



Chapter 7

Quality Assurance in the TIMSS 2003 Data Collection

Eugenio J. Gonzalez and Dana Diaconu

7.1 Overview

As part of its overall quality assurance efforts, TIMSS conducted an ambitious program of site visits to document the quality of the TIMSS 2003 data collection. Together with the IEA Secretariat and the national centers, the TIMSS & PIRLS International Study Center (ISC) identified and appointed one International Quality Control Monitor (QCM) in each country to observe data collection procedures at both national and classroom levels.

Quality Control Monitors had two major responsibilities: to interview the National Research Coordinator (NRC) about the survey operations and activities, and to conduct site visits to a random sample of 15 schools in the country at each grade assessed during test administration. The QCMs attended a two-day training session conducted by the ISC and the IEA Secretariat,¹ where they were introduced to the TIMSS 2003 survey operations procedures and instructed on how to conduct their interviews and site visit observations. At the training session, QCMs received a copy of the *TIMSS 2003 Manual for International Quality Control Monitors* (TIMSS, 2002a), which explained their duties in detail, and copies of the *Survey Operations Manual* (TIMSS, 2002b), *School Coordinator Manual* (TIMSS, 2002c), and *Test Administrator Manual* (TIMSS, 2002d).

Fifty QCMs were trained across the 49 countries and four Benchmarking participants where the international quality control program was conducted.² Where necessary, QCMs who attended the training session were permitted to recruit other QCMs to assist them in covering the territory and

1 Two training sessions were conducted, one for countries in the southern hemisphere and the other for northern hemisphere countries.

2 Iran and Israel were the only countries whose QCMs were not trained; Ontario and Quebec shared the same QCM.

meeting the testing timetable. All together, these monitors and those trained by them observed 1147 testing sessions (755 for grade 8 and 392 for grade 4),³ and conducted interviews with the National Research Coordinator in each of the participating countries. Exhibit 7.1 indicates the dates of data collection and the number of site visits by QCMs in each country.

7.2 Observing the TIMSS Test Administration

When visiting the school, the QCM had to complete a Classroom Observation Record Form. This form was organized into four sections as follows:

- Preliminary activities of the Test Administrator
- Test session activities
- Summary observations
- Interview with the School Coordinator

7.2.1 Preliminary Activities of the Test Administrator

Section A of the Classroom Observation Record addressed the extent to which the Test Administrator had prepared for the testing session. Monitors were asked to note the following activities of the Test Administrator: checking the testing materials, reading the administration script, organizing space for the session, and arranging for the necessary equipment (e.g., pencils, a watch for timing the testing session).

Exhibit 7.2 summarizes the results for Section A for the eighth grade. In almost all testing sessions, Test Administrators observed the proper preparatory procedures. When deviations occurred, the QCMs provided reasonable explanations for the discrepancies. For example, QCMs noted that the main reason for students receiving booklets with student identifications that did not correspond to the Student Tracking Form was because new students did not appear on the list, as the tracking forms had been created before the students were enrolled. In the few cases where there reportedly was not enough room for students, QCMs indicated that it was due to unavoidable circumstances (e.g., the test was administered in a small classroom, students had to sit two or three at one desk or in groups of five or six around a table).

³ Operational constraints did not permit QCM visits to be conducted in five testing sessions in Japan.

Exhibit 7.1 TIMSS 2003 International Quality Control Site Visits

Countries	Eighth Grade		Fourth Grade	
	Date of Data Collection	Number of Site Visits	Date of Data Collection	Number of Site Visits
Argentina	Nov. 2003	16		
Armenia	May 2003	15	May 2003	15
Australia	Oct. - Nov. 2002	15	Nov. 2002	15
Bahrain	Apr.-May 2003	15		
Belgium (Flemish)	May 2003	15	May 2003	15
Botswana	Oct. - Nov. 2002	15		
Bulgaria	Apr.-May 2003	15		
Chile	Nov. 2002	19		
Chinese Taipei	May 2003	15	June 2003	15
Cyprus	May 2003	15	May 2003	14
Egypt	May 2003	15		
England	June 2003	15	May 2003	15
Estonia	June 2003	15		
Ghana	Apr.-May 2003	14		
Hong Kong, SAR	May 2003	15	May - June 2003	15
Hungary	March 2003	15	March – Apr. 2003	15
Indonesia	May 2003	15		
Iran, Islamic Rep. of	Apr.-May 2003	15	Apr.-May 2003	15
Israel	May 2003	15		
Italy	Apr.-May 2003	16	Apr.-May 2003	14
Japan	Feb. 2003	10	Feb. 2003	11
Jordan	May 2003	15		
Korea, Rep. of	Apr. 2003	15		
Latvia	May 2003	15	May 2003	15
Lebanon	Apr. 2003	15		
Lithuania	May 2003	15	May 2003	15
Macedonia, Rep. of	May 2003	15		
Malaysia	Oct. 2002	15		
Moldova, Rep. of	May 2003	15	May 2003	15

Exhibit 7.1 TIMSS 2003 International Quality Control Site Visits (...Continued)

Countries	Eighth Grade		Fourth Grade	
	Date of Data Collection	Number of Site Visits	Date of Data Collection	Number of Site Visits
Morocco	June 2003	15	May 2003	15
Netherlands	Apr.-May 2003	13	Apr. 2003	14
New Zealand	Nov. 2002	13	Nov. 2002	16
Norway	Apr. 2003	20	Apr. 2003	10
Palestinian Nat'l Auth.	Apr.-May 2003	15		
Philippines	March 2003	16	March 2003	14
Romania	May – June 2003	15		
Russian Federation	Apr.-May 2003	15	Apr.-May 2003	15
Saudi Arabia	May 2003	15		
Scotland	Apr.-May 2003	15	March – May 2003	15
Serbia	May 2003	15		
Singapore	Oct. 2002	15	Oct. 2002	15
Slovak Republic	May 2003	15		
Slovenia	Apr.-May 2003	15	May 2003	15
South Africa	Oct. 2002	15		
Sweden	May 2003	15		
Syria	May 2003	15		
Tunisia	May 2003	14		
United States	Apr.-May 2003	17	Apr.-May 2003	14
Yemen			May 2003	15
Benchmarking Participants				
Basque Country, Spain	May 2003	16		
Indiana State, US ⁴				
Ontario Province, Can.	Apr. 2003	15	Apr. 2003	15
Quebec Province, Can.				
TOTAL		755		392

4 Data collection for Indiana was conducted by Westat, Inc., using the same procedures that it applied in collecting the data for the United States' national sample for TIMSS 2003.

Exhibit 7.2 Preliminary Activities of the Test Administrator - Eighth Grade

Question	Yes	No	N/A
Had the Test Administrator verified adequate supplies of the test booklets?	729*	22**	4
Did the student identification information on the test booklets and student questionnaires correspond with the Student Tracking Form?	741	11	3
Had the Test Administrator familiarized himself or herself with the test administration script prior to the testing?	729*	21**	5
Was there adequate seating space for the students to work without distractions?	737	17	1
Was there adequate room for the Test Administrator to move about during the testing to ensure that students were following directions correctly?	738	17	0
Did the Test Administrator have a stopwatch or timer for accurately timing the testing session?	723	24	8
Did the Test Administrator have an adequate supply of pencils and other necessary materials ready for the students?	646	102	7

* Represents the number of respondents answering either "Definitely Yes" or "Probably Yes"

** Represents the number of respondents answering either "Definitely No" or "Probably No"

The absence of a stopwatch was not considered a serious limitation. Test Administrators who did not have a stopwatch had a wristwatch available to monitor the time remaining in the test sessions. In about 14 percent of the testing sessions, the QCMs noted that the Test Administrators did not have an adequate supply of pencils for the students. However, in most of these cases, students provided their own. In general, QCMs observed no procedural deviations in test preparations severe enough to jeopardize the integrity of the test administration.

Exhibit 7.3 summarizes the results for Section A for the fourth grade. Similar to the eighth grade, in almost all testing sessions Test Administrators observed the proper preparatory procedures, and when deviations occurred the QCMs provided reasonable explanations for the discrepancies. As at the eighth grade, QCMs observed no procedural deviations in test preparations severe enough to jeopardize the integrity of the test administration.

7.2.2 Test Session Activities

Section B of the Classroom Observation Record addressed the activities that took place during the actual testing session. These activities included following the Test Administrator script, distributing and collecting test booklets, and making announcements during the testing sessions.

Exhibit 7.3 Preliminary Activities of the Test Administrator - Fourth Grade

Question	Yes	No	N/A
Had the Test Administrator verified adequate supplies of the test booklets?	369*	16**	6
Did the student identification information on the test booklets and student questionnaires correspond with the Student Tracking Form?	378	7	6
Had the Test Administrator familiarized himself or herself with the test administration script prior to the testing?	378*	9**	4
Was there adequate seating space for the students to work without distractions?	378	8	5
Was there adequate room for the Test Administrator to move about during the testing to ensure that students were following directions correctly?	382	4	5
Did the Test Administrator have a stopwatch or timer for accurately timing the testing session?	371	13	7
Did the Test Administrator have an adequate supply of pencils and other necessary materials ready for the students?	342	40	9

* Represents the number of respondents answering either "Definitely Yes" or "Probably Yes"

** Represents the number of respondents answering either "Definitely No" or "Probably No"

The achievement test for grade in 8 was administered in two sessions of 45 minutes each, with a short break between. Exhibit 7.4 documents the activities associated with the first testing session and shows that at least 80 percent of the Test Administrators followed their script exactly when preparing the students, distributing the test materials, and beginning testing. In the rare instances when changes were made to the script, these tended to be additions or revisions, rather than deletions.

In only about five percent of the sessions visited, the total testing time for Session 1 was not equal to the time allowed. However, in most of these sessions, this was because all students had completed Session 1 before the allotted time had elapsed, and so the Test Administrator reasonably went on with the next part of the session according to the prescribed procedures. The average testing time for Session 1 was approximately 45 minutes, same as the allocated time.

Exhibit 7.4 also shows that only in about half of the sessions did the Test Administrator collect booklets one at a time at the end of the session, as prescribed in the directions. While this may seem surprising, it turns out that when the booklets were not collected individually from each student, students were instructed to close their test booklets and leave them on their desks during the break. The room was then either secured or supervised during the break.

When asked whether the break between sessions was 20 minutes long, QCMs tended to interpret the question quite literally. As a result, QCMs reported that only about half of classrooms started the test after a break that was “exactly” 20 minutes. The remainder reported having breaks that ranged from no break at all (with all students’ agreement) to about one hour.

The achievement test for grade 4 was administered in two sessions of 36 minutes each with a short break in between. Exhibit 7.5 documents the activities associated with the first testing session and shows that about three-quarters of the Test Administrators followed their script exactly when preparing the students, distributing the test materials, and beginning testing. As at grade 8, in the rare instances when changes were made to the script, these tended to be additions, rather than revisions or deletions.

In almost all of the sessions visited the total testing time for Session 1 corresponded to the time allowed. Where it did not, it was because all students had completed Session 1 before the allotted time had elapsed, and the Test Administrator went on with the next session. The average testing time for Session 1 was approximately 36 minutes, identical to the allocated time.

Mirroring grade 8, Exhibit 7.5 also shows that in less than half of the sessions the Test Administrator collected booklets one at a time at the end of the session, as prescribed in the directions. Again, when the booklets were not collected individually from each student, students were instructed to close their test booklets and leave them on their desk during the break. The room was then either secured or supervised during the break, in some instances by the QCM.

Similar to grade 8, when asked whether the break between sessions was 20 minutes long, QCMs tended to interpret the question quite literally. As a result, only 35 percent of Test Administrators reported that the test started after a break that was “exactly” 20 minutes. The total break time across all countries ranged from one to 50 minutes.

Exhibit 7.4 Testing Session 1 – Eighth Grade

Question	Yes	No	N/A
Did the Test Administrator follow the Test Administrator’s script exactly in each of the following tasks?			
Prepare the students	619	119 (minor changes) 11 (major changes)	6
Distribute the materials	661	70 (minor changes) 13 (major changes)	11
Begin testing	661	69 (minor changes) 12 (major changes)	13
If the Test Administrator made changes to the script, how would you describe them?			
Additions	103	243	409
Revisions	100	245	410
Deletions	58	256	441
Did the Test Administrator distribute test booklets one at a time to each student?	692	52	11
Did the Test Administrator distribute the test booklets according to the booklet assignments on the Student Tracking Form?	738	12	5
Did the Test Administrator record attendance correctly on the Student Tracking Form?	728	11	16
Did the total testing time for Session 1 equal the time allowed?	715	36	4
Did the Test Administrator announce “you have 10 minutes left” prior to the end of Session 1?	721	31	3
Were there any other time remaining announcements made during Session 1?	124	620	11
At the end of Session 1, did the Test Administrator collect the test booklets one at a time from each student?	406	341	8
Was the total time for the break between Session 1 and Session 2 equal to 20 minutes?	344	402	9

Exhibit 7.5 Testing Session 1 – Fourth Grade

Question	Yes	No	N/A
Did the Test Administrator follow the Test Administrator's script exactly in each of the following tasks?			
Prepare the students	299	81 (minor changes) 5 (major changes)	5
Distribute the materials	341	40 (minor changes) 2 (major changes)	8
Begin testing	325	54 (minor changes) 2 (major changes)	8
If the Test Administrator made changes to the script, how would you describe them?			
Additions	90	117	184
Revisions	44	147	200
Deletions	17	166	208
Did the Test Administrator distribute test booklets one at a time to each student?	383	3	5
Did the Test Administrator distribute the test booklets according to the booklet assignments on the Student Tracking Form?	383	3	5
Did the Test Administrator record attendance correctly on the Student Tracking Form?	373	9	9
Did the total testing time for Session 1 equal the time allowed?	375	8	8
Did the Test Administrator announce "you have 10 minutes left" prior to the end of Session 1?	374	11	6
Were there any other time remaining announcements made during Session 1?	52	330	9
At the end of Session 1, did the Test Administrator collect the test booklets one at a time from each student?	180	199	12
Was the total time for the break between Session 1 and Session 2 equal to 20 minutes?	139	238	14

Exhibit 7.6 summarizes QCMs' observations from the second testing session for grade 8. In the vast majority of sessions, the Test Administrator kept to the time limits prescribed in the directions. Exhibit 7.6 also reveals that in about 70 percent of the sessions the Test Administrator kept to the testing script for signaling a break. Those who did make changes mostly made additions or other minor changes such as paraphrasing the directions. However, here too, QCMs took the question about time for restarting literally. In more than half of the sessions, the time spent to restart the testing session was the prescribed five minutes. For the rest, the session took up to 10 minutes longer to restart. Finally, this exhibit also shows that in only one-quarter of the sessions did students request additional time to complete the student questionnaire. In most cases, this request was granted.

Exhibit 7.6 Testing Session 2 – Eighth Grade

Question	Yes	No	N/A
Was the time spent to restart the testing with Session 2, 5 minutes?	437	314	4
Was the total time for testing in Session 2 correct as indicated in the Administrators' script?	718	27	10
Did the Test Administrator announce "you have 10 minutes left" prior to the end of Session 2?	729	21	5
Were there any other time remaining announcements made during Session 2?	113	631	11
At the end of Session 2, did the Test Administrator collect the test booklets one at a time from each student?	650	91	14
When the Test Administrator read the script to end the testing for Session 2, did the Test Administrator announce a break to be followed by the Student Questionnaire?	610	89	56
How accurately did the Test Administrator read the script to end the testing and signal a break?	531 (no changes)	135 (minor changes) 22 (major changes)	67
If there were changes, how would you describe them?			
Additions	45	179	531
Some minor changes	94	139	522
Omissions	41	165	549
At the end of the break, did the Test Administrator distribute the Student Questionnaires and give directions as specified in the script?	585	82	88
Did the students ask for additional time to complete the questionnaire?	192	494	69
At the end of the session, prior to dismissing the students, did the Test Administrator thank the students for participating in the study?	622	68	65

Exhibit 7.7 summarizes QCMs' observations from the second testing session for grade 4. In the large majority of sessions the Test Administrator kept to the time limits prescribed in the directions. About 60 percent of the Test Administrators stuck to the testing script for signaling a break. Of those who did make changes, most made minor changes such as paraphrasing the directions. Similar to grade 8, QCMs here also took the question about time for restarting literally. In about 40 percent of the sessions, the time spent to restart the testing session was the prescribed five minutes. For the rest, the session took up to 10 minutes longer to restart. Only about one-quarter of students requested additional time to complete the student questionnaire. In most cases, this request was granted.

Exhibit 7.7 Testing Session 2 – Fourth Grade

Question	Yes	No	N/A
Was the time spent to restart the testing with Session 2, 5 minutes?	169	215	7
Was the total time for testing in Session 2 correct as indicated in the Administrators' script?	372	10	9
Did the Test Administrator announce "you have 10 minutes left" prior to the end of Session 2?	367	15	11
Were there any other time remaining announcements made during Session 2?	48	333	10
At the end of Session 2, did the Test Administrator collect the test booklets one at a time from each student?	322	59	10
When the Test Administrator read the script to end the testing for Session 2, did the Test Administrator announce a break to be followed by the Student Questionnaire?	301	40	50
How accurately did the Test Administrator read the script to end the testing and signal a break?	242	84 (minor changes) 10 (major changes)	53
If there were changes, how would you describe them?			
Additions	29	68	294
Some minor changes	66	54	271
Omissions	25	77	289
At the end of the break, did the Test Administrator distribute the Student Questionnaires and give directions as specified in the script?	288	42	61
Did the students ask for additional time to complete the questionnaire?	96	243	52
At the end of the session, prior to dismissing the students, did the Test Administrator thank the students for participating in the study?	304	35	52

Responses to the remaining questions focusing on the test session activities for eighth grade are summarized in Exhibit 7.8. These questions dealt with topics such as student compliance with instructions and the align-

ment of the scripted instructions with their implementation. Exhibit 7.8 shows that in almost all of the sessions, the students complied well or very well with the instructions to stop testing. In addition, in at least 70 percent of the sessions, breaks were conducted exactly or nearly exactly as directed in the script. When this was not the case, it was mostly due to differences in the amount of time allocated for the break. It is also notable that in 95 percent of the testing sessions calculators were not allowed for Session 1 – as required in the script – while in 80 percent of cases calculators were allowed for Session 2.

Exhibit 7.8 Test Session Activities – Eighth Grade

Question	Very Well	Well	Fairly Well	Not well at all	N/A
When the Test Administrator ended Session 1, how well did the students comply with the instruction to stop work (close their booklets and put their pencils down)?	590	136	16	0	13
When the Test Administrator ended Session 2, how well did the students comply with the instruction to stop work (close their booklets and put their pencils down)?	584	142	21	0	8
	Exactly	Nearly the same	Somewhat differently	Not well at all	N/A
Was the first break conducted as directed in the script?	541	133	56	8	17
Was the second break conducted as directed in the script?	457	72	37	48	141
	Exactly the same	Longer	Shorter	N/A	
How did the actual break time compare to the recommended time in the script?	314	113	166	162	
How did the total time allocated for the administration of the Student Questionnaire compare to the time specified in the script?	420	150	111	74	
	Yes	No	N/A		
Were calculators allowed during Session 1?	43	702	10		
Were calculators allowed during Session 2?	604	142	9		
	Very orderly	Somewhat orderly	Not orderly at all	N/A	
In your opinion, how orderly was the dismissal of the students?	502	184	11	58	

Exhibit 7.9 presents the results of the remaining questions that focused on the test session activities for grade 4. Similar to the eighth grade, Exhibit 7.9 shows that in almost all the sessions the students complied well or very well with the instructions to stop testing. In addition, in at least two-thirds of the sessions breaks were conducted exactly or nearly exactly as directed in the script. When this was not the case, it was mostly due to differences in the amount of time allocated for the break. It is also notable that calculators were *not* allowed in almost all testing sessions.

Exhibit 7.9 Test Session Activities – Fourth Grade

Question	Very Well	Well	Fairly Well	Not well at all	N/A
When the Test Administrator ended Session 1, how well did the students comply with the instruction to stop work (close their booklets and put their pencils down)?	311	56	7	1	16
When the Test Administrator ended Session 2, how well did the students comply with the instruction to stop work (close their booklets and put their pencils down)?	317	58	6	0	10
	Exactly	Nearly the same	Somewhat differently	Not well at all	N/A
Was the first break conducted as directed in the script?	255	74	43	3	16
Was the second break conducted as directed in the script?	213	54	16	1098	
	Yes	No	N/A		
Were calculators allowed during Session 1?	1	382	8		
Were calculators allowed during Session 2?	21	358	12		
	Exactly the same	Longer	Shorter	N/A	
How did the actual break time compared to the recommended time in the script?	123	68	71	129	
How does the total time allocated for the administration of the Student Questionnaire compare to the time specified in the script?	158	110	68	55	
	Very orderly	Somewhat orderly	Not orderly at all	N/A	
In your opinion, how orderly was the dismissal of the students?	269	68	1	53	

7.2.3 Summary Observations

Section C of the Classroom Observation Record asked QCMs to reflect on their observations. The QCMs reported overall impressions of the test administration, including how well the Test Administrator monitored students' conduct, and any unusual circumstances that arose during the testing session (e.g., student refusal to participate, defective instrumentation, emergency situations, cheating).

The results presented in Exhibit 7.10 for grade 8 show that in almost all sessions the testing took place without any problems. In the few sessions where problems arose due to defective instrumentation, the Test Administrator replaced the instruments appropriately. In less than five percent of sessions, QCMs reported evidence of students attempting to cheat on the exam. However, when asked to explain the situation, QCMs generally indicated that students were merely looking around at their neighbors to see whether their test booklets were indeed different. Because the TIMSS 2003 test design involves 12 different booklets, students were unlikely to have the same booklet as their neighbors. The QCMs reported that on the rare occasions when they observed serious efforts to cheat, the Test Administrator intervened to prevent cheating. Most of the 31 students who were reported to leave the room for an "emergency" during the testing session had already completed the test. When students left the room for an emergency, Test Administrators handled the situation appropriately by ensuring the security of the test booklets until the students returned. Students were permitted to complete the test when they returned to the classroom.

Exhibit 7.10 also indicates that in almost all of the testing sessions at the eighth grade, QCMs found the behavior of students to be orderly and cooperative. The problem cited most often by QCMs as the reason for disorderly behavior was the noise level of students who had completed the test well before the prescribed 45 minutes had passed. In the few cases where students were disruptive, the Test Administrator was able to control the situation. For the great majority of sessions, QCMs reported that the overall quality of the sessions was either excellent or very good.

Exhibit 7.10 Summary Observations of the QCM – Eighth Grade

Question	Yes	No	N/A			
During the testing sessions did the Test Administrator walk around the room to be sure students were working on the correct section of the test and/or behaving properly?	727	17	11			
Did the Test Administrator address students' questions appropriately?	720	26	9			
Did you see any evidence of students attempting to cheat on the tests (e.g., by copying from a neighbor)?	39	708	8			
Were any defective test booklets detected and replaced before the testing began?	15	726	14			
Were any defective test booklets detected and replaced after the testing began?	20	706	29			
If any defective test booklets were replaced, did the Test Administrator replace them appropriately?	44	19	692			
Did any students refuse to take the test either prior to the testing or during the testing?	17	714	24			
If a student refused, did the Test Administrator accurately follow the instructions for excusing the student (collect the test book and record the incident on the Student Tracking Form)?	32	16	707			
Did any students leave the room for an "emergency" during the testing?	61	685	9			
If a student left the room for an emergency during the testing, did the Test Administrator address the situation appropriately (collect the test booklet, and if re-admitted, return the test booklet)?	56	19	680			
	Extremely	Moderately	Somewhat	Hardly	N/A	
To what extent would you describe the students as orderly and cooperative?	511	207	26	3	8	
	Definitely	Some effort	Hardly any effort	N/A		
If the students were not cooperative and orderly, did the Test Administrator make an effort to control the students and the situation?	199	37	2	517		
	No, there were no late students	No, they were not admitted	Yes, but before testing began	Yes, after testing began	N/A	
Were any late students admitted to the testing room?	659	25	32	27	12	
	Excellent	Very good	Good	Fair	Poor	N/A
In general, how would you describe the overall quality of the testing session?	380	278	68	15	6	8

Exhibit 7.11 presents QCMs' summary observations for fourth grade. Similar to eighth grade, in almost all sessions the testing took place without any problems. In the few sessions where problems arose due to defective instrumentation, the Test Administrator replaced the instruments appropriately. In less than four percent of the sessions, QCMs reported evidence of students attempting to cheat on the exam. Like at grade 8, when asked to explain the situation, QCMs indicated that students were merely looking around at their neighbors to see whether their test booklets were indeed different. The QCMs reported that on the rare occasions when they observed serious efforts to cheat, the Test Administrator intervened to prevent cheating. Most of the 58 students who were reported to leave the room for an "emergency" during the testing session had already completed the test. When students left the room for an emergency, Test Administrators handled the situation appropriately by ensuring the security of the test booklets until the students returned. Students were permitted to complete the test when they returned to the classroom.

Exhibit 7.11 also indicates that in almost all of the testing sessions at the fourth grade, QCMs found the behavior of students to be orderly and cooperative. In the few cases where students were disruptive, the Test Administrator was able to control the situation. For the great majority of sessions, QCMs reported that the overall quality of the sessions was either excellent or very good.

Exhibit 7.11 Summary Observations of the QCM – Fourth Grade

Question	Yes	No	N/A			
During the testing sessions did the Test Administrator walk around the room to be sure students were working on the correct section of the test and/or behaving properly?	371	11	9			
Did the Test Administrator address students' questions appropriately?	380	2	9			
Did you see any evidence of students attempting to cheat on the tests (e.g., by copying from a neighbor)?	16	368	7			
Were any defective test booklets detected and replaced before the testing began?	4	377	10			
Were any defective test booklets detected and replaced after the testing began?	11	362	18			
If any defective test booklets were replaced, did the Test Administrator replace them appropriately?	19	15	357			
Did any students refuse to take the test either prior to the testing or during the testing?	14	355	22			
If a student refused, did the Test Administrator accurately follow the instructions for excusing the student (collect the test book and record the incident on the Student Tracking Form)?	13	9	369			
Did any students leave the room for an "emergency" during the testing?	31	349	11			
If a student left the room for an emergency during the testing, did the Test Administrator address the situation appropriately (collect the test booklet, and if re-admitted, return the test booklet)?	23	13	355			
	Extremely	Moderately	Somewhat	Hardly	N/A	
To what extent would you describe the students as orderly and cooperative?	292	86	6	0	7	
	Definitely	Some effort	Hardly any effort	N/A		
If the students were not cooperative and orderly, did the Test Administrator make an effort to control the students and the situation?	103	12	0	276		
	No, there were no late students	No, they were not admitted	Yes, but before testing began	Yes, after testing began	N/A	
Were any late students admitted to the testing room?	344	13	5	5	24	
	Excellent	Very good	Good	Fair	Poor	N/A
In general, how would you describe the overall quality of the testing session?	194	146	37	0	0	355

7.2.4 Interview with the School Coordinator

The QCM recorded details of the interview with the School Coordinator in Section D of the Classroom Observation Record. The interview addressed the shipment of assessment materials, arrangements for the test administration, the responsiveness of the NRC to queries, the necessity for make-up sessions, and, as a validation of within school sampling procedures, the organization of classes in the school.

The results presented in Exhibit 7.12 for the eighth grade show that TIMSS 2003 was an administrative success in the eyes of School Coordinators. In more than 70 percent of the cases, school officials received the correct shipment of the test materials. Mistakes that did occur tended to be minor and could be remedied prior to testing. Furthermore, about 85 percent of School Coordinators reported that the NRCs were responsive to their questions or concerns, and that the relations were cordial and cooperative. About half of the School Coordinators reported that they were able to collect the completed Teacher Questionnaires prior to student testing. Of the rest, the vast majority reported that they were missing only one or two questionnaires and were expecting them to be turned in shortly. It was estimated that the Teacher Questionnaires would take about 60 minutes to complete. About 50 percent of the School Coordinators indicated that the estimate of 60 minutes was about right, while about 10 percent reported that the questionnaires took longer and about 22 percent that they took less time to complete.

In about 50 percent of the cases, School Coordinators indicated that students were given special instructions, motivational talks, or incentives prior to testing. The majority of students received motivational talks either by a school official, classroom teacher, or the TIMSS Test Administrator. Only a few classes received special instructions or practice, such as reading competitions or extra reading assignments prior to the testing session.

Because the sampling of classes requires a complete list of all mathematics classes in the school at the target grade, QCMs were asked to verify that the class list did indeed include all classes. Although a significant number of School Coordinators reported that this was not so, there may have been some misunderstanding since very few (about 3 percent) knew of any students not in the classes listed.

A tribute to the planning and implementation of TIMSS 2003 was the fact that about 85 percent of respondents said they would be willing to serve as a School Coordinator in future international assessments. Furthermore, the majority of School Coordinators believed the testing session went very well, and that school staff members had positive attitudes towards the TIMSS testing.

Exhibit 7.12 QCM Interviews with the School Coordinator – Eighth Grade

Question	Yes	No	N/A
Prior to the test day did you have time to check your shipment of materials from your TIMSS National Coordinator?	545	122	88
Did you receive the correct shipment of the following items?			
Test booklets	604	55	96
Test Administrator Manual	566	92	97
School Coordinator Manual	556	94	105
Student Tracking Forms	632	35	88
Student Questionnaires	609	51	95
Teacher Questionnaires	639	50	66
School Questionnaire	655	33	67
Test Administration Form	547	112	96
Teacher Tracking Form	470	176	109
Student-Teacher Linkage Form (if applicable)	264	279	212
Envelopes or boxes addressed to the National Center for the purpose of returning the materials after the assessment	453	189	113
Was the National Coordinator responsive to your questions or concerns?	642	21	92
Were you able to collect completed Teacher Questionnaires prior to the test administration?	356	337	62
Was the estimated time of 60 minutes to complete the Teacher Questionnaire a correct estimate?	373	81 (more time) 166 (less time)	135
Were you satisfied with the accommodations (testing room) you were able to arrange for the testing?	695	22	38
Do you anticipate that makeup sessions will be required at your school?	77	625	53
If yes, do you intend to conduct one?	94	119	542
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the assessment?	378	331	46
Is this a complete list of the mathematics classes in this grade in this school?	561	82	112
To the best of your knowledge, are there any students in this grade level who are not in any of these mathematics classes?	25	606	124
To the best of your knowledge are there any students in this grade level in more than one of these mathematics classes?	16	633	106
If there were another international assessment, would you be willing to serve as a School Coordinator?	647	45	63

Exhibit 7.12 QCM Interviews with the School Coordinator – Eighth Grade (...Continued)

	Very well, no problems	Satisfactorily, few problems	Unsatisfactorily, many problems	N/A
Overall, how would you say the session went?	616	94	4	41
	Positive	Neutral	Negative	N/A
Overall, how would you rate the attitude of the other school staff members towards the TIMSS testing?	549	159	10	37
	Worked well	Needs improvement	N/A	
Overall, do you feel the TIMSS School Coordinator Manual worked well or does it need improvement?	584	33	138	

Similar to the eighth grade, the administrative success of TIMSS 2003 at the fourth grade is exemplified by the results of the QCM interviews with School Coordinators, presented in Exhibit 7.13. School Coordinators received the correct shipment of the test materials in most cases. In cases where shipment errors occurred, they tended to be minor and were remedied prior to testing. More than 85 percent of School Coordinators reported that the NRCs were responsive to their questions or concerns.

About half the School Coordinators reported that they were able to collect the completed Teacher Questionnaires prior to student testing. Of those who did not, most reported that teachers completed their questionnaires during the testing sessions. Almost half of the School Coordinators indicated that the estimate of 60 minutes to complete the questionnaire was accurate, while only 11 percent reported that the questionnaires took longer and about 26 percent that they took less time to complete.

In about 35 percent of the cases, School Coordinators indicated that students were given special instructions, motivational talks, or incentives prior to testing. The majority of students received motivational talks either by a school official, classroom teacher, or the TIMSS Test Administrator. Only a few classes received special instructions or practice, such as reading competitions or extra reading assignments prior to the testing session.

As at the eighth grade, a large majority (more than 85 percent) of School Coordinators said they would be willing to serve again in future international assessments. Furthermore, the majority of School Coordinators believed that the testing session went very well, and that school staff members had positive attitudes towards the TIMSS testing.

Exhibit 7.13 QCM Interviews with the School Coordinator – Fourth Grade

Question	Yes	No	N/A
Prior to the test day did you have time to check your shipment of materials from your TIMSS National Coordinator?	282	72	37
Did you receive the correct shipment of the following items?			
Test booklets	347	15	29
Test Administrator Manual	321	42	28
School Coordinator Manual	288	43	60
Student Tracking Forms	341	20	30
Student Questionnaires	349	14	28
Teacher Questionnaires	334	15	42
School Questionnaire	362	0	29
Test Administration Form	320	42	29
Teacher Tracking Form	261	84	46
Student-Teacher Linkage Form (if applicable)	113	172	106
Envelopes or boxes addressed to the National Center for the purpose of returning the materials after the assessment	262	74	55
Was the National Coordinator responsive to your questions or concerns?	335	5	51
Were you able to collect completed Teacher Questionnaires prior to the test administration?	162	204	25
Was the estimated time of 60 minutes to complete the Teacher Questionnaire a correct estimate?	165	45 (longer) 102 (less time)	79
Were you satisfied with the accommodations (testing room) you were able to arrange for the testing?	371	6	14
Do you anticipate that makeup sessions will be required at your school?	44	334	13
If yes, do you intend to conduct one?	48	58	285
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the assessment?	138	237	16
Is this a complete list of the mathematics classes in this grade in this school?	320	30	41
To the best of your knowledge, are there any students in this grade level who are not in any of these mathematics classes?	15	325	51
To the best of your knowledge, are there any students in this grade level in more than one of these mathematics classes?	7	342	42
If there were another international assessment, would you be willing to serve as a School Coordinator?	338	35	18

Exhibit 7.13 QCM Interviews with the School Coordinator – Fourth Grade (...Continued)

Question	Very well, no problems	Satisfactorily, few problems	Unsatisfactorily, many problems	N/A
Overall, how would you say the session went?	329	42	0	20
	Positive	Neutral	Negative	N/A
Overall, how would you rate the attitude of the other school staff members towards the TIMSS testing?	287	86	6	12
	Worked well	Needs improvement	N/A	
Overall, do you feel the TIMSS School Coordinator Manual worked well or does it need improvement?	297	11	83	

7.3 Interview with the National Research Coordinator

In addition to observing testing sessions, QCMs conducted face-to-face interviews with the National Research Coordinators for their countries. The QCM who attended the training session was responsible for conducting this interview and for completing an *Interview with the NRC* form.⁵

The interview questions were designed to elicit NRCs' experiences in preparing for and conducting the TIMSS 2003 data collection with a focus on identifying and selecting samples, working with School Coordinators, translating the instruments, assembling and printing the test materials, packing and shipping the test materials, scoring constructed-response questions, entering and verifying data, choosing quality assurance samples, and suggesting improvements in the process.

7.3.1 Sampling

Section A of the NRC interview involved questions about the sampling process. Topics covered in this section included the extent to which the NRCs used the manuals and sampling software provided by Statistics Canada and the IEA Data Processing Center (DPC) and found them helpful, and the difficulties encountered by NRCs as they carried out this task.

Exhibit 7.14 shows that six countries did not use the school sampling manual provided, but that was because Statistics Canada selected their sample. In one case (Bahrain), no school sampling was necessary because the TIMSS sample included the entire school population. Four-fifth of the NRCs used the within-school sampling software provided by the DPC to select classes. In the cases where the sampling software was not used, it was generally because the country had its own software.

⁵ A total of 50 QCM interviews with the NRCs were conducted. One interview was conducted for Ontario and Quebec combined.

A small number of NRCs encountered organizational constraints in their systems that necessitated deviations from the sample design. In each case, a sampling expert was consulted to ensure that the altered design remained compatible with the TIMSS standards. Sixty percent of NRCs reported that the sampling procedures were not unduly difficult to implement, while nearly 40 percent found the process somewhat difficult. Nevertheless, all but two NRCs managed to deliver school and student samples of high quality for the data collection.⁶

Exhibit 7.14 Interview with the NRC – Sampling

Question	Yes	No	N/A	
Were you able to select a sample of schools and students within schools using the Survey Operations Manual and the Sampling Manual provided by the TIMSS International Study Center?	44	6	0	
Did you use the Within-School Sampling Software provided by the TIMSS International Study Center to select classes or students?	40	9	1	
Were there any conditions or organizational constraints that necessitated deviations from the basic sampling TIMSS design?	8	42	0	
	Very difficult	Somewhat difficult	Not difficult at all	N/A
In terms of the complexity of the procedures and number of personnel needed, how would you describe the process of sample selection?	0	19	30	1

7.3.2 Working with School Coordinators

Questions in Section B of the NRC interview asked about cooperation with the School Coordinators, specifically about communication, shipment of materials, and training. A summary of the responses to the questions in Section B is presented in Exhibit 7.15. At the time the interviews were conducted, nearly all NRCs had contacted the School Coordinators in the sampled school, and most had sent the appropriate materials explaining the testing procedures. Where this was not the case, it was often because a meeting had been scheduled but not yet held. Half the NRCs planned to conduct formal training sessions for School Coordinators prior to the test administration.

Exhibit 7.15 Interview with the NRC – School Coordinator

Question	Yes	No	N/A
Have all the School Coordinators for your sample been contacted?	45	3	2
If yes, have you sent them materials about the testing procedures?	34	14	2
Did you have formal training sessions for the School Coordinators?	25	24	1

⁶ See Chapter 9 for information regarding countries' samples.

7.3.3 Translating the Instruments

Section C of the NRC interview dealt with translation and adaptation of the assessment instruments and manuals. Exhibit 7.16 shows that most NRCs were about evenly split between those who reported little difficulty in translating and adapting the test booklets and questionnaires and those who reported that this was somewhat or very difficult. Most NRCs, however, reported little difficulty in translating the Test Administrator and School Coordinator manuals.

In translating the test booklets, NRCs generally reported using their own staff or a combination of staff and outside experts. Almost all NRCs reported that they had submitted the achievement test booklets to the translation verification program at the International Study Center. At the time of the interview, almost all had received a translation verification report back. More than half the NRCs reported that they had already translated or planned to translate the scoring guides for the mathematics and science constructed-response items. Of those who did not translate the scoring guides, two countries reported that translation was not necessary, since all the scorers were proficient in English.

Exhibit 7.16 Interviews with the NRC – Translation

Question	Own Staff	Outside Experts	Combination	N/A
Did you use your own staff or outside experts to translate the test booklets for verification?	15	10	25	0
	Very difficult	Somewhat difficult	Not difficult at all	N/A
How difficult was it to translate and/or adapt the test booklets?	4	20	26	0
How difficult was it to adapt the questionnaires?	3	22	25	0
How difficult was it to adapt the Test Administrator Manual?	0	9	40	1
How difficult was it to adapt the School Coordinator Manual?	0	8	35	7
	Yes	No	N/A	
Did you go through the process of submitting test booklets and receiving a translation verification report from the ISC?	48 ⁷	1	1	
Did you translate or do you plan to translate the scoring guides for mathematics and science constructed-response items?	28	20	2	

7 Contrary to the data reported by the NRCs, all countries went through the translation verification process. See Chapter 4 for details.

7.3.4 Assembling and Printing the Test Materials

Section D of the NRC interview addressed assembling and printing the test materials, as well as issues related to checking the materials and securely storing them. The results in Exhibit 7.17 show that almost all NRCs were able to assemble the test booklets according to the instructions provided, and that nearly all conducted the recommended quality control checks during the printing process. In the cases where the NRCs did not conduct quality assurance procedures during the printing process, it was because of a shortage of time. Thirty percent of the NRCs detected errors during the printing process. Most countries elected to send their test booklets and questionnaires to an external printer, but printed the manuals in-house. Nearly all NRCs reported having followed procedures to protect the security of the tests during assembly and printing. In no instance was there a breach of security reported.

Exhibit 7.17 Interview with the NRC – Assembling and Printing Test Materials

Question	Yes	No	N/A	
Were you able to assemble the test booklets according to the instructions provided by the International Study Center?	47	3	0	
Did you conduct the quality assurance procedures for checking the test booklets during the printing process?	47	3	0	
Were any errors detected during the printing process?	16	31	3	
If errors were detected, what was the nature of the errors?				
Print quality	10	10	30	
Pages missing	5	10	35	
Page order	1	14	35	
Upside down pages	2	13	35	
Did you follow procedures to protect the security of the tests during the assembly and printing process?	49	1	0	
Did you discover any potential breaches of security?	0	50	0	
	In-House	External	Combination	N/A
Where did you print the test booklets?	8	34	7	1
Where did you print the questionnaires?	10	32	8	0
Where did you print the manuals (TA, SC, Scoring)?	31	11	7	1

7.3.5 Packing and Shipping the Testing Materials

Section E of the NRC interview addressed the extent to which NRCs detected errors in the testing materials as they were packed for shipping to School Coordinators. As shown in Exhibit 7.18, very few errors were found in any of the materials. Errors that were discovered before distribution were remedied. In 18 percent of the cases, the NRCs reported that they had some concerns about confidentiality that restricted their freedom to put student names on the booklet covers. Almost half the NRCs reported having established a procedure to confirm the schools' receipt of the testing materials and for verification of their contents. In most countries, NRCs reported that the deadline for the return of materials from schools was within a day or two of testing. All NRCs reported that the deadline was within two weeks of testing.

Exhibit 7.18 Interview with the NRC – Packaging Test Materials

Question	No Errors, or not used	Errors found before distribution	Errors found after distribution	N/A
In packing the assessment materials for shipment to schools, did you detect any errors in any of the following items?				
Supply of test booklets	32	5	1	12
Supply of student questionnaires	37	1	0	12
Student tracking forms	36	0	1	13
Teacher tracking forms	36	0	0	14
Student-Teacher Linkage Form	35	1	0	14
Test Administrator manual	37	1	0	12
School Coordinator manual	33	1	1	15
Supply of Teacher Questionnaires	37	0	1	12
School questionnaire	36	1	1	12
Test book ID labels	36	1	0	13
Sequencing of booklets or questionnaires	37	1	0	12
Return labels	33	1	1	15
Self-addressed post-cards for test dates	35	1	0	14
	Yes	No	N/A	
Did concerns about confidentiality restrict your freedom to put student names on the booklet covers?	9	32	9	
Do you plan to or have already established a procedure requiring schools to confirm receipt of the testing materials and verification of the contents?	26	12	12	

7.3.6 Scoring Constructed-Response Questions

Section F of the NRC interview focused on the NRC's preparation for scoring the constructed-response items. The scoring process was an ambitious effort, requiring the recruitment and training of scoring staff to score student responses including double scoring to verify scoring reliability. Exhibit 7.19 indicates that, at the time of the NRC interview, about 60 percent of the NRCs had selected their scoring staff, and roughly half of those had already begun the training process. All NRCs reported that they understood the procedures for scoring the reliability sample as explained in the Survey Operations Manual. Two NRCs reported that their own staff would score the constructed-response items, six reported that teachers would do so, six reported that university students would be employed, and 37 reported that a combination of various professionals would score the constructed-response items.

Exhibit 7.19 Interview with the NRC – Scoring

Question	Yes	No	N/A		
Have you selected your scorers for the constructed-response questions?	29	20	1		
If yes, have you trained the scorers?	17	18	15		
Have you scheduled the scoring sessions for the constructed-response questions?	40	9	1		
Do you understand the procedure for scoring the reliability sample as explained in the Survey Operations Manual?	50	0	0		
	Own Staff	Teachers	University Students	Combination	Other
Who will primarily be scoring your constructed-response questions?	2	6	3	37	2

7.3.7 Data Entry and Verification

Section G of the NRC interview addressed preparations for data entry and verification. As shown in Exhibit 7.20, at the time of the interviews about two-thirds of the NRCs had selected their data entry staff and more than half of those selected had participated in training sessions. About two-thirds of the NRCs reported that they planned to enter the data from a percentage of booklets twice, as a verification procedure. The estimated proportion of booklets to be entered twice ranged from five to 50 percent, with two countries reporting that they planned to re-enter 100 percent of the data. Nearly all NRCs established a secure storage area for the returned tests after data entry. Twenty-two NRCs reported that members of their staff would enter the data from test booklets and questionnaires, six reported that an external agency would do so, and 18 reported that a combination of staff and external agency people would enter the data.

7.3.8 Quality Assurance Sample

As part of their national quality assurance activities, NRCs were required to send National Quality Control Observers to a 10 percent sample of the schools to observe the test administration and document compliance with prescribed procedures. These site visits were in addition to the visits to 15 schools conducted by the International Quality Control Monitors.

At the time of the NRC interviews, 64 percent of the NRCs had selected their 10 percent quality assurance sample for site visits. Two NRCs reported that an external agency would conduct the observations, 24 reported that a member of their staff would do so, and 12 reported that a combination of staff and external agency people would conduct the observations. Eight NRCs reported that other professionals, such as inspectors, retired teachers, mathematics and science supervisors or university professors, would be recruited to conduct the on-site observations.

Exhibit 7.20 Interview with the NRC – Data Entry and Verification

Question	Yes	No	N/A	
Have you selected the data entry staff?	37	11	2	
If yes, have you conducted training sessions for the data entry staff?	21	17	12	
Do you plan to key enter a percentage of test booklets twice as a verification procedure?	37	10	3	
Have you established a secure storage area for the returned tests after coding and until the original documents can be discarded?	48	2	0	
	Own Staff	External Firm	Combination	Other
Do you plan to use your own staff or outside experts to enter the data from the achievement test booklets and questionnaires onto computer files?	22	6	18	4

7.3.9 The Survey Activities Report

The final section of the NRC interview asked the NRC for comments on any aspects of the study they felt might improve the assessment process. A major concern expressed by many NRCs was a time constraint for accomplishing all that was required to keep up with the demanding TIMSS 2003 schedule particularly the translation and preparation of the instruments. Some NRCs indicated they did not have ample staff.

References

TIMSS (2002a), *TIMSS 2003 Manual for International Quality Control Monitors*, prepared by the TIMSS & PIRLS International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002b), *TIMSS 2003 Survey Operations Manual*, prepared by the TIMSS & PIRLS International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002c), *TIMSS 2003 School Coordinator Manual*, prepared by the TIMSS & PIRLS International Study Center, Chestnut Hill, MA: Boston College.

TIMSS (2002d), *TIMSS 2003 Test Administrator Manual*, prepared by the TIMSS & PIRLS International Study Center, Chestnut Hill, MA: Boston College.