develop procedures for reliably evaluating student responses within and across countries. The International Study Center prepared detailed guides containing the PIRLS scoring rubrics, and explanations of how to implement them – together with example student responses for the various rubric categories. These guides, along with training packets containing extensive examples of student responses for practice in applying the rubrics, were used as a basis for intensive training of national representatives in scoring the constructed-response items.

To gather and document empirical information about the within-country agreement among scorers, PIRLS arranged to have a sample of 200 students' responses to each item in each country scored independently by two readers. Scoring reliability within countries was high – the percentage of exact agreement, on average, across countries, was more than 90 percent. PIRLS also conducted a study of scoring reliability across countries, asking countries with scorers proficient in English to score a reference set of student responses chosen from students in English-speaking countries. This study revealed a high level of agreement between scorers also (85% on average).

1.12   Data Processing

To ensure the availability of comparable, high-quality data for analysis, PIRLS took rigorous quality control steps to create the international database. PIRLS prepared manuals and software for countries to use in creating and checking their data files, so that the information would be in a standardized international format before being forwarded to the IEA Data Processing Center (DPC) in Hamburg for creation of the international database. Upon arrival at the DPC, the data underwent an exhaustive cleaning process involving several iterative steps and procedures designed to identify, document, and correct deviations from the international instruments, file structures, and coding schemes. The process also emphasized consistency of information within national data sets, and appropriate linking among the student, parent, teacher, and school data files.

Throughout the process, the data were checked and double-checked by the IEA Data Processing Center, the International Study Center, and the national centers. The national centers were contacted regularly and given multiple opportunities to review the data for their countries. In conjunction with the IEA Data Processing Center, the International Study Center reviewed item statistics for each cognitive item in each country to identify poorly performing items. In general, the items exhibited very good psychometric properties in all countries.

1.13   IRT Scaling

The general approach to reporting the PIRLS achievement data was based primarily on item response theory (IRT) scaling methods. Student reading achievement was summarized using a family of IRT models (2-parameter, 3-parameter, and generalized partial credit models). The IRT methodology was preferred for developing comparable estimates of performance for all students, since students responded to different passages and items depending upon which of the test booklets they received (Booklet 1 through 9 or the PIRLS Reader). This methodology produces a score by averaging the responses of each student to the items that he or she took in a way that takes into account the difficulty and discriminating power of each item. The approach followed in PIRLS uses information from the background questionnaires to provide improved estimates of student performance (a process known as conditioning) and multiple imputation to generate student scores (or "plausible values") for analysis and reporting.

The IRT analysis provides a common scale on which performance can be compared across countries. In addition to providing a basis for estimating mean achievement, scale scores permit estimates of how students within countries vary and provide information on percentiles of performance. Treating all participating countries equally, the PIRLS scale average across countries was set to 500 and the standard deviation to 100. Since the countries varied in size, each country was weighted to contribute equally to the mean and standard deviation of the scale. The average and standard deviation of the scale scores are arbitrary and do not affect scale interpretation.

In the PIRLS analysis, achievement scales were produced for each of the two reading purposes, reading for literary experience and reading for information, as well as for reading overall.

1.14   Data Analysis and Reporting

The PIRLS 2001 International Report (Mullis, Martin, Gonzalez & Kennedy, 2003) summarizes fourth-grade students' student reading achievement in each country. This report presents average student achievement in reading overall as well as iTw(r)Tj(.Ml as f)19.7(o)[(ng f)19.7(o

ther, the PIRLS international report presents examples of questions from both literary and informational passages that anchor at each of the benchmarks (providing another perspective on the reading demands of the benchmarks), and also displays student performance in each country on the example questions.

PIRLS 2001 collected a wide array of information about the home and school context in which students learned to read (from parents, students, teachers, and school principals). The PIRLS international report summarized much of this information, combining data into composite indices showing an association with achievement where appropriate. In particular, student reading achievement is described in relation to literacy-related activities in the home, the school curriculum and organization for teaching reading, teachers and reading instruction, school contexts, and students' reading attitudes, self-concepts, and out-of-school activities.

Additional information about the countries participating in PIRLS 2001 may be found in the

as to achieve a representative sample of students. The IRT scaling methodology used with PIRLS 2001 also was applied in the trends in reading literacy study. The results of the trend study are presented in Martin, Mullis, Gonzalez, and Kennedy (2003).

References

Campbell, J.R., Kelly, D.L., Mullis, I.V.S., Martin, M.O., & Sainsbury, M. (2001). *F* *ı* *2001. (2* *.)*. Chestnut Hill, MA: Boston College.

Elley, W.B. (1992). *?* The Hague, Netherlands: IEA.

Elley, W.B. (Ed.). (1994). *EA* *ℰ* *-* *: A* *-* . Oxford, England: Elsevier Science Ltd.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., & Kennedy, A.M. (2003). *´* *ℰ* *1991-2001: EA´* *1991* *ℰı* . Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Kennedy, A.M. (2003). *ı 2001* *: EA´* *ℰ* *35* . Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M.., Kenn(and)Tjing lit-