Adams, R.J. and Gonzalez, E.J. (1996) "The TIMSS Test Design" in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development.* Chestnut Hill, MA: Boston College.

3. The TIMSS Test Design

Raymond J. Adams
Eugenio J. Gonzalez

3.1 OVERVIEW

This chapter describes the TIMSS test design for each of the three TIMSS student populations. Underpinning the design is the recognition that the main goal of TIMSS was to estimate various characteristics of the defined populations in participating countries over a broad range of outcomes in mathematics and science. Accommodating the research demands and priorities of participants around the world, the curricular differences among educational systems, the precision necessary for the estimation of parameters, the constraints on testing time, and the need for simple administrative procedures required a test design which was flexible enough to meet the needs of individual participants yet rigid enough to meet acceptable quality standards. These competing demands imposed on the TIMSS test design can be summarized as a tension between the desire for wide subject-matter coverage and the limitations imposed by available resources.

Fortunately, modern test-scaling methods such as item response theory (Lord, 1980; Wright and Stone, 1979) and plausible value technology (Rubin, 1987; Mislevy and Sheehan, 1987) have made it possible to deal with much of the tension among practicality, coverage, and precision. Most important, it has been shown (Lord, 1962; Beaton, 1987) that when these methods are used, accurate estimates can be obtained for populations without precise measurement of individual students. That is, not all students in an assessment sample need to respond to all the test items. To derive population estimates, a multiple matrix sampling design is used in which subsets of items selected from the total item pool are administered

to random subsamples of students. The testing time for each student, and consequently the number of items administered to each student, is limited to an acceptable level, yet the population performance on a range of dimensions can be characterized.

The TIMSS tests were designed to be administered to three TIMSS student populations.

Population 1

Students in Population 1 (the two adjacent grades containing the largest proportion of 9-year-old students) were to be given tests that contained items in both mathematics and science. Proficiency estimates were required for each of ten reporting categories in mathematics (six) and science (four). A total of 199 unique items were to be distributed across eight test booklets for this population.

Population 2

Students in Population 2 (the two adjacent grades containing the largest proportion of 13-year-old students) were also to be given tests that contained items in both mathematics and science. Proficiency estimates were required for each of eleven reporting categories in mathematics (six) and science (five). A total of 286 unique items were to be distributed across eight test booklets for this population.

Population 3

Students in Population 3 (the final year of secondary schooling) were to be given tests containing items of general mathematical and scientific knowledge and on applications of mathematical and scientific principles to everyday situations. Proficiency estimates were required for each of three reporting categories: mathematics literacy, science literacy, and reasoning and social utility. A total of 76 unique items were to be distributed across two test booklets for Population 3 students.

Students in Population 3 taking advanced courses in mathematics constituted a subpopulation for which separate reporting was required. These students were to be given tests in advanced mathematics. Proficiency estimates in five reporting categories were required. A total of 65 advanced mathematics items were to be distributed across four booklets for these students. Similarly, students in Population 3 taking advanced courses in physics constituted a subpopulation. These students were to be given tests in physics, for which proficiency estimates in five reporting categories were also required. A total of 65 physics items were to be distributed across four booklets for these students.

3.2 CONSTRAINTS OF THE TIMSS TEST DESIGN

In the design of the test booklets several constraints had to be taken into account. Because of the different characteristics of the populations, the constraints for the Population 3 test booklets differed from those for Populations 1 and 2. Some of the key constraints for the Population 1 and 2 tests are detailed below.

- The total testing time for individual students in Population 1 was not to exceed 70 minutes

- The total testing time for individual students in Population 2 was not to exceed 90 minutes

- There were to be a maximum of eight different test booklets each for Populations 1 and 2, to be distributed randomly ("rotated") in each sampled classroom

- There was to be a small subset of items common to all test booklets, to allow the verification of within-school booklet rotation and testing of the equivalence of the student samples assigned each booklet

- The booklets were to be approximately parallel in content coverage, although it was acceptable for some booklets to have more mathematics than science content and vice versa

- The booklets were to be approximately parallel in difficulty

- Each booklet was to contain some mathematics and some science items.[1]

- Content coverage in each subject (mathematics and science) was to be as broad as possible. In Population 1, items from 11 mathematics content areas and 7 science content areas of the TIMSS curriculum frameworks (Robitaille et al., 1993) were to be included in the tests. In Population 2, items from 12 mathematics content areas and 10 science content areas were to be included

- It was anticipated that resources would be insufficient to provide precise estimates of item-level statistics for all items in the TIMSS item pool. Therefore, it was decided that such estimates would be required only for a subset of test items. Those items should therefore appear in as many booklets as possible. Others would appear in only one or two booklets, and would have less precise item-level statistics

- For a subset of the items, all of the item intercorrelations were required, whereas for the remaining items only a sample of item intercorrelations was estimated

Some of the key constraints for the design of tests for Population 3 are listed below.

- The total testing time for individual students in Population 3 was not to exceed 90 minutes

- There were to be test booklets for the general Population 3 students, as well as test booklets for each subpopulation (students taking advanced mathematics and students taking physics)

- Booklets for the general population were to be approximately parallel in content, with a set of common items

- Booklets for the advanced mathematics subpopulation were to be approximately parallel in content, with a subset of items common to all of the advanced mathematics booklets

- Booklets for the physics subpopulation were to be approximately parallel in content, with a subset of items common to all of the physics booklets

- There was to be a booklet containing advanced mathematics items and physics items for students taking advanced courses in both mathematics and physics. This booklet

---

[1] While some mathematics and some science occurs in every booklet, it was not necessary to balance the booklets by having an equal proportion of mathematics and science, since this would reduce the number of within-subject inter-item covariances that can be estimated.

was also to contain reasoning and social utility items designed for the general Population 3 students

- The mathematics and science literacy booklets were to include items from mathematics, science, and reasoning and social utility

- Items from five mathematics content areas were to be included in the booklets for the students taking advanced mathematics

- Items from five physics content areas were to be included in the booklets for the physics students.

Aside from the constraints listed above, the test design for all three populations had to cater to the use of three test-item types: multiple-choice items, short-answer items, and extended-response items (short-answer and extended-response items are collectively referred to as free response items). Multiple-choice items require the student to recognize and identify the correct response to an item on the test form. The short-answer items require the student to give a short response such as a number, phrase, sentence, or diagram. The extended-response items involve more elaborate responses, and may require that students explain their reasoning or detail the steps they took in solving a problem. While the multiple-choice and most short-answer items can be adequately scored as either correct or incorrect, the extended-response items require a more complex system of partial-credit scoring to do justice to the wider range of student responses. Some free-response items had two parts, each analyzed separately. In TIMSS, short-answer and extended-response items can receive anywhere from 0 to 4 score points, depending on the complexity and quality of the response elicited and whether it is a multi-part item, while all multiple-choice items are scored 0 or 1.

The inclusion of a variety of item types required that assumptions be made, at the time of test construction, about the average response time of students to the different types of items. The working assumption for the required response time, for each item type for each population, are presented in Table 3.1.

Table 3.1   Minutes of Response Time Required

Item Type	Population 1	Population 2	Population 3 Literacy	Population 3 Advanced
Multiple-Choice	1	1	1	3
Short-Answer	1	2	2	3
Extended-Response	3	5	5	5

3.3 A CLUSTER-BASED DESIGN

The design chosen for the TIMSS tests for all three populations calls for two stages. Items in the item pool were first assigned to one of a set of mutually exclusive groups, or "clusters." The clusters of items were then systematically assigned to test booklets. An item cluster is a small group of items that are collected and then treated as a block for the purpose of test construction. A cluster might appear in more than one test booklet;

furthermore, a cluster that appears in more than one booklet might appear in a different position within each booklet (e.g., first in one booklet, second in another, third in another). In each of Populations 1 and 2 there is one cluster of items (the core cluster) that appears in all eight booklets for that population.

Each item cluster has an identical format and layout wherever it appears. A test booklet is made up of item clusters and corresponds to the set of items that will be administered to an individual student. Because item clusters may be allocated to more than one booklet, the booklets do not contain mutually exclusive subsets of the total item pool: some test items will appear in more than one booklet.

The number of items in each cluster varies, because the cluster design is based on the estimated number of minutes it would take a typical student to answer the items, rather than on the total number of items. For Population 1, one item cluster has been designed to take 10 minutes while the remaining 25 have been designed to take 9. The Population 2 item clusters vary in length: 8 of the clusters are estimated to take 10 minutes, 8 to take 12 minutes, and 10 to take 22 minutes. This means that there is a total pool of 235 unique testing minutes in Population 1 and a total of 396 unique testing minutes in Population 2. Half of these pools were allocated to mathematics and half to science in each population. For Population 3, the test items have been grouped into 12 clusters of varying length. There are 4 core clusters, each 30 minutes in length. The remaining 8 clusters vary in length, from 30 minutes (for the literacy rotated clusters) to 60 minutes (for the advanced mathematics and physics rotated clusters).

3.4 TIMSS POPULATION 1 TEST DESIGN

All test items for Population 1 were grouped into 26 mutually exclusive item clusters, each identified by a letter of the alphabet (A through Z). There is a core cluster (cluster A) that appears in the second position in every booklet. Each cluster contains items in mathematics, science, or both. Each test booklet for Population 1 comprises up to 7 item clusters. The booklets are divided into two parts and are administered in two consecutive testing sessions. Some clusters appear in all booklets, some in four, some in three, some in two, and some in only one booklet. Of the 26 clusters at Population 1, 18 take 9 minutes and 8 take 10 minutes.

In Population 1, it is convenient to regard the clusters as being of the six types described in the following.

1. Core Cluster	One Core cluster comprising five mathematics and five science multiple-choice items was assigned to all booklets. It is labeled cluster A.
2. Focus Clusters	There were seven Focus clusters, each containing nine minutes of science and mathematics items. Four of the Focus clusters contain five mathematics and four science multiple-choice items each, and three of them contain four mathematics and five multiple-choice items.[2] These clusters were assigned to each of the first seven booklets. They are called Focus clusters because each appears in at least three booklets, so that the items in them were answered by a relatively large fraction (three-eighths) of the student sample in each country–enough to permit accurate reporting of the item statistics. The Focus clusters are labeled B through H.
3. Mathematics Breadth Clusters	There were five Mathematics Breadth clusters, each containing nine minutes of mathematics items and consisting largely, but not exclusively, of multiple-choice items. These clusters appear in only one booklet, and consequently the number of students responding to each item was small. While the items in these clusters contribute to the breadth of content coverage of the tests, the accuracy of the item statistics would be relatively low. These clusters are labeled I through M.
4. Science Breadth Clusters	There were five Science Breadth clusters, each containing nine minutes of science items and consisting largely, but not exclusively, of multiple-choice items. These clusters appear in only one booklet, and consequently the number of students responding to each item was small. While the items in these clusters contribute to the breadth of the tests, the accuracy of their item statistics would be relatively low. These clusters are labeled N through R.
5. Mathematics Free-Response Clusters	There were four Mathematics Free-Response clusters, each containing nine minutes of short-answer and extended-response mathematics items. These clusters were each assigned to two booklets, so that item statistics of reasonable accuracy would be available. These clusters are labeled S through V.
6. Science Free-Response Clusters	There were four Science Free-Response clusters, each containing nine minutes of short-answer and extended-response science items. These clusters were each assigned to two booklets, so that item statistics of reasonable accuracy would be available. These clusters are labeled W through Z.

---

[2] In the final design two of the clusters (E and H) were assigned one short-answer science item instead of a multiple-choice item.

Tables 3.2 and 3.3 list the clusters and the number of items of each type allocated to each cluster for mathematics and science respectively. The number of items per cluster varies because items have been allocated to clusters on the basis of the estimated number of minutes that it would take a typical student to answer them, rather than on the total number of items.

Table 3.2   Number of Mathematics Items per Cluster, by Item Type, for Population 1

Cluster	Multiple-Choice	Short-Answer	Extended-Response	Total
A	5	-	-	5
B	5	-	-	5
C	4	-	-	4
D	5	-	-	5
E	4	-	-	4
F	5	-	-	5
G	4	-	-	4
H	5	-	-	5
I	9	-	-	9
J	9	-	-	9
K	9	-	-	9
L	8	1	-	9
M	7	2	-	9
S	-	3	2	5
T	-	3	2	5
U	-	3	2	5
V	-	3	2	5
Total	**79**	**15**	**8**	**102**

Table 3.3   Number of Science Items per Cluster, by Item Type, for Population 1

Cluster	Multiple-Choice	Short-Answer	Extended-Response	Total
A	5	-	-	5
B	4	-	-	4
C	5	-	-	5
D	4	-	-	4
E	4	1	-	5
F	4	-	-	4
G	5	-	-	5
H	3	1	-	4
N	9	-	-	9
O	7	2	-	9
P	8	1	-	9
Q	7	2	-	9
R	8	1	-	9
W	-	3	2	5
X	1	2	2	5
Y	-	-	3	3
Z	-	-	3	3
Total	**74**	**13**	**10**	**97**

3.4.1 ORGANIZATION OF THE TEST BOOKLETS

In Population 1, the test design specifies eight booklets, each estimated to take a student 64 minutes to complete. Each booklet was constructed from one ten-minute Core cluster (cluster A) and six nine-minute clusters. Table 3.4 shows the assignment of clusters to booklets, as well as the position of each cluster within the booklet.

Table 3.4   Assignment of Item Clusters to Population 1 Booklets[3]

Cluster Type	Cluster Label	Booklet							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Core (10 minutes)	A	2	2	2	2	2	2	2	2
Focus (10 minutes)	B	1				5		3	1
	C	3	1				5		
	D		3	1				5	
	E	5		3	1				
	F		5		3	1			
	G			5		3	1		
	H				5		3	1	
Breadth (Mathematics) (9 minutes)	I								5
	J	6							
	K			6					
	L					6			
	M							6	
Breadth (Science) (9 minutes)	N		6						
	O				6				
	P						6		
	Q								6
	R								3
Mathematics Free-Response (9 minutes)	S	4							7
	T	7		4					
	U			7		4			
	V					7		4	
Science Free-Response (9 minutes)	W		4					7	
	X		7		4				
	Y				7		4		
	Z						7		4

---

[3] Numbers in the cells indicate the position of the cluster within the booklet.

The order of the clusters within the Population 1 booklets is shown in Table 3.5. Cluster A is the Core cluster and has been assigned to all booklets. The rotation design used to assign clusters B through H to booklets 1 through 7 allows the estimation of all item covariances for the items in clusters A through H. Booklet 8 serves primarily to increase the content coverage of the tests. Apart from booklet 8 (which has three), each booklet has only one Breadth cluster, and each Breadth cluster appears in only one booklet.

Table 3.5   Ordering of Item Clusters Within Population 1 Booklets

Cluster Order	Booklet							
	1	2	3	4	5	6	7	8
1st	B	C	D	E	F	G	H	B
2nd	A	A	A	A	A	A	A	A
3rd	C	D	E	F	G	H	B	R
4th	S	W	T	X	U	Y	V	Z
BREAK								
5th	E	F	G	H	B	C	D	I
6th	J	N	K	O	L	P	M	Q
7th	T	X	U	Y	V	Z	W	S

The Population 1 test design has the following features.

· The Core cluster (cluster A) appears in the second position in all test booklets.

· The Focus clusters (clusters B through H) each appear in at least three booklets, each time in a different position. They are assigned to each of the first seven booklets following a Balanced Incomplete Block design. In booklets 1 through 7, each Focus cluster appears together once with each of the remaining Focus clusters.

· Each of the Focus clusters occurs once in the first, third, and fifth positions in booklets 1 through 7.

· All test booklets contain mathematics and science items. Test booklets 1, 3, 5, and 7 have more mathematics items; booklets 2, 4, 6, and 8 have more science items.

· The test booklets are designed to be administered in two consecutive testing sessions with a 15-20-minute break in between. The first four clusters of items in the Population 1 test booklets were administered during the first testing session (37 minutes); after the break the remaining three clusters were administered (27 minutes).

· There are Free-Response clusters in Part 1 as well as in Part 2 of each test booklet (fourth and seventh cluster in each booklet).

· The design provides a total of 235 minutes, 118 for mathematics and 117 for science.

3.4.2 CONTENT OF THE TEST BOOKLETS

Test items were included in the Population 1 tests from 11 content areas in mathematics and 7 content areas in science. Some of these content areas are merged, for the purpose of scaling and reporting, into 6 reporting categories for mathematics and 4 for science. That is,

it will be possible to characterize the TIMSS Population 1 with respect to 6 mathematics and 4 science achievement dimensions.

The 6 mathematics reporting categories are:

- Whole numbers
- Fractions and proportionality
- Measurement, estimation, and number sense
- Data representation, analysis, and probability
- Geometry
- Patterns, relations, and functions.

The 4 science reporting categories are:

- Earth science
- Life science
- Physical science
- Environmental issues and the nature of science.

The TIMSS test blueprints (see Chapter 2) and the TIMSS curriculum frameworks provide more information on the composition of these reporting categories.

The Core and Focus clusters contain multiple-choice items only. The Breadth clusters include multiple-choice items and some short-answer items. Free-Response clusters consist almost exclusively of short-answer and extended-response items.

When items from both mathematics and science were included in a cluster they were grouped so that all of the mathematics items appear in a contiguous sequence, as do all of the science items. In half of these clusters the mathematics items were presented first; in the other half, the science items. Within each sequence, items were placed in order of estimated difficulty.

Table 3.6 shows the number of items by type, and the associated number of score points, for each of the content-based reporting categories for Population 1 mathematics. Table 3.7 provides the same information for Population 1 science.

Table 3.6   Number of Mathematics Items of Each Type and Score Points, by Reporting Category, Population 1

Reporting Category	Multiple-Choice	Short-Answer	Extended-Response	Total Items	Score Points
Whole numbers	19	5	1	25	28
Fractions and proportionality	15	2	4	21	28
Measurement, estimation, and number sense	16	3	1	20	21
Data representation, analysis, and probability	8	2	2	12	15
Geometry	12	2	-	14	14
Patterns, relations, and functions	9	1	-	10	10
Total	**79**	**15**	**8**	**102**	**116**

Table 3.7   Number of Science Items of Each Type and Score Points, by Reporting Category, Population 1

Reporting Category	Multiple-Choice	Short-Answer	Extended-Response	Total Items	Score Points
Earth science	13	2	2	17	19
Life science	33	5	3	41	45
Physical science	23	4	3	30	33
Environmental issues and the nature of science	5	2	2	9	11
Total	**74**	**13**	**10**	**97**	**108**

Tables 3.8 and 3.9 show the number of items from each reporting category that are included in each of the eight test booklets for mathematics and for science. Tables 3.10 and 3.11 show the maximum number of possible score points for each reporting category.

Table 3.8   Number of Mathematics Items in Each Booklet by Reporting Category, Population 1

Reporting Category	Booklet							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Whole numbers	9	4	7	6	10	6	11	7
Fractions and proportionality	9	6	10	4	8	4	7	8
Measurement, estimation, and number sense	9	4	9	5	8	4	6	3
Data representation, analysis, and probability	5	3	3	1	4	1	4	2
Geometry	4	2	4	2	4	1	4	3
Patterns, relations, and functions	1	-	4	1	4	2	2	1
Total	**37**	**19**	**37**	**19**	**38**	**18**	**34**	**24**

Table 3.9   Number of Science Items in Each Booklet by Reporting Category, Population 1

Reporting Category	Booklet							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Earth science	6	7	7	10	7	8	5	4
Life science	6	16	6	13	6	14	10	12
Physical science	4	10	4	8	4	9	5	11
Environmental issues and the nature of science	3	4	2	4	1	3	2	3
Total	**19**	**37**	**19**	**35**	**18**	**34**	**22**	**30**

Table 3.10   Maximum Number of Mathematics Score Points in Each Booklet by Reporting Category, Population 1

Reporting Category	Booklet							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Whole numbers	9	4	7	6	13	6	14	7
Fractions and proportionality	13	6	15	4	11	4	8	9
Measurement, estimation, and number sense	9	4	10	5	9	4	6	3
Data representation, analysis, and probability	8	3	5	1	4	1	4	3
Geometry	4	2	4	2	4	1	4	3
Patterns, relations, and functions	1	-	4	1	4	2	2	1
Total	**44**	**19**	**45**	**19**	**45**	**18**	**38**	**26**

Table 3.11   Maximum Number of Science Score Points in Each Booklet by Reporting Category, Population 1

Reporting Category	Booklet							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Earth science	6	7	7	11	7	10	5	5
Life science	6	16	6	15	6	18	10	14
Physical science	4	12	4	9	4	10	6	12
Environmental issues and the nature of science	3	5	2	5	1	3	2	3
Total	**19**	**41**	**19**	**40**	**18**	**41**	**24**	**34**

Table 3.12 shows the rotation ratios (the number of times an item appears in a booklet) for items that belong to each cluster and the number of times that items from pairs of clusters appear together. The rotation ratios for the items in each cluster are shown in the diagonal elements of the matrix (only the diagonal and lower-triangular parts of the matrix are shown). For example, the Core cluster appears in all 8 booklets, so its rotation ratio is 8. For the Focus clusters the rotation ratio is 3 (except for cluster B, which has a rotation ratio of 4).

The nondiagonal elements of Table 3.12 give the number of times that a pair of clusters appears together in a booklet. If two clusters appear together, then the covariances between the items in them can be estimated (of course all the covariances between items within each cluster will be available). A dash indicates that the corresponding pair of clusters never appears in the same booklet. The first column of the matrix shows that the Core cluster appears at least once with each other cluster.   Further, the matrix shows that few covariances between the items in different Breadth clusters, and between those in different Mathematics and Science Free-Response clusters, will be available.   The design does, however, ensure that all covariances between items in Focus clusters, and most covariances between items in Focus and Free-Response clusters, will be available.

Table 3.12   Rotation Ratios for Item Clusters and Cluster Pairings, Population 1

```
A  8      1. Core

B  4 4        2. Focus
C  3 1 3
D  3 1 1 3
E  3 1 1 13
F  3 1 1 11 3
G  3 1 1 11 1 3
H  3 1 1 11 1 13

I  1 1 - - - - - -    1         3. Mathematics Breadth
J  1 1 1 -1 - - -    -1
K  1 - -11 -1 -    - -1
L  1 1 - - -11 -    - - -1
M  1 1 -1 - - -1    - - - -1

N  1 -1 1 -1 - -    - - - --    1       4. Science Breadth
O  1 - - -11 -1    - - - --    -1
P  1 -1 - - -11    - - - --    - -1
Q  1 1 - - - - - -    1 - - --    - - -1
R  1 1 - - - - - -    1 - - --    - - -11

S  2 2 1 -1 - - -    1 1 - - -    - - -11    2       5. Mathematics Free-Response
T  2 1 1 12 -1 -    -1 1 --    - - - - -    12
U  2 1 -11 1 2 -    - -1 1-    - - - - -    -1 2
V  2 2 -1 -1 11    - - -11    - - - - -    - -1 2

W  2 1 1 2 -1 -1    - - - -1    1 - - - -    - - -1    2             6. Science Free-Response
X  2 -1 11 2 -1    - - - - -    1 1 - - -    - - - -    12
Y  2 -1 -1 1 12    - - - --    -1 1 - -    - - - -    -1 2
Z  2 1 1 - - -11    1 - - --    - -1 1 1    1 - - -    - -1 2

   A B C D E F G H   I J K L M   N O P Q R   S T U V   W X Y Z                CLUSTER
```

3.5 TIMSS POPULATION 2 TEST DESIGN

The test design for Population 2 is very similar to that for Population 1. As in the Population 1 design, all test items in Population 2 were grouped into 26 mutually exclusive item clusters, each identified with a letter of the alphabet (A through Z). There is a core cluster (cluster A) that appears in the second position in every booklet.   Each cluster contains items in mathematics, science, or both. Each booklet comprises of up to seven item clusters. The booklets are divided into two parts and administered in two consecutive testing sessions. One cluster appears in all booklets and some clusters appear in all three, two, or only one booklet. Of the 26 clusters in Population 2, 8 take 12 minutes, 10 take 22 minutes, and 8 take 10 minutes. The design thus provides 396 unique testing minutes, 198 for science and 198 for mathematics.

In Population 2, it is convenient to regard the clusters as being of five types.

1. Core Cluster	One Core cluster comprising six mathematics and six science multiple-choice items was assigned to all booklets. It is labeled cluster A.
2. Focus Clusters	There were seven Focus clusters, each containing six mathematics and six science multiple-choice items. These Focus clusters are assigned to Booklets 1 through 7. They are called Focus clusters because each appears in at least three booklets, so that the items in them were answered by a relatively large fraction (three-eighths) of the student sample in each country–enough to permit accurate reporting of the item statistics. These clusters are labeled B through H.
3. Mathematics and Science Breadth Clusters	There were 10 Breadth clusters, each containing 11 minutes of mathematics and 11 minutes of science items. These clusters consist largely but not exclusively of multiple-choice items. They appear in only one booklet and consequently the number of students responding to each item was small. While the items in these clusters contribute to the breadth of content coverage of the tests, the accuracy of their item statistics would be relatively low. These clusters are labeled I through R.
4. Mathematics Free-Response Clusters	There were four Mathematics Free-Response clusters, each containing 10 minutes of short-answer and extended-response items. These clusters were each assigned to two booklets, so that item statistics of reasonable accuracy would be available. These items are labeled S through V.
5. Science Free-Response Clusters	There were four Science Free-Response clusters, each containing 10 minutes of short-answer and extended-response items. These clusters were each assigned to two booklets, so that item statistics of reasonable accuracy would be available. These clusters are labeled W through Z.

Tables 3.13 and 3.14 list the clusters and the number of items of each type allocated to each cluster for mathematics and science, respectively. The number of items in each cluster varies because the clusters have been designed on the basis of the estimated number of minutes that it would take a typical student to answer them, rather than on the total number of items.

Table 3.13   Number of Mathematics Items per Cluster, by Item Type, for Population 2

Cluster	Multiple-Choice	Short-Answer	Extended-Response	Total
A	6	-	-	6
B	6	-	-	6
C	6	-	-	6
D	6	-	-	6
E	6	-	-	6
F	6	-	-	6
G	6	-	-	6
H	6	-	-	6
I	7	2	-	9
J	7	2	-	9
K	7	2	-	9
L	9	1	-	10
M	7	2	-	9
N	7	2	-	9
O	7	2	-	9
P	9	1	-	10
Q	9	1	-	10
R	7	2	-	9
S	-	-	2	2
T	-	-	2	2
U	-	-	2	2
V	1	2	1	4
Total	**125**	**19**	**7**	**151**

Table 3.14   Number of Science Items per Cluster, by Item Type, for Population 2

Cluster	Multiple-Choice	Short-Answer	Extended-Response	Total
A	6	-	-	6
B	6	-	-	6
C	6	-	-	6
D	6	-	-	6
E	6	-	-	6
F	6	-	-	6
G	6	-	-	6
H	6	-	-	6
I	9	1	-	10
J	7	2	-	9
K	8	2	-	10
L	6	-	1	7
M	2	2	1	5
N	8	2	-	10
O	4	4	-	8
P	3	4	-	7
Q	5	3	-	8
R	2	2	1	5
W	-	-	2	2
X	-	-	2	2
Y	-	-	2	2
Z	-	-	2	2
Total	**102**	**22**	**11**	**135**

3.5.1 ORGANIZATION OF THE TEST BOOKLETS

As for Population 1, the Population 2 test design specifies eight booklets. While for Population 1 each booklet required 64 minutes for a student to complete, for Population 2 90 minutes were required. Except for booklet 8, each booklet was constructed from one 12-minute Core cluster (cluster A), three 12-minute Focus clusters, one 22-minute Breadth cluster, and two 10-minute Mathematics or Science Free-Response clusters. Table 3.15 shows the assignment of clusters to booklets, as well as the position of each cluster within the booklet.

Table 3.15   Assignment of Item Clusters to Population 2 Booklets[4]

Cluster Type	Cluster Label	Booklet							
		1	2	3	4	5	6	7	8
Core (12 minutes)	A	2	2	2	2	2	2	2	2
	B	1				5		3	1
	C	3	1				5		
Focus (12 minutes)	D		3	1				5	
	E	5		3	1				
	F		5		3	1			
	G			5		3	1		
	H					5		3	1
Breadth (Mathematics and Science) (22 minutes)	I	6							
	J		6						
	K			6					
	L				6				
	M					6			
	N						6		
	O							6	
	P								6
	Q								3
	R								5
Mathematics Free-Response (10 minutes)	S	4							
	T	7		4					
	U			7		4			
	V					7		4	
Science Free-Response (10 minutes)	W		4					7	
	X		7		4				
	Y				7		4		
	Z						7		

---

[4] Numbers in the cells indicate the position of the cluster within the booklet.

The order of the clusters within the Population 2 booklets is shown in Table 3.16. Cluster A is the Core cluster and has been assigned to all booklets. The rotation design used to assign clusters B through H to booklets 1 through 7 allows the estimation of all item covariances for the items in clusters A through H.

Table 3.16   Ordering of Clusters Within Population 2 Booklets

Cluster Order	Booklet							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1st	B	C	D	E	F	G	H	B
2nd	A	A	A	A	A	A	A	A
3rd	C	D	E	F	G	H	B	Q
4th	S	W	T	X	U	Y	V	
BREAK								
5th	E	F	G	H	B	C	D	R
6th	I	J	K	L	M	N	O	P
7th	T	X	U	Y	V	Z	W	

Booklet 8 serves primarily to increase the content coverage of the tests. Apart from booklet 8 (which has three), each booklet has only one Breadth cluster, and each Breadth cluster appears in only one booklet. This means that covariances between items in different Breadth clusters cannot be directly estimated. For each item in each Breadth cluster, covariances can be directly estimated with half of the items in the Focus clusters and with all of the items in the Core cluster.

Similarly, the rotation of the Free-Response clusters restricts estimation of the covariances between items in different Free-Response clusters and between free-response and multiple-choice items. Most of the covariances between the items in the Free-Response and Focus clusters can be directly estimated, as can more than half of those between items in the mathematics Free-Response clusters. The same situation applies to science. Only a small number of covariances between items in the Mathematics and Science Free-Response clusters can be estimated.

The Population 2 test design has the following features.

- The Core cluster (cluster A) appears in the second position in all test booklets.

- The Focus clusters (clusters B through H) each appear in at least three booklets, each time in a different position. They are assigned to each of the first seven booklets following a Balanced Incomplete Block design. In booklets 1 through 7, each Focus cluster appears together once with each of the remaining Focus clusters.

- Each of the Focus clusters occurs once in the first, third, and fifth positions in test booklets 1 through 7.

- All test booklets contain mathematics and science items. Test booklets 1, 3, 5, and 7 have more mathematics items; booklets 2, 4, 6, and 8 have more science items.

- The test booklets are designed to be administered in two consecutive testing sessions with a 15-20-minute break in between. The first four clusters of items in the Population 2 test booklets were administered during the first testing session (46 minutes); after the break the remaining three clusters were administered (44 minutes).

- There are Free-Response clusters in Part 1 as well as in Part 2 of each test booklet (fourth and seventh cluster in each booklet).

- The design provides a total of 396 testing minutes, 198 for science and 198 for mathematics.

3.5.2 CONTENT OF THE TEST BOOKLETS

Test items were included in the Population 2 tests from 12 content areas in mathematics and 10 in science. Some of these content areas are merged, for the purpose of scaling and reporting, into six reporting categories for mathematics and five reporting categories for science. That is, it will be possible to characterize the TIMSS Population 2 with respect to six mathematics and five science achievement dimensions. The six mathematics reporting categories are:

- Fractions and number sense

- Geometry

- Algebra

- Data representation, analysis, and probability

- Measurement

- Proportionality.

The five science reporting categories are:

- Earth science

- Life science

- Physics

- Chemistry

- Environmental issues and the nature of science.

The TIMSS test blueprints (see Chapter 2) and the TIMSS curriculum frameworks provide more information on these reporting categories.

The Core and Focus clusters contain multiple-choice items only. The Breadth clusters include multiple-choice items and some short-answer items. Free-Response clusters consist almost exclusively of short-answer and extended-response items.

When items from both mathematics and science were included in a cluster, all the mathematics items appear in a contiguous sequence, as do all of the science items. In half of these clusters the mathematics items are presented first; in the other half, the science items. Within each sequence, items were placed in order of estimated difficulty.

Table 3.17 shows the number of items by type, and the associated maximum number of score points, for each of the content-based reporting categories for Population 2 mathematics. Table 3.18 provides the same information for Population 2 science.

Table 3.17   Number of Mathematics Items of Each Type, and Maximum Score Points, by Reporting Category, Population 2

Reporting Category	Multiple-Choice	Short-Answer	Extended-Response	Total Items	Score Points
Fractions and number sense	41	9	1	51	52
Geometry	22	1	-	23	23
Algebra	22	3	2	27	30
Data representation, analysis, and probability	19	1	1	21	23
Measurement	13	3	2	18	23
Proportionality	8	2	1	11	12
Total	**125**	**19**	**7**	**151**	**163**

Table 3.18   Number of Science Items of Each Type, and Maximum Score Points, by Reporting Category, Population 2

Reporting Category	Multiple-Choice	Short-Answer	Extended-Response	Total Items	Score Points
Earth science	17	3	2	22	25
Life science	31	5	4	40	46
Physics	28	9	3	40	44
Chemistry	15	3	1	19	21
Environmental issues and the nature of science	11	2	1	14	15
Total	**102**	**22**	**11**	**135**	**151**

Tables 3.19 and 3.20 show the number of items from each reporting category that are included in each of the eight test booklets for mathematics and science. Tables 3.21 and 3.22 show the maximum number of possible score points for each reporting category in each booklet.

Table 3.19   Number of Mathematics Items in Each Booklet by Reporting Category, Population 2

Reporting Category	Booklet							
	1	2	3	4	5	6	7	8
Fractions and number sense	11	10	11	10	10	11	11	14
Geometry	5	6	6	3	6	4	5	6
Algebra	8	5	6	8	4	6	6	9
Data representation, analysis, and probability	5	4	4	6	7	6	7	5
Measurement	5	5	6	4	6	4	4	3
Proportionality	3	3	4	3	6	2	4	4
Total	**37**	**33**	**37**	**34**	**39**	**33**	**37**	**41**

Table 3.20   Number of Science Items in Each Booklet by Reporting Category, Population 2

Reporting Category	Booklet							
	1	2	3	4	5	6	7	8
Earth science	7	7	6	6	5	5	10	7
Life science	9	11	11	13	7	9	8	6
Physics	10	9	12	11	9	10	11	13
Chemistry	2	7	2	3	5	7	4	4
Environmental issues and the nature of science	6	3	3	2	3	7	1	2
Total	**34**	**37**	**34**	**35**	**29**	**38**	**34**	**32**

Table 3.21   Maximum Number of Mathematics Score Points in Each Booklet by Reporting Category, Population 2

Reporting Category	Booklet							
	1	2	3	4	5	6	7	8
Fractions and number sense	11	10	12	10	11	11	11	14
Geometry	5	6	6	3	6	4	5	6
Algebra	11	5	8	8	4	6	6	9
Data representation, analysis, and probability	5	4	4	6	9	6	9	5
Measurement	7	5	9	4	9	4	4	3
Proportionality	4	3	5	3	6	2	4	4
Total	**43**	**33**	**44**	**34**	**45**	**33**	**39**	**41**

Table 3.22   Total Number of Science Score Points in Each Booklet by Reporting Category, Population 2

Reporting Category	Booklet							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Earth science	7	9	6	6	5	5	13	7
Life science	9	13	11	15	9	9	8	8
Physics	10	9	12	14	9	12	11	14
Chemistry	2	7	2	3	5	9	4	4
Environmental issues and the nature of science	6	3	3	2	3	8	1	2
Total	**34**	**41**	**34**	**40**	**31**	**43**	**37**	**35**

Table 3.23 shows the rotation ratios for items that belong to each cluster and the number of times that items from pairs of clusters appear together. The rotation ratios for the items in each cluster are shown in the diagonal elements of the matrix in Table 3.23 (only the diagonal and lower-triangular parts of the matrix are shown). For example, the Core cluster appears in all 8 booklets, so its rotation ratio is 8. For the Focus clusters the rotation ratio is 3 (except for cluster B which has a rotation ratio of 4).

The nondiagonal elements of Table 3.23 give the number of times that a pair of clusters appears together in a booklet. If two clusters appear together, then the covariances between the items in them can be estimated (of course all the covariances between items within each cluster will be available). A dash indicates that the corresponding pair of clusters never appears in the same booklet. The first column of the matrix shows that the Core cluster appears at least once with each other cluster.   Further, the matrix shows that few covariances between the items in different Breadth clusters, and between those in different mathematics and science Free-Response clusters, will be available.   The design does, however, ensure that all covariances between items in Focus clusters and most covariances between items in Focus and Free-Response clusters, will be available.

Table 3.23   Rotation Ratios for Item Clusters and Cluster Pairings, Population 2

```
A   8    1. Core

B   4 4      2. Focus
C   3 1 3
D   3 1 1 3
E   3 1 1 13
F   3 1 1 11 3
G   3 1 1 11 13
H   3 1 1 11 11 3

I   1 1 1 -1 - - -   1      3. Breadth
J   1 - 1 1 - 1 - -    -1
K   1 - - 11 - 1 -    - -1
L   1 - - -1 1 - 1    - - -1
M   1 1 1 - -11 -    - - - -1
N   1 - - - - -1 1    - - - - -1
O   1 1 -1 - - -1    - - - - - - -1
P   1 1 - - - - - -    - - - - - - -1
Q   1 1 - - - - - -    - - - - - - -1 1
R   1 1 - - - - - -    - - - - - - -1 11

S   1 1 1 -1 - - -   1 - - - - - - - -    1      4. Mathematics Free-Response
T   2 1 1 12 -1 -   1 -1 - - - - - - -    1 2
U   2 1 - 11 12 -    - -1 -1 - - - - -    - 12
V   2 2 - 1 - 11 1    - - - -1 -1 - - -    - -1 2

W   2 1 1 2 -1 - 1    -1 - - - -1 - - -    - - -1    2      5. Science Free-Response
X   2 - 1 11 2 -1    -1 -1 - - - - - -    - - - -    1 2
Y   2 - 1 -1 11 2    - - -1 -1 - - - -    - - - -    - 1 2
Z   1 -1 - - -1 1    - - - - -1 - - - -    - - - -    - -11

    A B C D E F G H  I J K L M N O P Q R  S T U V  W X Y Z            CLUSTER
```

3.6 TIMSS POPULATION 3 TEST DESIGN

The TIMSS design for Population 3 requires the assessment of the mathematical and scientific literacy of students in their final year of secondary schooling, and of the mathematics and physics proficiency of students within that population who are taking advanced courses in those fields. The test design therefore differs significantly from those in Populations 1 and 2.

Because the educational backgrounds of the general Population 3 differs from that of the students taking advanced mathematics and physics courses, the test design had to ensure that each group received appropriate test materials. To achieve this, the students in Population 3 in each country were stratified by educational background. Each student was dichotomously characterized as being in advanced mathematics courses (M) or not (O), and as being in physics courses (P) or not (O). This two-way classification yielded four mutually exclusive and exhaustive categories:

OO   Students studying neither advanced mathematics nor physics

OP   Students studying physics but not advanced mathematics

MO   Students studying advanced mathematics but not physics

MP   Students studying both advanced mathematics and physics.

Four types of test booklets were designed to target these student categories:

- Two literacy booklets (booklets 1A and 1B) containing mathematics and science literacy items, as well as items in reasoning and social utility

- Three physics booklets (booklets 2A, 2B and 2C) containing physics items only

- Three mathematics booklets (booklets 3A, 3B and 3C) containing advanced mathematics items only

- One mathematics/physics booklet (booklet 4) containing items in physics, advanced mathematics, and reasoning and social utility items.

The design of the TIMSS tests for Population 3 builds 12 mutually exclusive clusters of items and distributes these clusters among the four types of test booklets in a systematic fashion. The 12 clusters are labeled A through L. Each cluster could appear in more than one test booklet and, in a few cases, in different positions within the booklets. The items within a cluster always appear in the same order and position.

The four types of item clusters as classified by domain in the Population 3 tests are described below.

1. One Reasoning and Social Utility cluster containing 12 items (30 minutes of testing time), of which 6 are related to concepts in mathematics and 6 to science concepts. These items may be multiple-choice, short-answer, or extended-response. This cluster is labeled cluster A.

2. Three Mathematics and Science Literacy clusters, each containing 30 minutes of testing time. These clusters are labeled B, C and D. The Core Literacy cluster (cluster B) appears in booklets 1A and 1B, and the other two clusters each appear in one of the literacy booklets.

3. The four clusters with physics items are labeled E, F, G, and H. Cluster E is a Core physics cluster that contains 30 minutes of multiple-choice items. Cluster E is the first cluster to appear in each of the physics booklets (booklets 2A, 2B, and 2C), and the second in the advanced mathematics/physics booklet (booklet 4). The remaining clusters contain multiple-choice, short-answer and extended-response items and are rotated amongst the physics booklets, with each appearing in one booklet only.

4. There are four clusters with advanced mathematics items. These clusters are labeled I, J, K and L. Cluster I is a Core mathematics cluster that contains 30 minutes of multiple-choice items. Cluster I is the first cluster to appear in each of the advanced mathematics booklets (booklets 3A, 3B, and 3C), and is the third in the advanced mathematics/physics booklet (booklet 4). The remaining clusters contain multiple-choice, short-answer, and extended-response items and are rotated amongst the advanced mathematics booklets, with each appearing in one booklet only.

5. Another way of classifying the clusters is as either Core or Rotated clusters. Table 3.24 presents this classification of the clusters by domains tested.

Table 3.24   Classification of the Clusters by Content and Cluster Type, Population 3

Domain	Core Cluster	Rotated Cluster
Reasoning and Social Utility (RSU)	A	(none)
Mathematics and Science Literacy (MSL)	B	C, D
Physics (P)	E	F, G, H
Advanced Mathematics (MA)	I	J, K, L

Again, the number of items per cluster varies because the clusters have been designed on the basis of the estimated number of minutes it would take a typical student to answer them. All four Core clusters (A, B, E and I) and the Rotated mathematics and science literacy clusters (C and D) are each 30 minutes in length. Each of the Rotated clusters for the physics and mathematics students (F, G, H, J, K and L) are 60 minutes in length. The total testing time per cluster is shown in Table 3.25.

Table 3.25   Allocation of Testing Time to Item Clusters, Population 3

Cluster Label	Cluster Type	Time Allocated
A	Reasoning and Social Utility (RSU)	30
B	Literacy Content Core (MSL)	30
C	Literacy Content Rotated a (MSLa)	30
D	Literacy Content Rotated b (MSLb)	30
E	Physics Core (PC)	30
F	Physics Rotated a (PRa)	60
G	Physics Rotated b (PRb)	60
H	Physics Rotated c (PRc)	60
I	Advanced Mathematics Core (MAC)	30
J	Advanced Mathematics Rotated a (MARa)	60
K	Advanced Mathematics Rotated b (MARb)	60
L	Advanced Mathematics Rotated c (MARc)	60

3.6.1 ORGANIZATION OF THE TEST BOOKLETS

In Population 3, the design calls for nine booklets, each estimated to require 90 minutes to complete. Each booklet has either two or three clusters of items. Table 3.26 shows the assignment of clusters to booklets, and the position of the clusters within each booklet.

Table 3.26   Assignment of Item Clusters to Population 3 Booklets[5]

Cluster Type	Cluster Label	1A	1B	2A	2B	2C	3A	3B	3C	4
RSU	A	1	2							1
MSL	B	2	1							
MSLa	C	3								
MSLb	D		3							
PC	E			1	1	1				2
PRa	F			2						
PRb	G				2					
PRc	H					2				
MAC	I						1	1	1	3
MARa	J						2			
MARb	K							2		
MARc	L								2	

[5] Number in cell indicates position of item cluster within the test booklet.

Table 3.27 summarizes the information in Table 3.26 by cluster order. It shows, for each booklet, the clusters assigned and the order in which they appear in the booklet.

Table 3.27   Ordering of Clusters Within Population 3 Booklets

Cluster Order	Booklet								
	1A	1B	2A	2B	2C	3A	3B	3C	4
1st	A	B	E	E	E	I	I	I	A
2nd	B	A	F	G	H	J	K	L	E
3rd	C	D	-	-	-	-	-	-	I

The design summarized in Tables 3.26 and 3.27 has the following features.

· Each test booklet comprises up to three item clusters, each of which can each appear in more than one booklet.

· Each of the mathematics and science literacy booklets (booklets 1A and 1B) contains the cluster with reasoning and social utility items (cluster A), the Core cluster for mathematics and science literacy (cluster B), and one of the mathematics and science literacy Rotated clusters (C or D).

· The physics booklets (booklets 2A, 2B, and 2C) contain the Core cluster for physics (cluster E), followed by one of the Rotated physics clusters (F, G, or H).

· The advanced mathematics booklets (booklets 3A, 3B and 3C) contain the Core cluster for advanced mathematics (cluster I), followed by one of the Rotated advanced mathematics clusters (clusters J, K, or L).

· The advanced mathematics/physics booklet (booklet 4) contains the reasoning and social utility cluster (cluster A), as well as the Core clusters for the physics and advanced mathematics items (clusters E and I).

· The expected completion time for reasoning and social utility (cluster A), the mathematics and science literacy (B through D), and the Core physics (E) and Core advanced mathematics (I) clusters is 30 minutes each. The expected completion time for each of the physics (F, G, H) and advanced mathematics (J, K, L) Rotated clusters is 60 minutes. As a result of the assignment of clusters to booklets, the expected completion time for each of the booklets is 90 minutes.

· Each booklet was administered in one 90-minute session with no break.

3.6.2 ASSIGNMENT OF BOOKLETS TO STUDENTS

In Populations 1 and 2, all of the test booklets were rotated through all students in the sample. In Population 3, it was necessary to specify a separate rotation scheme for each student classification: OO, MO, OP, and MP.

The booklet that a student was eligible to receive depended upon the classification of that student. Table 2.28 shows the booklets to be rotated for students of each type.

Table 3.28   Assignment of Test Booklets to Students, Population 3

Test Booklet	Student Grouping			
	OO	OP	MO	MP
1A	X	X	X	X
1B	X	X	X	X
2A		X		X
2B		X		X
2C		X		X
3A			X	X
3B			X	X
3C			X	X
4				X

The rotation ratios for items in each cluster for each student classification are shown in Table 3.29. These ratios give some indication of the relative precision of statistics that are expected at the item level. The higher the rotation ratio, the smaller the proportion of the sampled students that will respond to the item. While this figure does not address the absolute precision of item statistics, it does make it clear that items allocated to the Core clusters (RSU, MSL, MAC and PC) are likely to have more precise statistics than those assigned to the rotated clusters.

Table 3.29   Rotation Ratios for Items in Each Cluster, Assuming Uniform Rotation Within Domains, Population 3

	Student Classification			
Clusters	**OO**	**OP**	**MO**	**MP**
RSU	1	2.5	2.5	3
MSL	1	2.5	2.5	4.5
MSLa	2	5	5	9
MSLb	2	5	5	9
MAC	-	-	1.67	2.25
MARa	-	-	5	9
MARb	-	-	5	9
MARc	-	-	5	9
PC	-	1.67	-	2.25
PRa	-	5	-	9
PRb	-	5	-	9
PRc	-	5	-	9

3.6.3 CONTENT OF THE TEST BOOKLETS

Items were included in the Population 3 tests to cover several content areas in mathematics and science literacy, advanced mathematics, and physics. For the purpose of scaling and reporting, some of these content areas were merged into the reporting categories below.

The mathematics and science literacy reporting categories are:

· Mathematics literacy

· Science literacy

· Reasoning and social utility.

The physics reporting categories are:

· Mechanics

· Electricity and magnetism

· Heat

· Wave phenomena

· Particle, quantum, astrophysics and relativity.

The advanced mathematics reporting categories are:

· Numbers and equations

· Analysis (calculus)

· Geometry

- Probability and statistics

- Validation and structure.

The TIMSS test blueprints (see Chapter 2) and the curriculum frameworks describe in more detail the content areas upon which the reporting categories are based.

The Core literacy, advanced mathematics, and physics clusters contain multiple-choice items only. The Rotated clusters were composed of multiple-choice, short-answer, and extended-response items. In the literacy and reasoning and social utility clusters, the items were grouped by subject area (science and mathematics) within the cluster.

Tables 3.30 through 3.35 summarize the test design for Population 3. Tables 3.30 through 3.32 present for each scale, the number of items in each booklet for each reporting category. Tables 3.33 through 3.35 present for each scale, the maximum number of possible score points in each booklet for each reporting category.

Table 3.30   Number of Test Items per Booklet by Reporting Category, Population 3 Mathematics and Science Literacy

Reporting Category	Booklet								
	1A	1B	2A	2B	2C	3A	3B	3C	4
Mathematics literacy	26	25	-	-	-	-	-	-	-
Science literacy	21	18	-	-	-	-	-	-	-
Reasoning and social utility	12	12	-	-	-	-	-	-	12
Total	**59**	**55**	-	-	-	-	-	-	**12**

Table 3.31   Number of Test Items per Booklet by Reporting Category, Population 3 Physics

Reporting Category	Booklet								
	1A	1B	2A	2B	2C	3A	3B	3C	4
Mechanics	-	-	6	8	6	-	-	-	2
Electricity and magnetism	-	-	7	7	8	-	-	-	3
Heat	-	-	3	4	4	-	-	-	1
Wave phenomena	-	-	5	4	5	-	-	-	2
Particle, quantum, astrophysics, and relativity	-	-	6	6	6	-	-	-	2
Total	-	-	**27**	**29**	**29**	-	-	-	**10**

**Table 3.32   Number of Test Items per Booklet by Reporting Category,
Population 3 Advanced Mathematics**

Reporting Category	Booklet								
	1A	1B	2A	2B	2C	3A	3B	3C	4
Numbers, equations, and functions	-	-	-	-	-	7	8	8	3
Analysis (calculus)	-	-	-	-	-	7	7	5	2
Geometry	-	-	-	-	-	10	10	9	3
Probability and statistics	-	-	-	-	-	3	2	4	1
Validation and structure	-	-	-	-	-	2	1	2	1
Total	-	-	-	-	-	**29**	**28**	**28**	**10**

**Table 3.33   Number of Score Points per Booklet by Reporting Category,
Population 3 Mathematics and Science Literacy**

Reporting Category	Booklet								
	1A	1B	2A	2B	2C	3A	3B	3C	4
Mathematics literacy	28	30	-	-	-	-	-	-	-
Science literacy	24	21	-	-	-	-	-	-	-
Reasoning and social utility	21	21	-	-	-	-	-	-	21
Total	**73**	**72**	-	-	-	-	-	-	**21**

**Table 3.34   Number of Score Points per Booklet by Reporting Category,
Population 3 Physics**

Reporting Category	Booklet								
	1A	1B	2A	2B	2C	3A	3B	3C	4
Mechanics	-	-	7	10	6	-	-	-	2
Electricity and magnetism	-	-	9	8	10	-	-	-	3
Heat	-	-	4	5	5	-	-	-	1
Wave phenomena	-	-	5	4	7	-	-	-	2
Particle, quantum, astrophysics and relativity	-	-	7	7	7	-	-	-	2
Total	-	-	**32**	**34**	**35**	-	-	-	**10**

Table 3.35   Number of Score Points per Booklet by Reporting Category, Population 3 Advanced Mathematics

Reporting Category	Booklet								
	1A	1B	2A	2B	2C	3A	3B	3C	4
Numbers, equations, and functions	-	-	-	-	-	7	11	10	3
Analysis (calculus)	-	-	-	-	-	9	9	5	2
Geometry	-	-	-	-	-	12	12	11	3
Probability and statistics	-	-	-	-	-	3	2	5	1
Validation and structure	-	-	-	-	-	3	1	2	1
Total	**-**	**-**	**-**	**-**	**-**	**34**	**35**	**33**	**10**

Tables 3.36 through 3.38 present the number of items in the Population 3 item pool, organized by item type and by reporting category. It also presents the maximum number of score points in each of the Population 3 reporting categories.

Table 3.36   Number of Test Items of Each Type, and Maximum Score Points, by Reporting Category, Population 3 Mathematics and Science Literacy

Reporting Category	Item Type				
	Multiple-Choice	Short-Answer	Extended-Response	Number of Items	Score Points
Mathematics literacy	31	7	-	38	42
Science literacy	16	7	3	26	43
Reasoning and social utility	5	3	4	12	21
Total	**52**	**17**	**7**	**76**	**106**

Table 3.37   Number of Test Items of Each Type, and Maximum Score Points, by Reporting Category, Population 3 Physics

Reporting Category	Item Type				
	Multiple-Choice	Short-Answer	Extended-Response	Number of Items	Score Points
Mechanics	11	4	1	16	19
Electricity and magnetism	10	3	3	16	21
Heat	6	3	-	9	12
Wave phenomena	6	3	1	10	12
Particle, quantum, astrophysics, and relativity	9	2	3	14	17
Total	**42**	**15**	**8**	**65**	**81**

Table 3.38   Number of Test Items of Each Type and Score Points, by Reporting Category, Population 3, Advanced Mathematics

Reporting Category	Item Type				
	Multiple-Choice	**Short-Answer**	**Extended-Response**	**Number of Items**	**Score Points**
Numbers, equations, and functions	13	2	2	17	22
Analysis (calculus)	12	2	1	15	19
Geometry	15	4	4	23	29
Probability and statistics	5	2	-	7	8
Validation and structure	2	-	1	3	4
Total	**47**	**10**	**8**	**65**	**82**

REFERENCES

Beaton, A.E. (1987). *Implementing the New Design: The NAEP 1983-84 Technical Report.* Report No. 15-TR-2. Princeton, NJ: Educational Testing Service.

Lord, F.M. (1962). Estimating Norms By Item-sampling. *Educational and Psychological Measurement,* 22(2), 259-267.

Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Mislevy, R. and Sheehan, K. (1987). "Marginal Estimation Procedures" in A.E. Beaton (ed.), *Implementing the New Design: The NAEP 1983-84 Technical Report.* Report No: 15-TR-2. Princeton, NJ: Educational Testing Service.

Robitaille, D.F., Schmidt, W.H., Raizen, S.A., McKnight, C.C., Britton, E., and Nicol, C. (1993). *TIMSS Monograph No. 1: Curriculum Frameworks for Mathematics and Science.* Vancouver, Canada: Pacific Educational Press.

Rubin, D. B. (1987). *Multiple Imputation For Nonresponse in Surveys.* New York: John Wiley & Sons.

Wright, B. and Stone, M. (1979). *Best Test Design.* Chicago: MESA Press, University of Chicago.