

Third International Mathematics and Science Study

Quality Assurance in Data Collection

Edited by

Michael O. Martin
Ina V.S. Mullis

with contributors

Michael Bruneforth
Pierre Foy
Kelvin D. Gregory
Kathleen Haley
Craig D. Hoyle
Heiko Jungclaus
Dana L. Kelly
Teresa A. Smith

1996 International Association for the Evaluation of Educational Achievement (IEA).

Third international mathematics and science study: quality assurance in data collection/edited by Michael O. Martin, Ina V.S. Mullis
Publisher: Center for the Study of Testing, Evaluation, and Educational Policy,
Boston College.

Library of Congress Catalog Card Number: 96-71249

ISBN 1-889938-01-7

For more information about TIMSS contact:

TIMSS International Study Center
Center for the Study of Testing, Evaluation, and Educational Policy
Campion Hall
School of Education
Boston College
Chestnut Hill, MA 02167
United States

This report also is available on the World Wide Web:
<http://wwwcsteep.bc.edu/timss>

Boston College is an equal opportunity/affirmative action employer.
Funding for the international coordination of TIMSS is provided by the U.S.
National Center for Education Statistics, the U.S. National Science Foundation, the
IEA, and the Canadian government. Each participating country provides funding
for the national implementation of TIMSS.

Printed and bound in the United States.

INTRODUCTION

The Third International Mathematics and Science Study (TIMSS), a comparative study of student achievement in mathematics and science, is a huge, complex project involving 45 countries, three student populations incorporating five grade levels, and over half a million students. Although the study is directed from the International Study Center¹ at Boston College, each participating country was responsible for implementing the design in that country in accordance with the international standards. Survey instruments and field procedures were developed through a process of cooperation and consensus among the participants, and fieldwork was carried out by the National Research Coordinator (NRC) in each country. Each participating country or educational system was responsible for translating the instruments and procedures and adapting them to local conditions, drawing the school and student samples, and implementing the data collection plan. This combination of international cooperation and national implementation is an efficient and cost-effective approach to conducting international comparative studies, but it requires close collaboration among participants, and the validity and reliability of the results are crucially dependent on each participant adhering to the prescribed procedures at all times.

TIMSS has expended considerable effort in developing standardized materials and procedures so that the data collected in all countries are comparable to the greatest possible extent. Martin, Mullis, and Kelly (1996) have documented these efforts, which include the provision of extensive technical documentation, translation verification procedures, training seminars, individual consultation, and computer software. It is important not only that the TIMSS data be of high quality but that the project be able to demonstrate the quality of the data to readers and users of the TIMSS reports and data. Conscious of this and of criticisms that have been made of some procedures used in previous international studies, the main funders of the International Study Center, the U.S. National Center for Educational Statistics and the U.S. National Science Foundation, provided additional funds for a quality assurance program to document the quality of the TIMSS data in several areas that have been subject to criticism in past studies.

The quality assurance program specified a range of activities spread over a three-year period. The aim was to help ensure the comparability of results across participating countries, and to provide documentation to assist in the interpretation of the data. The main activities are described in this report. They were designed to enhance and document the quality of the TIMSS data, with particular emphasis on instrument translation and adaptation, sampling response rates, test administration and data collection, the reliability of the coding process, and the integrity of the database. The chapters dealing with the preparations for data collection, the data collection itself, and the checking and processing

¹ The study was coordinated from its inception until August 1993 by the International Coordinating Center (ICC) at the University of British Columbia, Vancouver, Canada. From August 1993, the study was directed from the International Study Center at Boston College.

of the data (Chapters 1, 3, 4, and 6) refer to all three student populations. However, those chapters that report results from the data (Chapter 2 on sampling and Chapter 5 on the reliability of the coding process) are restricted to Population 2,² which was compulsory for all participants.

In any comparative study of student achievement that takes place in more than one language there is a risk that the difficulty of the tests (which are usually constructed in one language) may be affected by translation into other languages. With the administration of tests and questionnaires in 31 languages, this was an issue of great concern for TIMSS. Maxwell (1996) describes the procedures that were developed to assist NRCs in producing high-quality translations, and to monitor the quality of the translation process. In Chapter 1 of this report, Mullis, Kelly, and Haley review the translation verification procedures, and report on the status of the translation effort.

International comparative studies like TIMSS, which seek to make inferences about national populations on the basis of sample survey methodology, rely on the quality of the national samples for the validity of those inferences. In TIMSS great attention was paid to all aspects of the population sampling process, from population definition through sample design and selection to computation of participation rates, sampling weights, and estimates of sampling variance. Foy, Rust, and Schleicher (1996) describe the sampling design in detail. In Chapter 2 of the present report, Foy, Martin, and Kelly document participants' compliance with prescribed procedures at each stage of the sampling process, and present data on population coverage and participation for each participant.

The TIMSS achievement tests were designed to be administered under uniform conditions throughout all participating countries. Documenting the uniformity of the test administration required that a sample of testing sessions in each country be observed. In order to visit schools and carry out such observations it was necessary to hire and train a quality control monitor for each country. Monitors had to be fluent both in English (the language of the training and monitoring materials) and in the language of the country. They had two major tasks. The first was to visit the TIMSS national center to interview the NRC about all aspects of the data collection, including sampling, instrument translation, production and shipping, and plans for receipt control, free-response coding, and data entry. The International Study Center prepared data collection instruments and a manual to be used by the quality control monitors, and organized regional training meetings to ensure that the monitors were well versed in all of their responsibilities. Martin, Hoyle, and Gregory, in Chapter 3, describe this activity in detail, including the development of data collection instruments for the visits and the design and implementation of a training program for the quality control monitors. This chapter also includes a summary of the results of the interview with NRCs.

² Population 2 is defined as the two adjacent grades with the largest proportion of 13-year-old students at the time of testing.

The second task of the quality control monitor was to visit a random selection of schools from the TIMSS sample at the time testing was taking place and determine whether the tests were being administered using uniform and secure procedures. Each quality control monitor was required to visit ten schools in the TIMSS sample, to observe a testing session, and to interview the school coordinator regarding the implementation of the TIMSS procedures. These school visits and test session observations are a central component of the TIMSS quality assurance effort. The quality control monitor completed one classroom observation record for each visit. Martin, Hoyle, and Gregory summarize the results of these observations in Chapter 4, and present the results in detail in an accompanying appendix.

The TIMSS achievement tests included both multiple-choice and free-response (open-ended) items. Many of the free-response items required an extended response from students, and all of them required that the student responses be coded by trained coders prior to data entry. Detailed coding rubrics with example codes were provided for each item, and regional training meetings were organized by the International Study Center to ensure that each participant had a full understanding of the application of the rubrics. To monitor the reliability of the coding process in each country, each participant was required to select a 10% random sample of student responses and code them twice, using different coders on each occasion. Reliability coefficients were computed for each item in every country that complied with this requirement. In order to provide an indication of the consistency of coding across countries, English-speaking coders from 21 countries came to a central location and coded samples of student responses from seven English-speaking countries. In Chapter 5, Mullis and Smith present the results of the reliability studies both within and across countries, and discuss their significance for the quality of the TIMSS data.

Accurate and reliable comparisons of international achievement require accurate and complete datasets from participating countries. Although each participant was responsible for coding, entering, and checking that country's data, and for ensuring that all data were in the prescribed international format, the IEA Data Processing Center (DPC) in Hamburg, Germany, was charged with verifying that participants had complied with the international standard. The scale and complexity of the TIMSS tests and questionnaires required an enormous data verification exercise. The quality assurance program provided support for the staff of the DPC as they engaged in an extensive series of quality control checks and communicated with each country regarding the nature and extent of the national deviations from prescribed international procedures. In Chapter 6, Jungclaus and Bruneforth describe the procedures used to verify the data and the actions taken to remedy any deviations.

The activities described in this report should provide assurance to readers of TIMSS publications and users of TIMSS data that the highest professional standards were applied in all phases of the data collection, and that a very high standard was attained in all stages of the endeavor.

REFERENCES

- Martin, M.O., Mullis, I.V.S., and Kelly, D.L. (1996). "Quality Assurance Procedures" in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.
- Maxwell, B. (1996). "Translation and Cultural Adaptation of the Survey Instruments" in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.
- Foy, P., Rust, K., and Schleicher, A. (1996). "Sample Design" in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.