

Martin, M.O., Hoyle, C.D., and Gregory, K.D. (1996). "Observing the TIMSS Test Administration" in M.O. Martin and I.V.S. Mullis *Third International Mathematics and Science Study: Quality Assurance in Data Collection*. Chestnut Hill, MA: Boston College.

4. OBSERVING THE TIMSS TEST ADMINISTRATION4-1

Michael O. Martin, Craig D. Hoyle, and Kelvin D. Gregory

4.1	OVERVIEW.....	4-1
4.2	SCHOOL VISITS AND TEST SESSION OBSERVATIONS.....	4-4
4.3	SUMMARY.....	4-11

4. OBSERVING THE TIMSS TEST ADMINISTRATION

Michael O. Martin
 Craig D. Hoyle
 Kelvin D. Gregory

4.1 OVERVIEW

In order to monitor compliance with international procedures in the administration of the TIMSS tests, quality control monitors visited a sample of schools where they observed testing sessions and interviewed School Coordinators. Table 4.1 summarizes the number of test administrations observed in each country. The complete program (visit to a school, observation of the test administration, and an interview with the School Coordinator) was implemented in 37 countries. The program was implemented only partially or not at all in the remaining countries for a variety of reasons.

- Because of the timing of the funding of the quality assurance program, countries on the Southern Hemisphere timeline for Population 1 and 2 (Australia, New Zealand, Republic of Korea, and Singapore) had already completed their testing program and therefore test sessions could not be observed by the quality control monitor. Australia and New Zealand carried out the program with samples of Population 3 testing sessions.
- Several other countries were scheduled to have completed testing before a quality control monitor could visit for classroom observations (Iceland, Japan, Kuwait, and Thailand). Iceland did conduct classroom observations for Population 3, but the materials arrived at the International Study Center too late to be included in this report. Japan and Thailand conducted interviews with samples of School Coordinators after testing had taken place.

- Denmark and Ireland, both countries where school participation was voluntary and testing was conducted by classroom teachers, were unwilling to ask schools to take part in the program of classroom testing observations.
- In the United States, the quality control monitor became indisposed, and was unable to conduct the classroom observations. Information about the testing sessions in the selected schools was collected at a later date from Test Administrators and School Coordinators.
- Germany was unable to nominate a quality control monitor in time to observe testing sessions.

Table 4.1
Classroom Observation Records

Country	Number of Observation Records	Country	Number of Observation Records
Argentina	10	Japan ⁴	6
Australia	8	Korea ³	-
Austria	10	Kuwait ³	-
Belgium (Fl)	10	Latvia	7
Belgium (Fr)	10	Lithuania	10
Bulgaria	10	Mexico	10
Canada (Alberta)	10	The Netherlands	10
Canada (Ontario)	10	New Zealand	9
Columbia	10	Norway	10
Cyprus	10	Philippines	10
Czech Republic	10	Portugal	10
Denmark ¹	-	Romania	10
England	10	Russian Federation	10
France	11	Scotland	10
Germany ²	-	Singapore ³	
Greece	10	Slovakia	10
Hong Kong	7	Slovenia	10
Hungary	10	South Africa	9
Iceland ²	-	Spain	9
Indonesia	11	Sweden	10
Iran	10	Switzerland	10
Ireland ¹	-	Thailand ⁴	5
Israel	10	United States ⁵	12
Italy	10		
Total = 384			

¹Unwilling to take part in classroom observations.

²Unable to nominate a quality control monitor to observe testing.

³Tests conducted before quality control monitoring programs were in place.

⁴Unable to conduct observations but did conduct interviews with School Coordinators.

⁵Unable to conduct observations but did conduct interviews with Test Administrators.

During each visit to a school, the quality control monitor documented the information he or she collected in a questionnaire called the Classroom Observation Record. This had four sections:

- Activities preliminary to the testing session, including preparation, test security, arranging accommodation

- Observation of the testing session
- Quality control monitor's general impressions of test administration
- Interview with the School Coordinator.

This chapter provides a general summary of the results of the school visits as reported by the quality control monitors. Detailed results are presented in Appendix E. The letter/number codes displayed after the headings in the commentary below correspond to the instrument questions and results provided in the appendix.

4.2 SCHOOL VISITS AND TEST SESSION OBSERVATIONS

4.2.1 SECTION A : PRELIMINARY ACTIVITIES OF THE TEST ADMINISTRATOR

Having become acquainted with the Test Administrator and located the room where the testing session would take place, the quality control monitor was to record observations on the condition of the testing materials, the Test Administrator's level of preparation, and the suitability of the testing room.

Overall, the quality control monitors reported very favorably upon the preliminary activities conducted by the Test Administrators. With very few exceptions, test conditions, booklets, and directions were in accordance with the study procedures. Where changes were made, they tended to be minor in nature and posed no threat the validity of the study.

- *Verification of the supply of test books (A.1).* Almost all Test Administrators (94%) had definitely (77%) or probably (17%) verified adequate supplies of test books prior to the test administration.
- *Seals on test books (A.2).* In every session where the national center had used booklet seals to enhance booklet security (128 of the 384 sessions observed), seals were intact before the testing session began.
- *Test booklet and Student Tracking Form correspondence (A.4).* Student identification on the test booklets and questionnaires corresponded to the information on the Student Tracking Form in almost every session (96%). In the remaining sessions, there were either minor problems involving only a few students or changes in procedure such as filling out the Student Tracking Form after the tests were distributed.
- *Correct version of the administration script (A.6).* In practically all sessions (99%), the Test Administrator had the correct version of the administration script for the session.
- *Familiarization with script (A.7).* Most Test Administrators (93%) had definitely (74%) or probably (19%) familiarized themselves with the administration script before the testing session.
- *Space for students to work (A.8).* In most sessions (93%), there was adequate seating space for the students to work without distractions. Comments from quality control monitors indicated that seating arrangements varied considerably, but that, for the most part, seating arrangement was not a problem. In a few sessions, however, it is clear that seating arrangements were less than ideal.

- *Adequate room for the Test Administrator to circulate (A.10).* In most of the sessions (97%), the Test Administrator had adequate room to move about during the testing session and ensure that students were following directions correctly.
- *Keeping track of time (A.12, A.14).* Most Test Administrators (90%) had a stop-watch or timer for accurately timing the testing sessions. Some of those who did not may, however, have had an ordinary watch. The presence of a wall clock for students to keep track of time was more the exception than the rule. Quality control monitors reported the presence of wall clocks in only 26% of the sessions. No data were collected, however, on the number of students with their own watches.
- *Supply of pencils (A.13).* The Test Administrator had an adequate supply of pencils and other necessary materials ready for the students in 77% of sessions. However, in many countries it is the responsibility of the student to bring pencils, pens, etc., to testing situations, so it is not necessarily the case that there was an inadequate supply in almost a quarter of the testing sessions.

4.2.2 SECTION B: OBSERVING THE TESTING SESSION

Following the preliminary activities, the quality control monitor was required to observe the testing session, and record the activities of the Test Administrator throughout the session. In Population 1 and 2 schools, a test administration consisted of two testing sessions, Session 1 and Session 2, separated by a short break, followed by a session for the student questionnaire after a second break. In Population 3 schools, there was a single testing session, followed after a break by a session for the student questionnaire.

In many cases, quality control monitors reported no changes in the script. Where changes were observed, they tended to be local adaptations. The testing sessions were orderly and well conducted. The time allowed for testing was generous, and few quality control monitors reported tests extending past the scripted time. Approximately half of the quality control monitors reported that more time was required to complete the student questionnaire.

SESSION 1 (ALL POPULATIONS)

- *Following the Administrator's Script (B.2a, B.2b, B.2c).* Most Test Administrators (95%) followed the instructions for preparing students in the administrator's script without making any changes (64%) or making only minor changes (31%). Most (97%) also followed the script with regard to distributing the materials, making no changes (83%) or minor changes (14%). Likewise, a very high percentage (97%) had made either no changes (82%) or only minor changes (15%) in the instructions to begin testing.
- *Changes to the script that were made (B.3a).* Approximately half of the Classroom Observation Records indicated that no changes were made in the script. Where changes were made, the Test Administrator essentially adapted the script in a manner pertinent to the students. No quality control monitor reported deviations that might be expected to affect the interpretation of test results.
- *Distribution of the test books (B.4).* In almost every session (97%), test booklets were distributed one at a time, as prescribed in the manual. In those few sessions where the procedure was not followed, some other acceptable method (in the judgment of the quality control monitors) of distribution was used.

- *Allocation of test books (B.6, B.7).* In most sessions (93%), test booklets were distributed according to the booklet assignments on the Student Tracking Form. In some of the sessions where they were not, it was because the Student Tracking Form was not available at the time of the session. Other cases involved one or two students who were issued spare booklets or whose booklet ID did not match the Student Tracking Form. In each of these cases, the actions of the Test Administrator were recorded on the Student Tracking Form.
- *Attendance (B.8).* The Test Administrator recorded attendance correctly on the Student Tracking Form in almost all sessions (98%).
- *Testing time (B.9, B.10., B.11).* In most test administration sessions (86%), the appropriate amount of testing time was allocated to Session 1. In many sessions, however, all of the students finished the test before the prescribed time had elapsed, and so the test administrator brought the session to an early conclusion.
- *Time announcements (B.12, B.13).* In most sessions (86%), the Test Administrator announced “you have 10 minutes left” prior to the end of Session 1, as instructed. Quality control monitors indicated that in 24% of the sessions other announcements regarding the time remaining were made during Session 1.
- *Instructions to stop work (B.15).* In almost every session (98%), students complied very well (77%) or well (22%) with the instructions to stop work.
- *Collection of the booklet at end of Session 1 (B.16).* Many Test Administrators chose not to collect the test booklets at the end of Session 1. Only 51% followed the prescribed procedure of collecting test booklets one-at-a-time from each student. In many cases, test booklets were left on the students’ desks between sessions. Sometimes this was necessary as the test booklets and questionnaires were in a single packet. None of the quality control monitors recorded any observations that would indicate that test integrity had been compromised.

SESSIONS 2 AND 3 (POPULATIONS 1 AND 2)

- *Break between Session 1 and Session 2 (B.19, B.20, B.21).* The recommended break time between testing sessions was 20 minutes, although in the majority of sessions (81%) some other interval was found to be more convenient. Frequently, breaks coincided with lunch or recess periods. In some instances, there was no break between sessions; in others, the break time was substantially shortened. In most sessions (84%), however, despite changes in its length, the break was conducted exactly (56%) or almost (28%) as prescribed.
- *Session 2 restart time (B.22, B.23, B.24).* Most sessions (54%) required less time than the prescribed five minutes to re-read instructions and settle students at the beginning of Session 2. Explanations for the deviation from the scripted 5 minutes included “the students had no questions,” and “students embarked immediately on the second part of the test.”
- *Testing time session 2 (B.25, B.26, B.27).* For most sessions (79%), the testing time for Session 2 was as prescribed in the administrator’s script. As with Session 1, Test Administrators sometimes brought the session to an early close if all students finished before the prescribed time had elapsed.
- *Time remaining announcement (B.28, B.29, B.30).* Generally the Test Administrators (81%) announced “you have 10 minutes left” prior to the end of Session 2, as prescribed in the manual. In most instances where this announcement was not made, an acceptable procedure was substituted. Such a substitution was made in about 20% of the sessions.

In several sessions, administrators kept track of time by marking off intervals on a blackboard.

- *Ending Session 2 (B.31).* In almost all sessions (97%), students complied with the instruction to stop work either very well (81%) or well (16%).
- *Collection of test booklets after Session 2 (B.32).* In most cases (81%), the Test Administrator collected the test books one at a time at the end of Session 2. Where the books were not collected as scripted, the test administrator used an alternative method that did not compromise the integrity of the test administration.
- *Announce break before student questionnaire (B.34).* In two-thirds of the sessions, the Test Administrators announced a break at the end of the testing session, to be followed by the student questionnaire. In many of the remaining sessions, the administration of the student questionnaire followed without a break. Sometimes the student questionnaire did not follow the testing sessions immediately but was completed on another occasion.
- *Read script for break (B.36, B.37).* In most of the sessions (89%), the Test Administrators followed the script to end the testing and signaled a break either verbatim (63%) or with minor changes (26%). Minor changes in the script were noted in 44 sessions, with additions in 24 sessions and omissions in 38 sessions (some sessions had both additions and omissions).
- *Break conducted (B.38, B.39, B.40, B.41).* Most (82%) of the Test Administrators held the break as directed in the manual (68% exactly as directed; 14% nearly as directed).
- *Distribution of the student questionnaire (B.43).* The majority (67%) of Test Administrators distributed the student questionnaire and gave directions as specified in the script. In many countries, the student questionnaire was distributed in a packet with the test booklets, not separately. There was no indication of any problems with the distribution of the student questionnaires.
- *Time allocated to Student Questionnaires (B.46, B.47, B.48, B.49).* In more than half the sessions (60%), the student questionnaire required more time than was prescribed in the administration script. The Test Administrator Manual made provision for more time for the questionnaire as necessary. Extra time allowed ranged from 1 to 45 minutes, with a median of 20 minutes.
- *End of session (B.50, B.51).* In 80% of the observed sessions, the Test Administrator thanked students for participating in the study. Dismissal of students was generally an orderly affair. Quality control monitors described 94% of the session dismissals as either very orderly (62%) or somewhat orderly (32%).

SESSION 2 (POPULATION 3)

- *Break announcement after testing session (B.69).* Test Administrators announced a break at the end of 64% of the testing sessions, to be followed by the student questionnaire. In many of the remaining sessions the administration of the student questionnaire followed without a break. Occasionally the student questionnaire was completed at another time.
- *Script (B.71, B.72).* For most sessions (91%), Test Administrators followed the script to end the testing and signal a break either verbatim (52%) or with minor changes (39%). Minor changes in the script were noted in 12 sessions, with additions in 2 sessions and omissions in 9 sessions (some sessions had both additions and omissions).
- *Break time (B.73, B.74, B.75, B.76, B.77).* The break time differed from the time recommended in the script in 42% of the questionnaire sessions. Most of these (33% of all sessions) involved a shorter break time. Most (77%) of the Test Administrators

conducted the break as directed (66% exactly as directed; 11% nearly as directed). The most common reason given for not including a break was that the country's Test Administrator Manual did not provide for one.

- *Distribution of Student Questionnaires (B.78, B.79).* Test Administrators distributed the Population 3 student questionnaire and gave directions as specified in the script in 71% of sessions. As in Populations 1 and 2, in many countries the student questionnaire was distributed in a packet with the test booklets. There was no indication of any problems with the distribution of the student questionnaires.
- *Time allocated to Population 3 student questionnaires (B.80, B.81, B.82, B.83, B.84).* There was considerable variation in the amount of time required to complete the Population 3 student questionnaire. In 40% of sessions, quality control monitors reported that the time allowed was less than the prescribed amount, whereas in 23% of the sessions, additional time was requested.
- *Dismissal of Population 3 students (B.85, B.86).* Most (82%) of the Test Administrators thanked students for participating at the end of the study. Dismissal of students was described as “very orderly” (75%) or “somewhat orderly” (20%).

4.2.3 SECTION C: SUMMARY OBSERVATIONS OF THE QUALITY CONTROL MONITORS

Following observation of the testing session, quality control monitors were asked to give their impressions of several aspects of the test administration, including the behavior of the students and the activities of the Test Administrator.

With few exceptions, quality control monitors commented favorably on test administrations. They stated that the Test Administrator conducted the test sessions in a well organized and professional manner. They found that students were well motivated and challenged by the test items. Where the quality control monitors noted deviations from the administration script, these deviations posed little if any threat to the validity of the results. Rather, the changes mostly represented acceptable adaptations in the test administration.

- *Student conduct (C.1, C.2).* In 94% of the sessions, students were described as either extremely (65%) or moderately (29%) orderly and cooperative. In the rare situations where students were not cooperative, quality control monitors indicated that the Test Administrator almost always made some effort to exert control.
- *Supervision by the Test Administrator (C.3).* Test Administrators walked around the room to monitor student behavior in 94% of the observed sessions. Where this did not occur, it was often because of lack of space.
- *Student Questions (C.5).* Test Administrators addressed students' questions appropriately in almost all sessions (97%).
- *Evidence of cheating (C.7).* In most sessions (87%), there was no evidence of students attempting to cheat. Where evidence was reported, it was usually that students attempted to talk to their neighbors or attempted to copy from a neighbor's booklet. Because eight different booklet versions were in use in a classroom, copying responses from a neighbor's booklet was unlikely to help a student's performance.

- *Defective booklets (C.9, C.10, C.11)*. In 6% of the observed sessions, defective test booklets were identified and replaced before the session began. In a further 6% of sessions, defective booklets were found and replaced after the sessions began. On most of the occasions where booklets needed replacement, the Test Administrator replaced them appropriately. Occasionally booklets were not replaced because of a lack of spare copies.
- *Late students (C.13)*. In most sessions (88%), no late students were reported. In 2% of sessions, late students were not admitted; in 5% of sessions, late students were admitted before the testing began; and in 5% of sessions, they were admitted after testing had begun.
- *Refusals to take the test (C.14, C.15, C.16)*. In just 3% of sessions did students refuse to take the test, and then usually just one student. In only one session was it reported that more than a few students refused to take the test. This case is described in more detail below. Test Administrators accurately followed the instructions for excusing students in 5 of the 9 sessions where this was necessary. In the single case where more than five students were excused, the entire class refused to take the test. The quality control monitor noted that the students were all of low ability and wanted to give up because the test was too difficult for them, and that the Test Administrator had persuaded them to attempt the test. In addition, the monitor noted that the school was for students who could not gain entry to "good schools."
- *Emergency during testing (C.17, C.18)*. In 15% of sessions, at least one student left the room during testing because of an "emergency." Usually the "emergency" was merely a need to visit the bathroom.
- *Overall quality of the test administration session (C.19, C.20)*. The overall quality of the testing sessions was rated high, with 94% of sessions rated "good" or better. The Test Administrator usually was praised, as were the students. It was commonly observed that students were well disciplined, motivated, and challenged by the test. In many cases, quality control monitors noted that the Test Administrator had conducted the test in a well organized, professional manner. Critical comments by monitors focused predominantly upon the time allocated to the test and the language used in the test, and not upon the actual test administration.

4.2.4 SECTION D: INTERVIEW WITH THE SCHOOL COORDINATOR

Following the completion of the testing in the school, quality control monitors were asked to interview the School Coordinator to collect information on experiences with the test administration, attitudes and reactions of school staff, and suggestions for improvements for the future.

The comments of the School Coordinators tended to be very positive. They were happy with the shipping of TIMSS materials, and overwhelmingly made positive comments regarding the National Research Coordinators. Negative criticisms centered mainly upon the teacher questionnaire, the mismatch between test items and curriculum, and the timing of the testing program.

- *Overall impression (D.1)*. School Coordinators almost unanimously (99%) indicated that the testing sessions went well (70% very well).

- *Attitude of other school staff (D.2)*. Most School Coordinators (71%) rated the attitude of other school staff members towards the TIMSS testing as positive. Negative attitudes (4%) were predominantly attributed to the date of testing, which caused disruptions in the regular schedule. A further 25% of the School Coordinators rated the attitude of other school staff as neutral to the TIMSS testing.
- *Checking materials (D.3)*. Most School Coordinators (87%) found time to check the shipment of materials from the National Research Coordinator prior to the day of testing.
- *Items received (D.5)*. In most cases, School Coordinators reported receiving the correct shipment of test booklets (99%), Test Administrator Manuals (100%), School Coordinator Manual (98%), Student Tracking Forms (93%), Student Questionnaires (98%), Teacher Questionnaires (91%), School Questionnaires (99%), and Test Administration Forms (92%). Teacher Tracking Forms (70%), Student-Teacher Linkage Forms (26%), and envelopes or boxes for the purpose of returning the materials after the assessment (71%) were less frequently reported to be correct, but in some instances these items may have been purposely omitted by the NRC.
- *Responsiveness of National Research Coordinators (D.6)*. School Coordinators felt that the National Research Coordinator was responsive to questions and concerns in most (93%) cases.
- *Collecting teacher questionnaires (D.7, D.8)*. In many schools (60%), it was not possible for the School Coordinator to collect completed teacher questionnaires before the test administration. These usually were completed during or after the test administration, with several observations noting that Test Administrators were unaware that the questionnaire was to be collected before the test administration. Many of the teachers (60%) commented that the questionnaire took more time than expected to complete.
- *Satisfaction with testing room (D.11)*. Most School Coordinators (96%) were satisfied with the testing room that they were able to arrange for the testing session.
- *Make-up-sessions (D.12, D.13)*. Most School Coordinators (84%) anticipated that make-up sessions would not be required at their school. Most (93%) of those who anticipated the need for make-up sessions planned to conduct one.
- *Selection and training of Test Administrators (D.14)*. School Coordinators predominantly made positive comments regarding the training of Test Administrators. The *Test Administrator Manual* was generally found to be very useful. In some cases, they suggested improvements such as adding flow diagrams to the manual, and a more extensive training period.
- *Motivational talk (D.16)*. Almost half of the School Coordinators (46%) reported that students received special instructions, motivational talks, or incentives to prepare them for the assessment. Most of these consisted of introductory presentations by the school principal, class teacher, or other test administrator.
- *Practice questions (D.18)*. Only 2% of the School Coordinators reported that students were given an opportunity to practice on questions like those in the tests before the testing session.
- *School Coordinator Manual (D.20)*. The majority of School Coordinators (92%) believed that the *School Coordinator Manual* worked well.
- *Completeness of class lists (D.23)*. Most (93%) of the School Coordinators reported that the classes listed on the Class Tracking Form for the school represented a complete list of the mathematics classes in that school at those grades.

- *Students not in math classes (D.25).* School Coordinators almost universally (98%) reported that to the best of their knowledge there were no students in their schools at the required grade levels who were not in any of the mathematics classes listed on the Class Tracking Form.
- *Students in more than one math class (D.27).* Most School Coordinators (96%) also believed that there were no students in the required grade levels who were in more than one mathematics class.
- *Willingness to serve again (D.29).* Most of the School Coordinators (90%) indicated that if there were another international assessment, they would be willing to serve as School Coordinators again.

4.3 SUMMARY

In order to monitor compliance with international procedures in the administration of the TIMSS achievement tests, the International Study Center dispatched a quality control monitor to each country to visit a sample of schools where they observed a testing session and interviewed the School Coordinator. Test administrations were observed and School Coordinators interviewed in 37 countries, and interviews were conducted with School Coordinators or Test Administrators in three further countries.

The Classroom Observation Record completed by the quality control monitor for each school visit had four sections:

- Activities preliminary to the testing session
- Observation of the testing session
- Quality control monitor's general impressions of the test administration
- Interview with the school coordinator.

In general, quality control monitors reported very favorably on the test administration effort. Test Administrators were well prepared, and, with few exceptions, test conditions, instruments, and directions were in accordance with prescribed procedures. Test administrations were reported to be orderly and well conducted. The time allowed for testing was found to be generous, with very few reports of students needing more time. With very few exceptions, quality control monitors commented favorably on the test administrations. Generally, they reported that Test Administrators were well organized and performed their duties in a professional manner, and that students were orderly and applied themselves to their tasks. School Coordinators also tended to be very positive in their remarks. Despite the disruption to school schedules, school staff were generally reported to have favorable attitudes towards the project. The burden of completing the teacher questionnaire drew adverse comment from quite a few teachers.

On the evidence provided by the quality control monitors from their school visits the TIMSS test administration was generally a very successful endeavor. Readers and reviewers can be assured that the TIMSS data were collected following standard procedures and under standard conditions to the greatest extent possible.